Vegetatio 71: 41–47, 1987

A comparison of reciprocal averaging and non-centred principal components analysis

Exequiel Ezcurra
Instituto de Ecología, Apartado 18-845, 11800 México, D.F., Mexico

Accepted 9.2.1987

Key words: Biplot, Ordination, Rare species: effects

Abstract

Non-centred Principal Components Analysis (NPCA) ordinates sites and species simultaneously, and can be solved either by direct iteration or by eigenvector calculation. The weight of sites and species in the analysis is proportional to their overall abundance. Because of this, the method is not susceptible to distortion by rare species, as is the case with Reciprocal Averaging (RA). Detrending techniques can also be applied to this method to eliminate arch effects.

When NPCA was tried with field data, it produced ordination axes that were significantly associated to independently measured environmental variables. In contrast, RA failed to produce axes related to environmental factors, even after the main rare species had been eliminated from the analysis.

Abbreviations: NPCA Non-centred Principal Components Analysis; RA Reciprocal Averaging

Introduction

It is usually accepted that weighted averages ordination methods, including Reciprocal Averaging (RA; Hill, 1973, 1974), and the derived Detrended Correspondence Analysis (DCA; Hill & Gauch, 1980), have a "considerable superiority" over Principal Component Analysis (PCA) when the analysis of floristic data is considered (Gauch, 1982). The advantages seem to lie in the fact that (a) RA and DCA can be solved by direct iteration, and (b) environmental gradients with nonlinear species responses are recovered better by weighted averages ordination than by PCA (Gauch, 1982). In this paper it is argued that non-centred, non-standardised principal component analysis (NPCA; Noy-Meir, 1973) can often give better floristic ordinations than weighted averages methods. It is shown that direct iteration and detrending techniques can also be applied to NPCA.

Methods

NPCA: Two-way matrix scoring as an ordination procedure

Probably the simplest eigenvector ordination of a floristic data matrix can be achieved with a straightforward two-way scoring system. If an initial vector of site scores ordinating the sites along any real or assumed axis is given, the corresponding species scores can be calculated as the weighted sum of the scores of the sites where the species is found, such that

$$\varphi x_i = \sum_{j}^{n} a_{ij}\, y_j \qquad (1)$$

where x_i is the score of species i, y_j is the score of site j, a_{ij} is the abundance value of species i in site j, and φ is a scale factor. Similarly, stand scores can be calculated from an initial vector of species scores, such that

$$\varphi\, y_j = \sum_{i}^{s} a_{ij}\, x_i \qquad (2)$$

In terms of matrix algebra, equations (1) and (2) can be written as

$$\varphi\, \mathbf{x} = \mathbf{A}\, \mathbf{y} \qquad (3)$$

$$\varphi\, \mathbf{y} = \mathbf{A}'\mathbf{x} \qquad (4)$$

Replacing **x** and **y** in equations (3) and (4), we get

$$\lambda\, \mathbf{x} = \mathbf{A}\, \mathbf{A}'\mathbf{x} \text{ and} \qquad (5)$$

$$\lambda\, \mathbf{y} = \mathbf{A}'\mathbf{A}\, \mathbf{y} \qquad (6)$$

where $\lambda = \varphi^2$.

As can be seen, **x** is an eigenvector of the species dispersion matrix **A A'**, while **y** is an eigenvector of the sites dispersion matrix **A'A**. Clearly, λ is the eigenvalue for both solutions. Hence, any set λ, **x** and **y** that satisfies equations (5) and (6) is a valid solution to the floristic ordination problem. Although a direct iteration calculation as explained above will eventually converge to the largest eigensolution (Hotelling, 1936), it is not necessary in practice to follow this procedure. Given a floristic matrix **A**, the smallest dispersion matrix can be calculated, and the associated eigenvectors found. As demonstrated by Gower (1966) for the general dispersion-matrix problem, if there are more species than sites ($s > n$), it is simpler to solve eq. (6) for **y** (site scores). If there are more sites than species in the data set ($n > s$), then solve eq. (5) for **x** (species scores). Once any vector (**x** or **y**) is known, the other can be calculated from eqs. (3) and (4). The dispersion matrices (**A A'** and **A'A**) measure the dispersion of the data cluster around the origin of the data set. From equations (3) and (4) we can define

$$\text{Disp}(\mathbf{y}) = (\mathbf{A}'\mathbf{x})'\,(\mathbf{A}'\mathbf{x}) = \mathbf{x}'\mathbf{A}\,\mathbf{A}'\mathbf{x}$$

or, alternatively

$$\text{Disp}(\mathbf{x}) = (\mathbf{A}\,\mathbf{y})'\,(\mathbf{A}\,\mathbf{y}) = \mathbf{y}'\mathbf{A}'\mathbf{A}\,\mathbf{y}$$

