PBDB Intensive Summer Course 2007
Paleoecology Section – Thomas Olszewski

Parametric Ordination

Introduction

Unlike nonparametric ordinations, which are based on geometric search algorithms, parametric ordination techniques are based on least-squares fitting closely related to bivariate regression and correlation

The formal mathematics takes advantage of eigen decomposition to create an orthogonal frame of reference that maximizes variance along the principal axes.

Principal Components Analysis

Object: Find linearly independent combinations of variables (referred to as principal components). Assuming that samples are not random mixes of taxa – i.e., taxa show distinct and consistent associations related to environment, preservation, age, etc. – much of the information in a complex data set can be reduced (i.e., projected) to a small number of interpretable axes.

Although not regarded as appropriate for most ecological community data because it does not handle sparse matrices very well, PCA is the foundation of virtually all other parametric techniques. In addition, it is appropriate for studies like morphometrics.

Step #1: Center the data matrix. Calculate the mean of each column in the data matrix ($X_{N\times p}$ = data matrix; N = number of rows/samples, p = number of columns/taxa): $\bar{x}_j = \sum_{i=1}^{N} \frac{x_{ij}}{N}$. For each element of X, subtract the mean of its column:

$$Y_{N\times p} = X - \frac{\vec{1}'\cdot X}{\vec{1}'\cdot \vec{1}} \qquad (1).$$

Y is the centered data matrix; each of its elements equals: $y_{ij} = x_{ij} - \bar{x}_j$.

Geometric Interpretation. Each sample (i.e., row in X) describes a point in a multivariate space whose Cartesian axes (x, y, z, etc.) represent the abundances of taxa (i.e., taxon counts of each sample are its x-y-z coordinates). A point defined by the mean of each taxon (i.e., the mean of each column in X) identifies the centroid of the cloud of sample points. Subtracting the centroid from all the columns moves the cloud of points so that the centroid is at the origin.

Step #2 (optional): Scale the column vectors of the centered matrix using the standard deviation of each column. Calculate the variance of each column: $S_j = \sum_{i=1}^{N} \frac{(x_{ij} - \bar{x}_j)^2}{N}$; $S^{1/2}$ is the standard deviation of each column (i.e., the square root of the variance). Divide each value in Y by the standard deviation of its column:

$$Z_{N\times p} = Y \cdot diag(S^{-1/2}) \qquad (2).$$

This step results in a z-transform of each of the columns in X: $z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}$.

Geometric Interpretation. Recall that the square root of the sum of squared differences is a Euclidean distance – i.e., the standard deviation of each column is the length of that column vector. Because different taxa have different scales of abundance (some may be rare, some may be very abundant), the cloud of points in *Y* may be stretched along some axes and compressed along others. Dividing the values of all the sample points along each axis (i.e., taxon) by the standard deviation of the points along that axis means that each axis is scaled so that it is in units of standard deviations and the origin coincides with the centroid of the cloud of sample points.

Step #3: Calculate the correlation matrix (if step #2 was not applied, this calculates a covariance matrix). Take the inner product moment of *Z* to find *R*, the matrix of correlation coefficients between every pair of columns in *X*:

$$R_{p\times p} = \frac{Z'\cdot Z}{N} \qquad (3).$$

The value of each element of *R* is $r_{jk} = \sum_{i=1}^{N} \frac{(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\Sigma(x_{ij} - \bar{x}_j)^2}\sqrt{\Sigma(x_{ik} - \bar{x}_k)^2}}$, the Pearson product moment correlation coefficient between columns *j* and *k*.

Geometric Interpretation. Correlation is closely related to regression. By centering and scaling the data matrix, any regression line through the cloud of sample points conforms to a *reduced major axis regression* (unlike a bivariate least-squares regression, which minimizes the sum of squared deviations in the y-direction, a reduced major axis regression minimizes the sum of squared deviations perpendicular to the regression line). Sample points in the ordination space are scattered around the regression line in an ellipsoidal shape (assuming they are not randomly scattered, in which case they would be spheroidally distributed). The correlation coefficient equals the square root of the complement of the ratio of the minor axis of the ellipsoid to its major axis: $r = \sqrt{1 - \frac{\min or}{major}}$ evaluated in directions parallel to the axes of the multivariate space (which, again, represent the taxa). In other words, the *R* matrix describes the shape of the cloud of sample points in the multivariate space by describing the degree to which the cloud is ellipsoidal in a series of oblique cross-sections parallel to planes defined by all the axis pairs.

