



Principal Components Biplots and Alpha and Beta Diversity

Cajo J. F. ter Braak

Ecology, Vol. 64, No. 3 (Jun., 1983), 454-462.

Stable URL:

<http://links.jstor.org/sici?sici=0012-9658%28198306%2964%3A3%3C454%3APCBAAA%3E2.0.CO%3B2-9>

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

Ecology is published by The Ecological Society of America. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/esa.html>.

Ecology

©1983 The Ecological Society of America

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2003 JSTOR

PRINCIPAL COMPONENTS BILOTS AND ALPHA AND BETA DIVERSITY¹

CAJO J. F. TER BRAAK

*Institute TNO for Mathematics, Information Processing, and Statistics,
P.O. Box 100, 6700 AC Wageningen, The Netherlands*

Abstract. Attention is drawn to some useful but not generally known properties of principal components analysis (PCA). Noncentered PCA of proportion data gives site ordinations that display approximate *alpha diversities* of sites and *beta diversities* of groups of sites, as measured by the Simpson index and mean squared Euclidean distance, respectively. Species centering allows a better approximation to beta diversities. Alpha diversities can still be visualized after centering if the true origin is projected into the plane of the ordination. The approximate species composition of each site can also be visualized if the site ordination is combined with a species ordination. The resulting plot of site scores and species loadings is called a PCA *biplot*. Finally, in a PCA biplot that displays both species composition and diversity, diversity values can be explained in terms of the main species contributing to diversity. In such a biplot the sum of squares of the species loadings must be scaled to unity, while the site scores must be scaled to a sum of squares equal to the corresponding eigenvalue. This type of biplot is termed a "distance biplot." For a simple illustration noncentered and species-centered distance biplots were produced for some diatom samples taken from Dutch moorland pools in the 1920s and 1978. The distance biplot is concluded to be among the most powerful analytical tools for species-composition data and derives some of its power from properties not possessed by, for example, reciprocal averaging. One problem is that it attaches little weight to rare species, but this problem can be solved by various possible data transformations based on the theory of diversity indices.

Key words: *biplot; dissimilarity; diversity; graphical representation; ordination; principal components analysis; Simpson index; species composition.*

INTRODUCTION

Simultaneous site and species ordinations are not unique to reciprocal averaging (Hill 1973a). Gabriel (1971) has shown that in ordinary principal components analysis (PCA) both objects and variables, i.e., sites and species, can be combined into one plot: the biplot. The biplot is a powerful graphical tool (see also Laurec et al. 1979).

Diversity has an elegant mathematical theory (Engen 1978, Kempton 1979, Patil and Taillie 1979) and an elaborate ecological theory as well (Mac Arthur and Wilson 1967, Whittaker 1972, May 1973, Pielou 1975). Diversity measures describe community structure without reference to named species; sites with totally different species lists may have identical diversity. A diversity measure thus gives a very limited summary of data. It is also important to know which species contribute to the diversity of each site.

This paper presents a method, based on PCA biplots, to display sites and species in such a way that the diversity values of sites can be interpreted in terms of the contributing species. For simplicity I shall concentrate on measures related to the Simpson index (Simpson 1949). The method is illustrated with a small example data set consisting of diatom assemblages from six Dutch moorland pools, sampled in the 1920s and again in 1978. These data were collected in connection

with a project of the Research Institute for Nature Management (The Netherlands) to investigate the impact of acidification on these shallow water bodies (Van Dam et al. 1981).

THEORY AND METHODS

Geometrical representation of proportions and diversity

Suppose a survey of n sites lists the proportions of m species. Let p_{ki} be the proportion of individuals of species k at site i with respect to all individuals at site i ($\sum_{k=1}^m p_{ki} = 1$). Each of the n sites may now be represented by a point or vector \mathbf{p}_i in a m -dimensional Euclidean space: $\mathbf{p}_i = (p_{1i}, \dots, p_{mi})'$. The axes of this space are generated by the "basis vectors" $\mathbf{e}_1 = (1, 0, 0, \dots, 0)'$, $\mathbf{e}_2 = (0, 1, 0, \dots, 0)'$, \dots , $\mathbf{e}_m = (0, 0, 0, \dots, 1)'$, each basis vector (say \mathbf{e}_k) being associated with a species (k).