From equations (5) and (6) it is now clear that

$$\text{Disp}(\mathbf{x}) = \text{Disp}(\mathbf{y}) = \lambda$$

That is, any eigenvalue of the dispersion matrix is a measure of the dispersion in the data explained by that particular axis. Let us define two matrices **X** and **Y** that contain, respectively, all the r possible **x** and **y** eigenvectors, where r is the rank of the data matrix **A**. By using Lagrange multipliers, it can be proved that this reciprocal scoring procedure produces orthonormal axes (i.e. $\mathbf{X}'\mathbf{X} = \mathbf{Y}'\mathbf{Y} = \mathbf{I}$), and that if all possible eigensolutions to the ordination problem are ranked from the largest to the smallest eigenvalue, each new axis maximises the residual dispersion in the data (Kendall, 1975; Morrison, 1976; Anderson, 1958).

If we define the diagonal matrix Λ of dimensions ($r \times r$) containing the r eigenvalues of **A** ($\lambda_1, \lambda_2, \lambda_3 \ldots \lambda_r$), then from (3) and (4) it is obvious that $\mathbf{X}\Lambda^{1/2} = \mathbf{A}\,\mathbf{Y}$, and hence $\mathbf{A} = \mathbf{X}\Lambda^{1/2}\mathbf{Y}'$, that is, $a_{ij} = \sum_{k}^{r} \sqrt{\lambda_k}\, x_{ik}\, y_{jk}$. This decomposition of the floristic table into a product of species scores, site scores and eigenvalues is known as a singular value decomposition (SVD; Good, 1969; Gabriel, 1978; Gabriel & Zamir, 1979). The interesting characteristic of SVD's is the fact that if only the first m eigensolutions are taken into consideration, then a least squares approximation of the floristic matrix results, i.e. $\| \mathbf{A} - \mathbf{X}_m \Lambda_m^{1/2} \mathbf{Y}_m' \|^2$ is minimum, or in simple algebraic terms, $\sum_{i}^{s} \sum_{j}^{n} (a_{ij} - \sum_{k}^{m} \sqrt{\lambda_k} x_{ik}\, y_{jk})^2$ is minimum (Good, 1969; Gabriel, 1971, 1978; Gabriel & Zamir, 1979). Thus, the procedure is really a criss-cross regression of columns of **A** onto **x** to obtain **y** as coefficients, and simultaneously of rows of **A** onto **y** to obtain **x** as coefficients. Each new **x**, **y** axis is a dyadic step in this least squares procedure (Gabriel & Zamir, 1979), decreasing the residual variance of the model.

This procedure is similar to Noy-Meir's (1973, see also Noy-Meir *et al.*, 1975) NPCA, with the only difference that species scores are presented and plotted in the same scale as site scores in the form of a biplot (Gabriel, 1971; Ter Braak, 1983). Note that the problem can be solved both by direct iteration or by calculating the eigensolutions of the dispersion matrices. To give a graphic idea of the dispersion removed by each axis, it is advisable to plot the scores (i.e. **x** and **y**) multiplied by the square root of the dispersion value (i.e. $\sqrt{\lambda}$).

The greatest virtue of this method is that it produces a corresponding and simultaneous ordination of both sites and species, as the analysis

makes no difference between loadings for attributes and positions for individuals, the scores for both rows and columns being defined in the same scale. The main difference from centred, or centred and standardised, PCA is that NPCA cannot be visualised exclusively as an axis rotation for the site cluster around its centroid in species space. It is a procedure that ordinates simultaneously the untransformed site cluster in species-space and the untransformed species cluster in site-space.

In this method the scores for the first axis will be all positive (or zero in some cases), as this axis extracts the dispersion of the data cluster with respect to the origin. Subsequent axes will extract the residual dispersion, i.e. approximately variation about the centroid. Noy Meir (1973) has demonstrated that, if there are disjunctions or near-disjunctions in the data set (i.e. if there are two or more distinct and separate submatrices), more than one asymmetric axis will appear. In this sense, NPCA helps to detect disjunct clusters in the data, which form distinct floristic sets with few intermediate or transitional sites.