Step #4: Find the multivariate regression lines. Due to redundancy among taxa, the space with axes representing taxa is not the best way to see the shape of the cloud of sample points. As an alternative way of viewing the cloud, rotate its longest ellipsoidal axis to the x-axis, its second longest to the y-axis, etc. To do this, it is necessary to find the orientation of the regression lines (i.e., ellipsoid axes) in the taxon-defined space. Do this by decomposing *R* into its eigenvectors (*U*) and eigenvalues (Λ):

$$R = U\Lambda U' \qquad (4).$$

Geometric Interpretation. U is the matrix of eigenvectors; these are orthonormal – i.e., they are mutually perpendicular and all of length one – and oriented in the direction of the ellipsoid's axes. Λ represents the length of the axes as measured in units of variance.

Step #5: Rotate the cloud of sample points so that the eigenvectors serve as the new frame of reference: To complete the PCA, we need to calculate the *factor loadings* (A) and *factor scores* (F):

$$A = U\Lambda^{1/2}$$
$$F = ZA\Lambda^{-1} = ZU\Lambda^{-1/2} \qquad (5)$$

Geometric Interpretation. $A = U\Lambda^{1/2}$ describes the ellipsoid's axes scaled to standard deviation (an orthonormal matrix post-multiplied by a diagonal matrix has its column vectors stretched by the scalar values in the diagonal matrix). A is the new frame of reference (i.e., the eigenvectors are the new x-y-z axes). $F = ZA\Lambda^{-1} = ZU\Lambda^{-1/2}$ projects the rows of Z (i.e., sample points) in the old frame of reference into the new frame of reference. A defines composite taxa (i.e., linear combinations of observed taxonomic abundances) and F describes the abundance of each of these composite taxa in each sample.

Note that PCA is a rotation of the data to a more optimal frame of reference. In fact, the original centered, scaled matrix can be easily recovered: $Z = FA'$, which shows that no information about species associations has been lost. This is also confirmed by the following relationship: sum(Λ) = sum(diag(R)), which indicates that the lengths of the axes equal the bounds of the ellipsoid in the original frame of reference.

Additional Comments.

1) If step #2 is not applied, covariances are used instead of correlations. Effectively, this means that the axes are not normalized and the orientation of the eigenvectors and the magnitude of the eigenvalues will reflect differences in the magnitude of the measurement scale of different taxa.
2) In effect, PCA uses r as a similarity metric. r is not an appropriate measure for sparse data sets. Application of eigen decomposition to similarity matrices based on other measures is given the general term *Principal Coordinates Analysis*. If the metric does not follow the triangular inequality, this can lead to imaginary eigenvalues.
3) PCA has an implicit assumption of linear change in taxonomic abundance along PC axes. This is not generally a good assumption for ecological community data, so various types of transformations (e.g., logarithmic) are necessary to scale the data appropriately. In addition, linearity also assumes continuous representation of species at all sites (i.e., non-sparse data), which is not consistent with many natural gradients (these oftern show unimodal, bell-shaped curves along gradients).

4) Number of non-zero eigenvalues = rank of R = rank of X – i.e., the minimum number of dimensions required to include all the information in the data matrix
5) Q-mode can be conducted by starting with X'; it will have identical eigenvalues with the R-mode

Factor Analysis

PCA reorganizes the information in a data matrix so that the largest proportion of variance is constrained on the first axis, the next largest amount that is independent of the first axis (i.e., perpendicular) is on the second, the third largest amount of variance is perpendicular to the plane defined by the first and second axes is on the third axis, etc. The first axis is the best one-dimensional least-squares summary of the data. The first plus second axis is the best two-dimensional least-squares summary of the data.

The total amount of information in the eigen decomposition is the sum of the eigenvalues; the amount of information captures by a subset of axes is the sum of the corresponding subset of eigenvalues.

Factor analysis is simply a focus on a subset of principal components regarded as a sufficient approximation of the full data matrix.