If there are just three species, the sites are points in a three-dimensional space. The closed circles in Fig. 1 represent five sites. They are located on the triangular plate, shown in Fig. 1, formed by connecting the endpoints of the basis vectors \mathbf{e}_1 , \mathbf{e}_2 , and \mathbf{e}_3 . For each site, the proportion of individuals of a particular species is equal to the length of the (orthogonal) projection of the site onto the axis that represents that species, as illustrated in Fig. 1 for site 2.

The length of each site vector \mathbf{p}_i is the distance between the site vector and the origin and is denoted by

¹ Manuscript received 18 November 1980; revised 27 March 1982; accepted 26 May 1982.

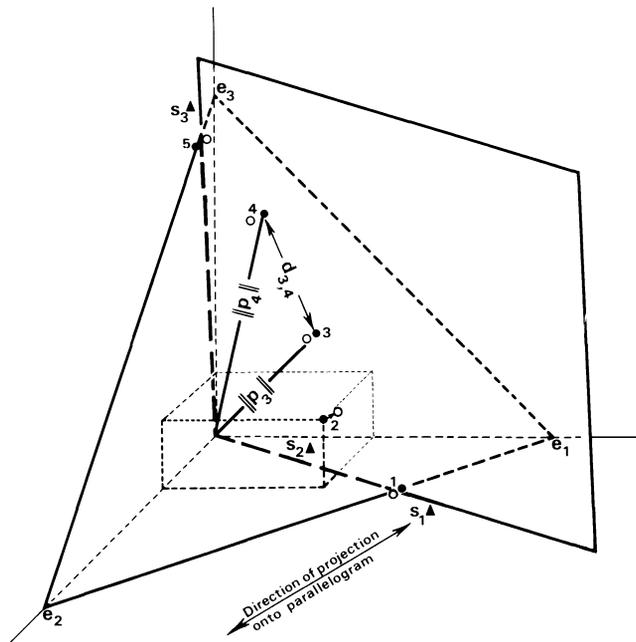


FIG. 1. Geometrical representation of the proportions of three species at five sites (●) with the projection points of the sites (○) and the species (▲) in the noncentered PCA plane (parallelogram). Artificial data from Jöreskog et al. (1976). For explanation see text.

$\| \mathbf{p}_i \|$. In Fig. 1 the lengths of sites 3 and 4 are indicated. The squared length of each site vector \mathbf{p}_i is given by

$$\| \mathbf{p}_i \|^2 = \sum_{k=1}^m p_{ki}^2, \tag{1}$$

which is exactly the formula for the (uncorrected) Simpson index (Simpson 1949). The Simpson index, henceforth abbreviated to SI, can easily be transformed into a useful measure of alpha (within-site) diversity: for example $1 - SI$ (Pielou 1969) or $1/SI$ (Hill 1973b). McIntosh (1967) developed his index of diversity, given by one minus the length of a site vector, in a similar geometrical context, and Bhargava and Doyle (1974) studied the geometry of a more general class of species diversity measures.

Squared Euclidean distance between sites,

$$d_{ij}^2 = \sum_{k=1}^m (p_{ki} - p_{kj})^2, \tag{2}$$

is a measure of dissimilarity between sites (as illustrated for sites 3 and 4 in Fig. 1) and may be regarded as a measure of beta (between-site) diversity (Whittaker 1960). "Beta or between-habitat diversity refers to the . . . degree of contrast in species composition among samples of a set taken from a landscape" (Whittaker 1973:41). I define the beta diversity of two samples as the dissimilarity between them, according to Eq. 2. If there are more than two samples, their

mean dissimilarity is a measure of their beta diversity. This definition is useful for the present purpose of displaying the actual data. In the definition that is in common use, beta diversity is the amount of species turnover along an environmental gradient (Gauch 1973) and cannot be measured until the underlying structure of the data has been recovered successfully (Noy-Meir and Whittaker 1977). With the present definition beta diversity is not restricted to coenoclines and is easy to measure.

Noncentered PCA

The geometrical configuration of the sites has now been completely specified in terms of alpha and beta diversity measures. However, if there are more than two or three species, this configuration has too many dimensions for convenience. We would like a two- or three-dimensional plot representing as accurately as possible the original configuration. We are anxious to preserve both the lengths of site vectors and the distances between them; we therefore need to know the position in the plot of the "true" origin, i.e., the intersection of $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m$, a notional site without species. Noncentered PCA guarantees this (Noy-Meir 1973). Noncentered PCA simply projects the sites onto the plane of best fit through the true origin.