In NPCA, the weight of each species is proportional to its total abundance. For a similar site distribution, species with higher abundances are given more extreme scores than species with lower abundances. This means that, implicitly, the method considers the more abundant species to be more reliable indicators of environment-induced floristic variation than the rarer species. The main advantage of this method lies in its lack of sensitivity to rare species.

Reciprocal averaging

RA can also be described as a reciprocal scoring system, but in this case the scores are averaged by the row (species or sites) totals. The main advantage of this procedure over NPCA is that the averaging (i.e. standardising by row totals) removes the effect of total abundances from the analysis. In RA the site score is an index of its position along the floristic gradient, independently of the total site abundance. Likewise, the species score can be regarded as an index reflecting the mean position of the species distribution along the site gradient, with independence of the overall species abundance or the amplitude of its distribution. In fact, RA calculates the species-scores by standardising the weighted sum of site-scores by the species total

$$\rho\, x_i = \sum_j^n a_{ij}\, y_j \,/\, \sum_j^n a_{ij}$$

and standardises site-scores by site-totals

$$\rho\, y_j = \sum_i^s a_{ij}\, x_i \,/\, \sum_i^s a_{ij}$$

If we define $a_{i.} = \sum_j^n a_{ij}$, and $a_{.j} = \sum_i^s a_{ij}$, then the standardisation matrices $\mathbf{R} = \mathrm{diag}(1/a_{i.})$ and $\mathbf{C} = \mathrm{diag}\,(1/a_{.j})$ can be defined, such that

$$\rho\, \mathbf{x} = \mathbf{R}\,\mathbf{A}\,\mathbf{y}, \text{ and} \tag{7}$$

$$\rho\, \mathbf{y} = \mathbf{C}\,\mathbf{A}'\mathbf{x}, \tag{8}$$

where ρ is, again, a scale factor. The complete eigenvector solution can be derived from these equations as

$$\lambda\, \mathbf{x} = \mathbf{R}\,\mathbf{A}\,\mathbf{C}\,\mathbf{A}'\mathbf{x} = \mathbf{S}\,\mathbf{x}, \text{ or as} \tag{9}$$

$$\lambda\, \mathbf{y} = \mathbf{C}\,\mathbf{A}'\mathbf{R}\,\mathbf{A}\,\mathbf{y} = \mathbf{Q}\,\mathbf{y}, \tag{10}$$

where $\mathbf{S} = \mathbf{R}\,\mathbf{A}\,\mathbf{C}\,\mathbf{A}'$ is the $(s \times s)$ species similarity matrix, and $\mathbf{Q} = \mathbf{C}\,\mathbf{A}'\mathbf{R}\,\mathbf{A}$ is the $(n \times n)$ site similarity matrix (note that, as with PCA, the problem can be solved by direct iteration or by calculating the eigensolutions of the similarity matrices). Both matrices are non-symmetric, i.e. the similarity of species i compared to species k is different to the similarity of species k in relation to species i.

$$s_{ik} = \left[\sum_{j_n}^{n} (a_{ij}\, a_{kj} \,/\, a_{.j})\right] \,/\, a_{i.} \tag{11}$$

$$s_{ki} = \left[\sum_{j}^{j_n} (a_{ij}\, a_{kj} \,/\, a_{.j})\right] \,/\, a_{k.} \tag{12}$$

If, say, species i is rare and species k is abundant (i.e. $a_{i.} << a_{k.}$) *then, obviously,* $s_{ik} >> s_{ki}$. That is, the rare species will show a high similarity with the frequent species, but the latter will have a low similarity with the former. Intuitively, this is an interesting and rarely discussed property of RA. It means that species restricted to only one site share

all of their distribution with widespread species, while widespread species share only a small fraction of their ecological variation with restricted species. Greenacre (1984) has shown that RA can also be interpreted as a singular value decomposition problem and, if the last axes are dropped, results in a least-squares approximation of the data matrix. The largest RA eigensolution is $\lambda_1 = 1$, both $\mathbf{x}_1$ and $\mathbf{y}_1$ being vectors of ones (Hill, 1973, 1974; Greenacre, 1984). This first RA axis is a trivial solution equivalent to the first asymmetric axis in NPCA, in this case all sites and species having equal values.