Communality = $a_{i1}^2 + a_{i2}^2 + ... + a_{im}^2$; m = number of factors included; a_{ij} = element of A, the factor loadings matrix; communality represents that part of a variable's variance related to factor i
Specificity = that part of a variable's variance that is not related to factor i

How many factors?

1) enough to explain a total amount of cumulative variance (sum of eigenvalues)
2) include all PC's larger than a certain eigenvalue
3) include all factors with high factor loadings (i.e., they incorporate most of the pattern for most of the taxa)
4) where a scree plot (eigenvalue versus factor number) flattens out
5) test for significance of residuals
6) take only factors with eigenvalue > 1 (i.e., more correlated than diagonal elements in R matrix) – NOT A SOUND CRITERION!
7) test eigenvalues for significance

Rotation of Axes

PCA axes are perpendicular and maximize variance – however, the PCs can be rotated to maximize other aspects of the pattern.

Methods: varimax, promax, equimax, or quartimax

Varimax: rigid rotation of subset of PCA axes; rotate first two axes until variances of a^2_{ij} are maximized, then go to next pair; repeat for each possible pair until total sum of a^2_{ij} is maximized – axes now point at clusters with a few taxa maximized on each factor and minimized on the rest

Correspondence Analysis

CA is a synonym of *Reciprocal Averaging*

This is perhaps the most heavily used ordination technique in ecology at the present time

Unlike PCA and related techniques, R and Q modes are depicted *in the same multivariate space*, allowing taxa to be directly related to sites

All elements of the data matrix must be > 0 (i.e., no negative values)

Step #1: Rescale elements of X to proportions of the whole. Divide every element of the data matrix ($X_{N\times p}$ = data matrix; N = number of rows/samples, p = number of columns/taxa) by the sum of all terms in the matrix:

$$P = \frac{X}{\vec{1}'\cdot X\cdot \vec{1}} \qquad (1);$$

the elements of P equal $p_{ij} = \frac{x_{ij}}{\sum_{i=1}^{N}\sum_{j=1}^{p} x_{ij}}$.

Such a matrix of proportions is called a contingency table – each element represents the probability that a given specimen comes from that taxon (column) in that sample (row).

Geometric Interpretation. This operation rescales the data matrix – i.e., relative positions of sample points in taxon space are not changed because all have been divided by the same scalar value. The rows of P indicate the positions of sample points in a Cartesian coordinate space; the position of each point is determined by the proportions of each taxon in each sample (i.e., the x-y-z values). Note that the distance of each sample point from the origin is determined by its size – larger samples plot further from the origin

Step #2: Calculate a matrix of <u>expected</u> values of P. The matrix of expected values is calculated based on the row and column totals of matrix P – these vectors represent the probability that a randomly chosen specimen will come from a given sample (row totals vector) or taxon (column totals vector):

$$\overline{P} = row\cdot column = (P\cdot\vec{1})\cdot(\vec{1}\cdot P) \qquad (2).$$

The row total vector represents the average sample in the data matrix; the column total vector represents the average species distribution for the entire data matrix.

Geometric Operation. In the orthogonal space where each axis corresponds to a taxon, the rows of $\overline{P}$ indicate points with exactly the taxonomic proportions as the entire data set – i.e., the sample centroid – but at distances from the origin corresponding to the size of each actual sample as recorded in the rows of P.

Step #3. Calculate the matrix of deviates. This is a matrix of the observed proportions minus the expected proportions:

$$B_{N\times p} = P - \overline{P} = \frac{X}{\vec{1}'\cdot X\cdot\vec{1}} - (P\cdot\vec{1})\cdot(\vec{1}\cdot P) \qquad (3);$$

the elements of *B* equal $b_{ij} = p_{ij} - p_{i+}p_{+j}$

p_{i+} = sum of all terms in row *i*
p_{+j} = sum of all terms in column *j*

Geometric Interpretation. *B* is a matrix of vectors pointing from the line of points defined by the rows of $\overline{P}$ (the scaled sample centroids) to the positions of the corresponding actual samples indicated in the rows of the *P* matrix.