The plane of best fit through the origin is indicated in Fig. 1 by a parallelogram intersecting the triangular plate according to the line joining approximately sites

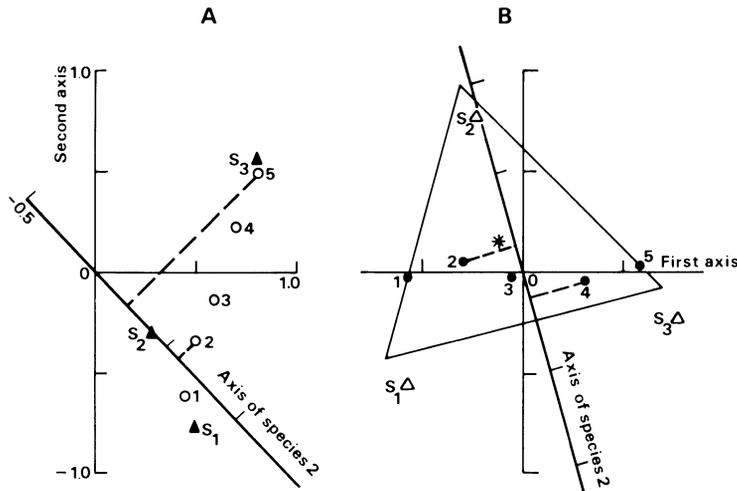


FIG. 2. PCA biplots of Fig. 1 that display (approximate) Simpson index values of sites, Euclidean distances between sites, as well as species proportions. (A) noncentered PCA. (B) species-centered PCA. For explanation see text (horizontally and vertically the first and second principal component axes; ● and ○ = sites; ▲ and △ = species; * = projection of origin).

1 and 5. The sites are projected onto this plane along the direction of projection, indicated in Fig. 1 (see also site 2). The projection point of each closed circle is given as an open circle; the open circles lie in the plane of best fit through the origin. Fig. 2A shows this plane in front view after a rotation around the origin. (The rotation is such that the horizontal axis in Fig 2A is the *line* of best fit through the origin.) In fact, Fig. 2A is an ordinary ordination plot, based on noncentered PCA, with horizontally the first, and vertically the second principal component axis. The coordinates of a site on these axes are usually termed the scores of the site on the first and second principal component axes.

The result of a noncentered PCA is thus an ordination plot in which sites with high alpha diversity (low SI) are likely to be near the origin, and sites with low alpha diversity (high SI) are likely to be far from the origin. Sites that are similar in species composition, i.e., with low beta diversity as measured by mean squared Euclidean distance, will be close together, while sites that differ in species composition will be more spread out, so long as the reduction in dimensionality has not resulted in too much distortion of the original configuration. In Fig. 2A sites 2 and 3 thus have a low SI; site 5 has the highest SI.

The biplot

Distances between sites in an ordination plot are, of course, due to differences in species composition (Eq. 2, Fig. 1). These differences in species composition can be displayed explicitly, as they were in the original *m*-dimensional space, by combining the site ordination with a species ordination. The site ordination resulted from projection of the sites onto an optimal plane. The

appropriate species ordination now follows from projection of the basis vectors of the *m*-dimensional space onto this plane. In Fig. 1 the basis vectors e_1 , e_2 , and e_3 are projected onto the plane of the parallelogram along the direction indicated; the projection points are labeled s_1 , s_2 , and s_3 . In front view (Fig. 2A) the result is a principal-component biplot (Gabriel 1971). The coordinates of a species with respect to the principal component axes in this plot are usually termed the loadings of the species on the first and second components.

The proportions of each species at each site can be recovered approximately from the biplot in very much the same way as in Fig. 1, viz, by projecting each site onto each species-axis. The axis of a species in a biplot is obtained by drawing a line joining the species-point to the origin of the coordinate system. This process is illustrated in Fig. 2A for species 2. Scale values along this axis are simply (signed) distances from the origin; the points at which the scale values along this axis are $-.5$, $.5$, and 1.0 are shown in Fig 2A by scale marks. Note that the length of the vector s_2 of species 2 is $\approx .4$. The projection of site 5 onto the axis gives a projection point that has a scale value of $\approx .2$. The proportion of species 2 at site 5, as approximated by this biplot, is now simply $.4 \times .2 = .08$; the true proportion was $.10$. Similarly, the proportion at site 2 is $\approx .4 \times .6 = .24$. More generally, the proportion of a species at a site is given approximately by the product of (1) the length of the species vector and (2) the scale value at the projection point of the site onto the appropriate species-axis. Moreover the rôle of species and site may be interchanged. Hence the species composition of a site can be approximated by projecting

all the species onto the appropriate site vector, and the proportions of a species at all the sites can be approximated by projecting the sites onto the appropriate species vector.