In RA, because every species has equal weights, the final position of a species on the ordination axes depends on the mean of its distribution, independently of its total abundance or of its distributional amplitude. This makes RA sensitive to the occurrence of rare species (i.e. species with low totals which usually occur in very few sites).

A test of the methods on field data

The data from a floristic matrix (100 species × 110 sites) obtained at the Pinacate Desert in NW Sonora, Mexico, was used as a mean to compare the efficiency of both methods (NPCA and RA) in generating floristics axes which could be statistically related to environmental data (for a detailed description of the area see Ezcurra *et al.,* 1987). The whole data matrix was subjected to both NPCA and RA. The relationship between the ordination axes and nine environmental variables (altitude, landform, soil unit, surface rockiness, rockiness class, bulk rockiness, electrical conductivity, % clay, % sand) was analysed through an additive step-wise linear regression procedure. Quadratic terms (allowing for non-linearity) were also tried in the regression function. Discrete variables (landform, soil unit, and rockiness class) were included in the analysis by decomposing them into binary dummy variables (Draper & Smith, 1981). The stepwise procedure was continued while the addition of new variables to the polynomial model decreased the probability (P) of the variance ratio (F) test. The complete matrix was analysed first, and a masked analysis (eliminating disjunct sites and rare species) was performed afterwards.

Complete analysis

NPCA. The reciprocal scoring analysis showed some sites with zero scores on the first axis, indicating disjunctions in the data set. Two asymmetric components (axes 3 and 6) were subsequently extracted. While axis 1 showed high scores for typical inland desert sites (and species), axis 3 separated riparian sites, and axis 6 showed high scores for coastal sites with halophytic species. In this desert, coastal halophytic vegetation, riparian sites and inland desert vegetation form disjunct floristic subsets with few transitional sites. The other, symmetric, axes (2, 4 and 5) revealed the variation within the main data cluster (inland desert), a problem which will be discussed in the next section. The regression model showed axis 1 to be highly associated ($P < 0.001$) with inland desert soils. Axis 3 was significantly associated ($P < 0.001$) with riparian landforms. Axis 6, the third asymmetric component, showed a strong association ($P < 0.001$) with saltflats.

Reciprocal Averaging. RA failed to detect the disjunctions in the data and concentrated the analysis on the distribution of "rare" species. Most axes separated only one or two sites (possessing the rare species) from the rest of the cluster. In some cases, rare species distinguished truly rare or a-typical sites; but in many other cases sites which were clearly similar to a larger cluster in soil type, landform and floristic composition were placed at the extreme of a gradient solely because of the presence of an unfrequent rare species in them. None of the axes showed a significant statistical association with the environmental variables under consideration.

Masked analysis

NPCA. A second ordination was performed, masking all sites and species that had zero or near-

zero scores along axis 1 of the NPCA ordination with the complete data (i.e. disjunct data sets and rare species). As expected, only one asymmetric component (axis 1) was extracted. This first axis goes approximately from the origin of the data space to the centroid of the cluster, and sites having higher total abundances project further away than sites with low abundances. In our case (presence-absence data), axis 1 reflected the species-richness of each site, the more diverse sites having higher scores. Axis 2, analysing the residual floristic variation, showed a significant linear correlation (r = 0.35, $P < 0.001$) with axis 1. The orthogonality condition does not necessarily mean lack of linear correlation between axes when an asymmetric component is considered: total species abundance in this desert is correlated with floristic variation (Ezcurra *et al.*, 1987).

Axis 2 (explaining 14.4% of the residual dispersion in the floristic data) was highly associated to both landform and altitude (r^2 = 0.55, $P < 0.001$). This axis extracted the main floristic variation in the area: from rocky communities at higher elevations to sandy, deep-soil communities at lower altitudes. Axis 3 (explaining 8.0% of the residual dispersion), also showed a significant relation with landform (r^2 = 0.38, $P < 0.001$). An interpretation of the results is given in Ezcurra *et al.* (1987). The main topographic sequence in this desert was recovered by axes 2 and 3, forming an arched sample cluster. The same configuration could be seen on the corresponding species ordination.