Step #4. Divide the deviate matrix B by the square roots of the expected values. This procedure rescales the vectors in *B* so that the scale of all of the taxon axes is equal. The $\frac{p_{ij}}{\sqrt{p_{i+}p_{+j}}}$ term corresponds to the rescaled points of matrix P, and the $\sqrt{p_{i+}p_{+j}}$ term corresponds to the rescaled points of $\overline{P}$:

$$Q_{N\times p} = diag(P\cdot\vec{1})^{-1/2}\cdot B\cdot diag(\vec{1}'\cdot P)^{-1/2} \qquad (4);$$

the elements of *Q* equal $q_{ij} = \frac{p_{ij} - p_{i+}p_{+j}}{\sqrt{p_{i+}p_{+j}}} = \frac{p_{ij}}{\sqrt{p_{i+}p_{+j}}} - \sqrt{p_{i+}p_{+j}}$.

Geometric Interpretation. Compositional data can be summarized on a *simplex*. A familiar example of a simplex is a triangular ternary diagram on which every sample can be plotted based on the proportions of three components (e.g., minerals in a rock). With four components, a compositional simplex becomes a tetrahedron and occupies three dimensions. The same reasoning can be extrapolated into any number of dimensions. The compositional simplex concept can be tied to the Cartesian space of parametric ordination in the following way: the apices of the simplex each intersect an orthogonal axis. In other words, a sample with 100% of one taxon should be at the simplex apex that corresponds to that taxon; in Cartesian space, this point should fall exactly on the axis corresponding to that taxon. Note that a set of samples with different total abundances will fall at different distances along the taxon axis – this indicates that samples of different size result in simplexes of different size.

Each sample in the *P* matrix falls on a simplex plane (although at different distances from the origin because samples are of different size and the only way to accommodate a larger simplex is to move farther from the origin). $\overline{P}$ contains the position on each simplex of the sample centroid adjusted for the size of each real sample. Note that the centroid does NOT correspond to the center of the simplex (i.e., a point at which each taxon is of equal proportion). Because different taxa are of different abundance, the *P* simplex is distorted and tilted (i.e., it is not an equilateral triangle in which each taxon is given equal importance). In

Q, the $\frac{p_{ij}}{\sqrt{p_{i+}p_{+j}}}$ term shifts the sample points in *P* and the $\sqrt{p_{i+}p_{+j}}$ term shifts the points in $\overline{P}$ so that they lie in a space in which the orthogonal axes representing taxa are all the same scale. *Q* itself then contains the rescaled vectors of *B.*

Step #5. Take the inner product moment of the scaled, centered, weighted matrix. This step is analogous to finding the *R* matrix in PCA, except that instead of using the correlation coefficient *r*, the *C* matrix if filled with values of the so-called χ^2 coefficient. (Note how the form of the elements in Q resemble the square root of the metric used to test the significance of contingency tables: $\chi^2 = \Sigma(\text{Observed} - \text{Expected})^2/\text{Expected}$ – hence the name.) As with all forms of ordination, *C* is a form of similarity matrix:

$$C_{p\times p} = Q'\cdot Q \qquad (5)$$

the elements of *C* are $c_{jk} = \sum_{i=1}^{N} \frac{(p_{ij} - p_{i+}p_{+j})(p_{ik} - p_{i+}p_{+k})}{\sqrt{p_{i+}p_{+j}}\sqrt{p_{i+}p_{+k}}} = \sum_{i=1}^{N} \frac{(p_{ij} - p_{i+}p_{+j})(p_{ik} - p_{i+}p_{+k})}{p_{i+}\sqrt{p_{+j}}\sqrt{p_{+k}}}$.

Geometric Interpretation. C records the pairwise dot product between each of the columns in *Q* (i.e., the taxa). Remember that the dot product of two vectors is the product of their lengths times the cosine of the angle between them: $\vec{x}'\cdot\vec{y} = |\vec{x}||\vec{y}|\cos\theta$. The diagonal elements of *C* record the sum of squared sample distances along each orthogonal axis in the space defined by *Q*. The off-diagonal elements express the degree of similarity in the distribution of sites along axes *j* and *k.*