Diversity values can now be explained in terms of species. By and large, species with large vectors pointing in roughly the same direction as a site vector are present in high proportions at that site. A site with low diversity will have only a few long species vectors pointing in its direction. A site with high diversity will have several, shorter species vectors that point in its direction. The biplot thus displays which species make a site as diverse as it is.

Species-centered PCA

Species-centered PCA (Orlóci 1966) is more commonly used than noncentered PCA and is termed ordinary PCA by Prentice (1980). In species-centered PCA $p_{ki} - p_k$. (where $p_k = [1/n] \sum_i p_{ki}$, the species mean) is used to derive the principal components, instead of p_{ki} (Jöreskog et al. 1976, Noy-Meir 1973). Geometrically this means that the origin of the coordinate system generated by e_1, e_2, \dots, e_m is translated to the centroid of the sites. The axes of the plane of best fit now pass through the centroid of the sites, instead of through the true origin as they do in noncentered PCA.

In Fig. 1 the centroid has coordinates (.36, .20, .44), with respect to $e_1, e_2,$ and e_3 , and lies near site 3 on the triangular plane. This plane is in fact the plane of best fit. As all sites lie on this plane, projection of the sites onto this plane leaves the positions of the sites unchanged. Fig. 2B is the usual ordination plot and shows this plane after a rotation in front view. The species-points are the projections of the (translated) basis vectors onto this plane. In Fig. 2B those species-points are indicated by $s_1, s_2,$ and s_3 , but they are not the ones shown in Fig. 1.

The proportions of a species at each site are approximated in a species-centered biplot as deviations from the mean proportion of the species in the sample set. These deviations can be derived as in the noncentered biplot: each species-axis is obtained by drawing a line joining the species-point to the origin of the coordinate system, i.e., now the centroid. Scale values along this axis are signed lengths from the centroid, and the approximate deviation can be found as before by multiplying (1) the length of the species vector by (2) the scale value at the projection point of the site onto the appropriate species-axis of the species. For species 2 in Fig. 2B these deviations from the mean (.20) are at site 2 (.8 × .14 = .11) and at site 4 (.8 × -.13 = -.10), which give .20 + .11 = .31 and .20 - .10 = .10 as the proportions of species 2 at site 2 and site 4, respectively. The ranking of proportions of a given species among sites can thus be seen immediately from the order of the projection-points of sites along the species-axis. In particular, the proportion is

higher than the species mean if the projection point lies, along this axis, on the same side of the centroid as the species-point does, and lower than the species mean if the centroid lies between the projection point and the species-point.

In species-centered PCA the origin of the coordinate system is shifted from the true origin to the centroid of the sites. It is clear that distances between sites are not affected by this shift, but lengths are; hence, Euclidean distances are displayed, indeed more accurately than before, but not the SI values. For these we need to know the position of the true origin. This position is determined by projecting the true origin onto the plane of the biplot and calculating the distance from the true origin to its projection. If the loadings of the m species on the r^{th} principal component are $v_{1r}, v_{2r}, \dots, v_{mr}$, respectively ($\sum_{k=1}^m v_{kr}^2 = 1$), the projection z of the true origin has the score:

$$z_r = -(v_{1r}p_{1.} + v_{2r}p_{2.} + \dots + v_{mr}p_{m.}) \quad (3)$$

on the r^{th} principal component, i.e., z_r is derived by the usual formula (Orlóci 1966) for the score of a point with coordinates $(-p_{1.}, -p_{2.}, \dots, -p_{m.})$ with respect to the centroid. Now z can be plotted in the same way as could the sites. The squared distance from the origin to z in a two-dimensional biplot is equal to:

$$s^2 = \left\{ \sum_{k=1}^m p_{k.}^2 \right\} - (z_1^2 + z_2^2). \quad (4)$$

The origin therefore lies exactly s above its projection onto the biplot. The SI of a site in the biplot is approximated by s^2 plus the squared distance between the site and the projection of the origin. The order of SI's of the sites can therefore be seen at once by looking at distances from z . High-diversity sites will be near to z ; low-diversity sites will be further away.