Reciprocal Averaging. RA also produced axes which related statistically to environmental data, but, in spite of the masking procedure, the rare-species effect still had a great impact on the analysis. Axis 1 (explaining 7.6% of the dispersion in the data cluster) showed a significant linear relation with landform (r^2 = 0.59, $P < 0.001$), but this relation was mostly due to the extreme scores of three dune sites. Axis 2 (explaining 6.3% of the dispersion) presented also a high association with landform (r^2 = 0.34, $P < 0.001$), but it was found again that this relation was due to extreme scores of three river-bank sites which had not been masked from the analysis as they also supported some non-riparian species. These first two RA axes detected true floristic variation, but failed to analyse the main floristic gradient and concentrated on a few atypical sites. Subsequent axes continued to separate one or a few sites, on the basis of the presence of one or a few species of restricted distribution. None of these axes showed a significant relation to environmental variables.

Discussion

It has been demonstrated (Gauch *et al.*, 1977; Del Moral, 1980) that given a strongly non-linear gradient, RA is better than PCA at recovering the original environmental gradient. The main reason for this lies in the particular form of standardisation used in the weighted averages scoring system of RA, which is capable of unbending the horseshoe configuration of non-linear data clusters in both stand-space and species-space. However, the arguments supporting this conclusion have been based either on theoretical reasoning (Hill, 1973, 1974) or on simulated data for species with Gaussian response curves (Gauch *et al.*, 1977; Del Moral, 1980). In practice, though, simple non-standardised PCA often gives more reliable descriptions of the floristic variation in the data, as RA is frequently distorted by the rare-species effect (e.g. Oksanen, 1983; Ezcurra, 1984).

Because of its particular form of score standardisation, RA gives the same weight to all species. Hence, rare species are treated as being extremely distinctive. RA places such species (and the sites containing them) at the extreme end of ordination axes. Gauch (1982) suggested that this problem can be easily overcome by deleting rare species from the data matrix. In practice, deciding when a species can be considered rare is a difficult and highly subjective task. Some species might occur in just a few sites because their particular habitat is poorly represented in the data matrix. Notwithstanding, they represent a true and well-defined floristic gradient, and should not be disregarded from the analysis. Other species may appear in only a few sites because their density in the area is low for a given stand size. Finally, some weedy, or "oportunistic",

species may show a more or less random occurrence in the sample, being present in only a few sites that do not present a distinct environmental characteristic. It would clearly be desirable to eliminate the last two categories from a floristic data matrix, as they add little or no information to the ordination problem, but there is no easy way to separate a priori the three groups and many intermediate cases may be found. To eliminate "uninformative" species from the analysis, a good knowledge of the environmental preferences of each species is needed. But one of the main objectives of a vegetation study is, precisely, to discriminate between species with a strong association with environmental factors from species with no clear environmental preferences. A priori deletion of all species occurring in, say, less than five sites might obscure important information.

It has been clear since the publication of Noy-Meir's (1973) and Noy-Meir *et al.*'s (1975) work, that any data transformation in PCA will change the configuration of the sample cluster in the alternative space (e.g. standardisation by species to get a less bended configuration of the site cluster in species-space will produce unexpected distortions of the species cluster in site-space). RA overcomes this problem in a mathematically elegant way, by using site standardisation when site scores are calculated, and species standardisation when species scores are calculated. Geometrically, this means that data transformations in one space are not reflected as undesirable changes on the other, alternative, space (note that the standardisation in RA is not a double standardisation but rather an alternating system of simple one-way standardisations). This type of standardisation means that all species are given equal weights in the analysis. In theory, this is a desirable property: species occupying the same position along the gradient will be given similar scores independently of their overall abundance; sites with a similar species composition will present similar scores even if they differ in their absolute total abundances.

But RA's main virtue can also be its main defect. Rare species with no clear environmental preference will be considered as indicators of large differences between sites, as they are given the same importance as the more abundant species with well-defined preferences. Geometrically, standardising by row totals is equivalent to a radial projection of the data points onto a unit hypotenuse hyperplane (Noy-Meir *et al.*, 1975; Greig-Smith, 1983). Obviously, species present in only one site will be projected on the edge of this hyperplane, and will appear as occupying an extreme of the floristic gradient. The same, of course, can happen if rare sites with only one species are incorporated to the analysis. This can be understood in more mathematical terms by comparing the asymmetric similarity measurements (eqs. 11 and 12) and their associated eigenvector problem (eqs. 9 and 10). The standardisation by row totals will give any rare species an extremely high similarity with the other species with which it shares its distribution. Therefore, the scores of rare species will amplify the scores of the other species in that particular site. If an eigensolution λ, $\mathbf{x}$, $\mathbf{y}$ is thought of as the product of the iterative calculation described in eqs. (7) and (8), it becomes clear that rare species (or depauperated sites) will present very high scores and can distort the analysis as they are placed at the extreme ends of a sample gradient.