Step #6. Define an orthonormal frame of reference onto which to project the simplex. Do this by eigen decomposing the matrix of similarities:

$$C = U\Lambda U' \qquad (6).$$

Geometric Interpretation. U is the matrix of eigenvectors; these are oriented in directions that maximize the variance of species along the length of each of the eigenvectors (i.e., the sum of squared sample positions along the eigenvector) subject to the constraint of orthonormality. One eigenvector (the one corresponding to an eigenvalue of 0) will follow the trend of scaled centroid points (the $\sqrt{p_{i+}p_{+j}}$ term in the equation for each element of *Q*). Λ represents the length of the axes as measured in units of variance. In CA, the eigenvalues are referred to as *interias*; total intertia (i.e., $tr\Lambda$) equals $\bar{1}'\cdot Q^2\cdot\bar{1} = trC$ (note that this is the actual χ^2 value for matrix *P*, which is not equal to the so-called χ^2 similarity metric). Inertias describe the portion of the total variance in the data matrix accounted for by each CA axis.

Step #7. Define the new frame of reference using the eigenvectors (combinations of taxa) and project samples into this new frame of reference. The orthonormal frame of reference defined by the eigenvectors can be used to project both samples and taxa can be projected into the same space:

$$\begin{aligned} A &= U\Lambda^{1/2} \\ A^* &= QU \end{aligned} \qquad (7),$$

A = taxon scores
A^* = sample scores.

Geometric Interpretation. A and A^* are equivalent to the factor loadings and factor scores of PCA, respectively. A^* contains the projections of samples in the original orthonormal reference frame onto the new reference frame defined by the eigenvectors. A contains the contribution of each taxon to each eigenvector (as in PCA, these can be considered composite taxa).

Step #8. Stretch the sample and taxon vectors to account for difference in total proportion. The final step of CA rescales the positions of the samples and taxa on the eigenvectors to account for differences in sample size and taxon abundance:

$$\begin{aligned} A_a &= diag(\vec{1}' \cdot P)^{-1/2} A \\ A_a^* &= diag(P \cdot \vec{1})^{-1/2} A^* \end{aligned} \qquad (8).$$

Geometric Interpretation. The values given by A and A^* use Cartesian coordinates to describe the positions of taxa and samples, respectively. This means that the simplexes of each sample have not been projected onto each other at the same scale, so samples of identical composition but different size appear at different locations. The procedure used to obtain A_a and A^*_a rescales the simplexes so that they reflect strictly taxon proportions – i.e., samples of different size but with the same taxonomic proportions plot at the same locations. This also shifts taxon positions so that they reflect the weighted mean of the samples in which they occur (weighted by their proportion in each sample).

Presenting CA results

1) include inertia percentages on plots or in captions
2) can use CA first axis to reorder data matrices – result is similar to a 2-way cluster analysis and related to a method of matrix ordering called TWINSPAN
3) triangular shapes when crossplotting axes are an expression of the simplex geometry of CA – if samples preferentially occupy only the edges of the simplex this can result in "bending" of a sample gradient; this is what has been referred to an "arch" or "horseshoe"; it is endemic to all ordination techniques (see discussion of detrended CA below)

Reciprocal Averaging

Correspondence analysis is a true parametric ordination method, but it was originally described by ecologists as an algorithm called *reciprocal averaging*.

Recipe for reciprocal averaging:

1) Starting with *X*, find the row totals and column totals (i.e., marginal vectors)
2) Assign each sample an arbitrary ordination score (the closer these are to the final scores, the faster RA will converge)
3) Calculate the species scores by finding the weighted averages (i.e., centroids) of the sites they belong to
4) Use these newly calculated species scores to back-calculate new site scores in an analogous manner
5) The back-calculated site scores will result in a decreased range of values, so rescale them to the original range (so they have the same minimum and maximum values as you started with)
6) Repeat until scores stabilize
7) To derive higher axes, it is necessary to remove information already represented by the previously calculated axes – do this by subtracting a multiple of the lower axis scores from the new axis scores:

$$v_a = \sum_{i=1}^{n} \frac{x_{+i} s_i f_{ai}}{x_{++}}$$

$$s_{i,new} = s_{i,old} - \sum_{a=1}^{i} v_a f_{ai}$$

8) To achieve the proper CA scaling (which is not necessary in traditional RA), calculate the centroid of the site scores on each axis ($\bar{x}_a = \frac{\sum n_i x_i}{\sum n_i}$), the variances of the site scores on each axis ($v_a = \frac{\sum n_i (x_i - \bar{x}_a)^2}{\sum n_i}$) and use these to z-transform the site scores: $x_{i,new} = \frac{x_{i,old} - \bar{x}_a}{\sqrt{v_a}}$.