In Fig. 1 the projection of the origin onto the triangular plate has coordinates ($1/3, 1/3, 1/3$) in three-dimensional space, and the squared distance from the origin to this point of highest diversity equals $1/3$. SI values can be derived easily from distances to the point ($1/3, 1/3, 1/3$). This point is indicated in Fig. 2B by an asterisk. By recording the distance of the sites to the asterisk, it can be seen that sites 2 and 3 are more diverse than sites 1 and 5.

Because the data of Fig. 1 are proportions of just three species, the sites can be plotted in the form of a triangular compositional diagram (e.g., Jöreskog et al. 1976:93). This diagram is shown in a rotated fashion in Fig. 2B. The asterisk is the midpoint of this diagram, while the species-points coincide with the corner points when shifted over the distance and the direction that are given by the centroid and the asterisk. A triangular compositional diagram is in fact a biplot in which the species-axes each go through the midpoint and a corner point and make an angle of 120° .

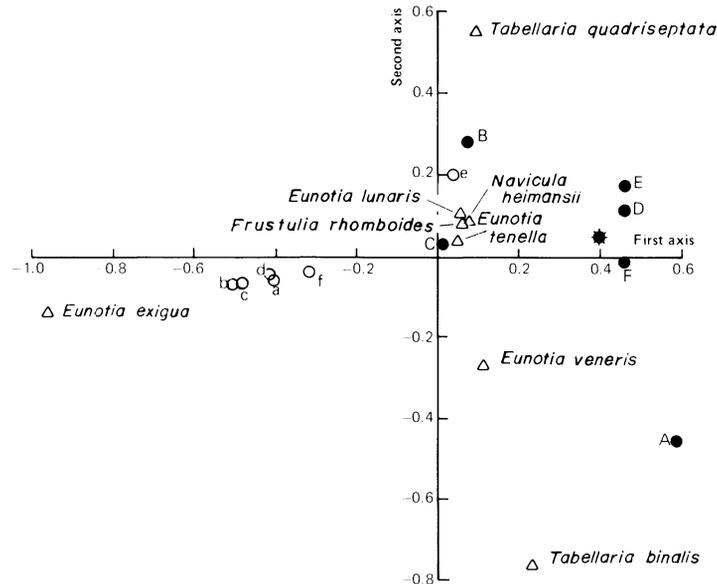


FIG. 3. Distance biplot by species-centered PCA of the diatom samples, using all 82 species. The eight most abundant species (Δ) are shown. Simpson index values are approximated by $.077$ plus squared distances from the projection of the origin (*). Codes as in Table 1.

tors) is often possible but is unwanted in the present methods.

PCA programs produce species loadings, site scores, and eigenvalues for each principal component. (Loadings are sometimes termed eigenvector coefficients.) The scores for the projection point of the origin can easily be calculated from the loadings and the species means. In order to preserve Euclidean distances in the biplot (1) the loadings on each component must have a sum of squares equal to unity, and (2) the scores on a component must have a sum of squares equal to the corresponding eigenvalue (Gabriel 1971). I propose calling this type of biplot a "distance biplot," to distinguish it from the "covariance biplot" considered in detail by Corsten and Gabriel (1976). Some programs return scores that have a sum of squares of unity, in which case the scores should be multiplied by the square root of the appropriate eigenvalue before drawing a scatter diagram. Similarly it may be necessary to divide the loadings by the square root of the eigenvalue in order to get correctly scaled loadings. The scale units along abscissa and ordinate must have the same physical length. Laurec et al. (1979) illustrate the error made if the loadings and scores are scaled wrongly with respect to each other. In practice it may be useful to draw separate plots of sites and species on transparent paper, each one with its own scaling. Comparison can then be made by superimposing the plots.

RESULTS

The samples are a subset of those analysed in Van Dam et al. (1981), and data on the eight most abundant

species appear in Table 1. The last row of Table 1 shows SI values for each diatom sample; Table 2 gives squared Euclidean distances. All 82 species have been used in the computation of SI values, squared Euclidean distances, and the principal components.

As species-centered PCA is more efficient than non-centered PCA in approximating beta diversities in a two-dimensional scatter diagram, the species-centered distance biplot of the data is shown in Fig. 3. It is now of interest to compare squared Euclidean distances between samples with the corresponding squared distances in Fig. 3 and SI values for samples with their approximation in Fig. 3:

The 1978 samples (except sample e) have low pairwise squared Euclidean distances ($< .04$), hence their proximity in Fig. 3. Their squared distances from sample e are higher (.2 to .5); hence, sample e is placed somewhat apart. Mean squared distances for samples from the 1920s and 1978 are .41 and .12, respectively; the 1920 samples are therefore more spread out in Fig. 3. The fraction of the squared intersite distances accounted for by Fig. 3 is .83.