The use of NPCA does not exclude the use of RA. Rather, they both are complementary descriptive techniques. RA is a good and reliable method if most species are related to some main environmental gradient, and if it is wanted to give the same weight to every species. If, alternatively, it is thought that the more frequent species should have more weight in the analysis, NPCA is the adequate tool for exploring variation in the data. Detrending techniques to remove the arch effect can, of course, be applied to NPCA in the same way as they are applied to correspondence analysis, by using the iterative calculation approach. In our experience with field data, however, the rare-species effect on ordination analyses is frequently a bigger problem than the nonlinear relationship between ordination axes and environmental factors.

Acknowledgement

This research was supported by the Consejo Na-

cional de Ciencia y Tecnologia (CONACYT), Mexico, the Ford Foundation and the Regional Program for Technological and Scientific Development of the Organization of American States (OAS).

This paper was rewritten from the author's Ph.D. thesis, which was presented at the School of Plant Biology, Univ. Coll. of North Wales, Bangor. I thank Prof. P. Greig-Smith for his direction and guidance throughout this project and Mike Austin for critically reading the manuscript. The comments and criticism from two anonymous referees are greatly appreciated.

References

Anderson, T. W., 1958. An introduction to multivariate statistical analysis. J. Wiley & Sons, New York.

Del Moral, R., 1980. On selecting indirect ordination methods. Vegetatio 42: 75–84.

Draper, N. & Smith, H., 1981. Applied regression analysis. 2nd ed. John Wiley & Sons, New York.

Ezcurra, E., 1984. The vegetation of El Pinacate, Sonora. A quantitative study. Unpublished Ph.D. Thesis, Univ. Coll. of North Wales, Bangor, U.K.

Ezcurra, E., Equihua, M. & López-Portillo, J., 1987. The desert vegetation of El Pinacate, Sonara, Mexico. Vegetatio 71: 49–60.

Gabriel, K. R., 1971. The biplot display of matrices with application to principal component analysis. Biometrika 58: 453–467.

Gabriel, K. R., 1978. Least squares approximations of matrices by additive and multiplicative models. J. R. Statist. Soc. B 40: 186–196.

Gabriel, K. R. & Zamir, S., 1979. Lower rank approximation of matrices by least squares with any choice of weights. Technometrics 21: 489–498.

Gauch, H. G., 1982. Multivariate analysis in community ecology. Cambridge Univ. Press, Cambridge.

Gauch, H. G., Whittaker, R. H. & Wentworth, T. R., 1977. A comparative study of reciprocal averaging and other ordination techniques. J. Ecol. 65: 157–174.

Good, I. J., 1969. Some applications of the singular decomposition of a matrix. Technometrics 11: 823–831.

Gower, J. C., 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53: 323–338.

Greenacre, M. J., 1984. Theory and applications of correspondence analysis. Academic Press, London.

Greig-Smith, P., 1983. Quantitative plant ecology (3rd ed.). Blackwell Scientific Publications, Oxford.

Hill, M. O., 1973. Reciprocal averaging: an eigenvector method of ordination. J. Ecol. 61: 237–249.

Hill, M. O., 1974. Correspondence analysis: a neglected multivariate method. Appl. Stat. 23: 340–354.

Hill, M. O. & Gauch, H. G., 1980. Detrended correspondence analysis: An improved ordination technique. Vegetatio 42: 47–58.

Hotelling, H., 1936. Simplified calculation of principal components. Psychometrika 1: 27–35.

Kendall, M. G., 1975. Multivariate analysis. C. Griffin & Co., London.

Morrison, D. F., 1976. Multivariate statistical methods. McGraw-Hill Kogakusha, Tokyo.

Noy-Meir, I., 1973. Data transformation in ecological ordination: I. Some advantages of non-centering. J. Ecol. 61: 329–341.

Noy-Meir, I., Walker, D. & Williams, W. T., 1975. Data transformation in ecological ordination: II. On the meaning of data standardization. J. Ecol. 63: 779–800.

Oksanen, J., 1983. Ordination of boreal heath-like vegetation with principal component analysis, correspondence analysis and multidimensional scaling. Vegetatio 52: 181–189.

Ter Braak, C. J. F., 1983. Principal components biplots and alpha and beta diversity. Ecology 64: 454–462.