In order to infer approximate SI values we require the position of the true origin. The origin lies .277 above its projection which is indicated by an asterisk in Fig. 3. The SI for any site is therefore approximated by $.277^2$ plus the squared distance from the site to the asterisk. The 1920 samples D, E, and F (with SI .1 to .2 in Table 1) are shown to have high diversity: they lie near the asterisk. Samples B, C, and e (SI .2-.4) are shown to be less diverse. Sample A has SI .46 and is accordingly positioned farther away from the asterisk.

isk. The remaining 1978 samples have low diversity (SI .5–1.0) and are placed a long way from the projection of the origin. In Fig. 3 there is no indication of distortion of the original configuration.

The points representing species in Fig. 3 transform the site ordination in a biplot. In this biplot the proportions of the species are approximated as deviations from the species mean (Table 1: last column). The origin of the coordinate system, the point (0, 0), is the centroid of the sites. The positions of the eight most abundant species are shown: the remaining species have positions very close to the centroid.

Eunotia exigua points west (left) from the centroid. Projection of the samples onto this direction indicates that *Eunotia exigua* is present with high proportions in all the 1978 samples except e, with a proportion about equal to its mean (.48) in sample C, and with lower proportions in the remaining samples. These predictions agree with Table 1. *Tabellaria quadriseptata* has mean .10 and points north (up) from the centroid. Projection onto this direction shows that *Tabellaria quadriseptata* occurs with higher proportions in the samples B, E, e, and D, with proportions of \approx .10 in samples C, and F, and with lower proportions in the remaining samples. Table 1 does not confirm these statements fully; sample D in particular contains less of the species than predicted. The high proportions of *Tabellaria binialis* and *Eunotia veneris* in sample A are nicely represented in Fig. 3. The overall lack of fit is .48, i.e., 8% of the total sum of squares of all proportions (5.77). The lack of fit in the noncentered PCA (not shown) is worse (.74 or 13%). The percentage of the variance accounted for by Fig. 3 is 83%.

In conclusion, from Fig. 3 the impoverishment of the pools since the 1920s can be seen at once: many species have declined as *Eunotia exigua* has become the only dominant species. As a result the alpha diversity of the pools and the beta diversity among the pools have both decreased. Van Dam et al. (1981) discuss these findings in more depth.

DISCUSSION

The utility of distance biplots

Patterns of species composition are commonly summarized by means of ordination methods, by measures of diversity, or simply by reporting the main species with their abundances. The distance biplot unites these summaries: it is a form of PCA that allows approximate species composition and diversity to be displayed simultaneously and allows alpha and beta diversity values to be associated with the main contributing species. It is therefore among the most powerful tools available for the analysis of species-composition data. Yet the biplot is simply a specially scaled combination of principal component scores and loadings. The display of alpha diversities in species-centered PCA requires just one extra point to be plot-

ted, indicating the position of the true origin, i.e., the "site without species." Correlations between PCA results and values of alpha diversity measures have been found empirically by Goldsmith (1973) and Miracle (1974) and are evident in Ritchie and Yarranton (1978). This paper explains these findings.

The importance of species loadings in interpreting the site ordinations has already been stressed by Laurec et al. (1979). The positions of the species in a species-centered PCA biplot summarize the familiar series of diagrams, one for each species, in which the abundance values of each species are superimposed on a site ordination (e.g., Orłóci 1966, Goldsmith 1973). If a species' values show a marked trend in the ordination, that species' axis in the biplot will be oriented in the direction of the trend. This property is in contrast with reciprocal averaging, which tends to place each species-point in the center of gravity of the sites in which the species occurs.

Possible modifications to increase sensitivity to rare species

The use of SI and squared Euclidean distance as alpha and beta diversity measures has certain advantages. Both measures have a convenient probabilistic interpretation in terms of random draws of two organisms from communities (Simpson 1949, Dyer 1978); their estimators have a small coefficient of variation in comparison with many other diversity measures, and sampling error in species composition is conveniently related to SI. The main disadvantage of these indices is their relative insensitivity to rare species (Peet 1974). The corresponding PCA tends to emphasize the species-poor sites that are dominated by a few dominant species, whereas the species-rich sites appear more similar (e.g., McIntosh 1967), although this phenomenon was hardly discernible in Fig. 3. Among the possible solutions to this problem is the use of some kind of data transformation before applying PCA; for an excellent review see Noy-Meir and Whittaker (1977). From a diversity point of view a rigorous solution to this problem was the creation of families of alpha diversity measures (Hurlbert 1971, Hill 1973b, Smith and Grassle 1977, Patil and Taillie 1979), indexed by a parameter that determines the weight given to common as against rare species. A similar approach to beta diversity measures is possible too (cf. Grassle and Smith 1976). The question arises whether these alternative diversity measures can also be linked to PCA through biplots.

The basic idea of Hurlbert (1971) and Smith and Grassle (1977) was that a sample tends to contain more species with increasing sample size; hence, species richness (the number of species in a sample) should be standardized in terms of sample size. This idea leads to a definition of the expected number of species in a random sample of, say, r individuals from a commu-

nity. The probability that species k will be present in a random sample of r individuals is

$$y_{ki} = 1 - (1 - p_{ki})^r, \quad r = 1, 2, \dots \quad (5)$$

The sum of these probabilities over the species is the required expected number of species (Smith and Grassle 1977). Elsewhere I shall show that in a geometrical representation of the sites, with these probabilities as coordinates, the squared length of a site is strongly positively correlated with the expected number of species in a random sample of r individuals, provided r is reasonably large (say $r \geq 10$). Hence, when for such value of r these probabilities are depicted in a species-centered biplot, the low-diversity sites are likely to be near the projection point of the origin, while the high-diversity sites will tend to be further away from this point. In contrast, using Euclidean distances the high-diversity sites were near the projection of the origin. The ordination of presence-absence data by PCA is just the special case where $r \rightarrow \infty$. Sites with few species will be near the projection point of the origin, while sites with many species are likely to be further away, as shown by Banfield and Gower (1980) in a different context. Another similar approach is possible for the families of alpha diversity measures proposed by Hill (1973*b*) and Patil and Taillie (1979) through the power transformation p_{ki}^a ($-\infty < a < \infty$). Note however that ordination methods that involve any kind of standardization of species or sites, such as reciprocal averaging (Hill 1973*a*), do not produce results that can be interpreted in terms of alpha diversities.

ACKNOWLEDGMENTS

I would like to thank Mr. H. van Dam, who suggested to me the relation between beta diversity and ordination-plots. He kindly permitted use of his data here. I am grateful to Drs. I. C. Prentice, H. J. B. Birks, M. O. Hill, H. J. P. Eijsackers, J. B. van Biezen, my colleagues, and the referees for valuable comments.

LITERATURE CITED

- Alvey, N. G. et al. 1977. GENSTAT: a general statistical program. Rothamsted Experimental Station, Harpenden, Hertfordshire, England.
- Banfield, C. F., and J. C. Gower. 1980. A note on the graphical representation of multivariate binary data. *Applied Statistics* 29:238-245.
- Barr, A. J., J. H. Goodnight, J. P. Sall, and J. T. Helweg. 1976. A user's guide to SAS 76. SAS Institute, Raleigh, North Carolina, USA.
- Bhargava, T. N., and P. H. Doyle. 1974. A geometric study of diversity. *Journal of Theoretical Biology* 43:241-251.
- Corsten, L. C. A., and K. R. Gabriel. 1976. Graphical exploration in comparing variance matrices. *Biometrics* 32: 851-863.
- Dixon, W. J., and M. B. Brown. 1977. BMDP: biomedical computer programs. P-series. University of California Press, Berkeley, California, USA.
- Dyer, D. P. 1978. An analysis of species dissimilarity using multiple environmental variables. *Ecology* 59:117-125.
- Engen, S. 1978. Stochastic abundance models. Chapman and Hall, London, England.
- Gabriel, K. R. 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58:453-467.
- Gauch, H. G., Jr. 1973. The relationship between sample similarity and ecological distance. *Ecology* 54:618-622.
- . 1977. ORDIFLEX—a flexible computer program for four ordination techniques: weighted averages, polar ordination, principal components analysis, and reciprocal averaging. Release B. Ecology and Systematics, Cornell University, Ithaca, New York, USA.
- Goldsmith, F. B. 1973. The vegetation of exposed sea cliffs at South Stack, Anglesey. I. The multivariate approach. *Journal of Ecology* 61:787-818.
- Grassle, J. F., and W. Smith. 1976. A similarity measure sensitive to the contribution of rare species and its use in investigation of variation in marine benthic communities. *Oecologia* (Berlin) 25:13-22.
- Hill, M. O. 1973*a*. Reciprocal averaging: an eigenvector method of ordination. *Journal of Ecology* 61:237-249.
- . 1973*b*. Diversity and evenness: a unifying notation and its consequences. *Ecology* 54:427-432.
- Hurlbert, S. H. 1971. The non-concept of species diversity: a critique and alternative parameters. *Ecology* 52:577-586.
- Jöreskog, K. G., J. E. Klován, and R. A. Reymont. 1976. Geographical factor analysis. Elsevier Scientific Publishing, Amsterdam, The Netherlands.
- Kempton, R. A. 1979. The structure of species abundance and measurement of diversity. *Biometrics* 35:307-321.
- Laurec, A., P. Chardy, P. de la Salle, and M. Rickaert. 1979. Use of dual structures in inertia analysis: ecological implications. Pages 127-174 in L. Orlóci, C. R. Rao, and W. M. Stiteler, editors. *Multivariate methods in ecological work*. Volume S7. Statistical ecology. International Cooperative Publishing House, Fairland, Maryland, USA.
- MacArthur, R. H., and E. O. Wilson. 1967. The theory of island biogeography. Princeton University Press, Princeton, New Jersey, USA.
- May, R. M. 1973. Stability and complexity of model ecosystems. Princeton University Press, Princeton, New Jersey, USA.
- McIntosh, R. P. 1967. An index of diversity and the relation of certain concepts to diversity. *Ecology* 48:392-404.
- Miracle, M. R. 1974. Niche structure in freshwater zooplankton: a principal components approach. *Ecology* 55: 1306-1316.
- Nie, N. H., C. A. Hull, J. G. Jenkins, K. Steinbrenner, and D. H. Bent. 1975. SPSS: statistical package for the social sciences. McGraw-Hill, New York, New York, USA.
- Noy-Meir, I. 1973. Data transformations in ecological ordination. I. Some advantages of non-centering. *Journal of Ecology* 61:329-341.
- Noy-Meir, I., and R. H. Whittaker. 1977. Continuous multivariate methods in community analysis: some problems and developments. *Vegetatio* 33:79-98.
- Orlóci, L. 1966. Geometric models in ecology. I. The theory and application of some ordination methods. *Journal of Ecology* 54:193-215.
- . 1978. *Multivariate analysis in vegetation research*. Second edition. W. Junk, The Hague, The Netherlands.
- Patil, G. P., and C. Taillie. 1979. An overview of diversity. Pages 3-27 in J. F. Grassle, G. P. Patil, W. Smith, and C. Taillie, editors. *Ecological diversity in theory and practice*. Volume S6. Statistical ecology. International Cooperative Publishing House, Fairland, Maryland, USA.
- Peet, R. K. 1974. The measurement of species diversity. *Annual Review of Ecology and Systematics* 5:285-307.
- Pielou, E. C. 1969. An introduction to mathematical ecology. Wiley-Interscience, New York, New York, USA.
- . 1975. *Ecological diversity*. J. Wiley and Sons, New York, New York, USA.

- Prentice, I. C. 1980. Multidimensional scaling as a research tool in quaternary palynology: a review of theory and methods. *Review of Paleobotany and Palynology* **31**:71–104.
- Ritchie, J. C., and G. A. Yarranton. 1978. The late-quaternary history of the boreal forest of central Canada, based on standard pollen stratigraphy and principal components analysis. *Journal of Ecology* **66**:199–212.
- Simpson, E. H. 1949. Measurement of diversity. *Nature* **163**:688.
- Smith, W., and J. F. Grassle. 1977. Sampling properties of a family of diversity measures. *Biometrics* **33**:283–292.
- Van Dam, H., G. Suurmond, and C. J. F. ter Braak. 1981. Impact of acidification on diatoms and chemistry of Dutch moorland pools. *Hydrobiologia* **83**:425–459.
- Whittaker, R. H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs* **30**:279–338.
- . 1972. Evolution and measurement of species diversity. *Taxon* **21**:213–251.
- . 1973. Direct gradient analysis: results. Pages 33–51 in R. H. Whittaker, editor. *Ordination and classification of communities*. Volume 5. *Handbook of vegetation science*. W. Junk, The Hague, The Netherlands.