Detrended Correspondence Analysis

Long, linear gradients can show so much compositional turnover that the ends are no different from each other than either is from the middle – this results in the gradient being "bent" when projected into an ordination space, regardless of the particular method of ordination used.

PCA and PCOORD can result in gradients that are entirely folded over on themselves.

In NMDS, "flexible shortest path adjustment" and "global NMDS", which are methods for focusing on similarity relationships among samples that share species rather than those that do not, can be used to address the issue of truncated maximum dissimilarity (see references for NMDS).

In CA, "detrending" is used to straighten the arch and decompress the points at the ends of the arch so that the gradient is displayed linearly

1) *Detrending by segments* (an algorithmic solution to the perceived "arch" problem):
    a. Break ordinated gradient relating axes 1 and 2 into pieces along axis 1
    b. Find the local average of each segment and move it to zero
    c. Find a regression line in each segment and use it to straighten the segment (i.e., plot the residuals of the points)
    d. Do this repeatedly with a sliding window along axis 1 until point positions stabilize
    e. Calculate opposite mode (R or Q) using weighted averaging
2) *Detrending by polynomials*
    a. Fit a polynomial (i.e., a parabola) to the arch
    b. Find the residuals of each point perpendicular to the arch
    c. Find the position of each point along the length of the arch
    d. Plot the points using their position and residual

NOTE: Detrending is very useful for extracting linear gradients, but it requires external knowledge of the data to justify the interpretation that gradients are in fact linear and continuous. *Do not apply detrended CA without first examining CA results and being sure that detrending is appropriate!* When used inappropriately, detrending can destroy the ordinated depiction of complex multidimensional gradients (i.e., coenoplanes) and clustering among samples or taxa.

References for Parametric Ordination
(including Canonical Methods and Additional General References)

Albrecht, G.H., 1980, Multivariate analysis and the study of form, with special reference to canonical variate analysis. American Zoologist, v. 20, p. 679-693.

Austin, M.P. and Noy-Meir, I., 1971, The problem of non-linearity in ordination: experiments with two-gradient models. Journal of Ecology, v. 59, p. 763-773.

Benzécri, J.-P. 1992, Correspondence Analysis Handbook. Marcel Dekker, New York 665 p. (Benzécri and his group developed correspondence analysis in its modern form during the 1970's but almost all the literature was in French until they published this volume.)

Campbell, N.A. and Atchley, W.R., 1981, The geometry of canonical variate analysis. Systematic Zoology, v. 30, p. 268-280.

Clarke, K.R. and Warwick, R.M., 2001, Change in Marine Environments: An Approach to Statistical Analysis and Interpretation, 2nd ed. Primer-E Ltd., Plymouth, UK.

Gauch, H.G., Whittaker, R.H., and Wentworth, 1977, A comparative study of reciprocal averaging and other ordination techniques. Journal of Ecology, v. 65, p. 157-174.

Gittins, R., 1985, Canonical Analysis: A Review with Applications in Ecology. Springer-Verlag, New York, 351 p.

Greenacre, M.J. and Hastie, T., 1987, The geometric interpretation of correspondence analysis. Journal of the American Statistical Association, v. 82, p. 437-447. (The initial section on the geometry of simple CA presents some surprising expectations for the interpretation of ordination plots derived using this method.)

Greenacre, M.J., 1984, Theory and Application of Correspondence Analysis. Academic Press, New York, 364 p. (A very useful and thorough manual if you are really into CA.)

Hill, M.O. and Gauch, H.G., Jr., 1980, Detrended correspondence analysis: an improved ordination technique. Vegetatio, v. 42, p. 47-58. (A development of RA that **supposedly** takes care of certain problems in the method.)

Hill, M.O., 1973, Reciprocal averaging: an eigenvector method of ordination. Journal of Ecology, v. 61, p. 237-251. (How to do RA by hand – a clear explanation and set of instructions.)

Hill, M.O., 1974, Correspondence analysis: a neglected multivariate method. Applied Statistics, v. 23, p. 340-354.

Jackson, D.A., 1993, Multivariate analysis of benthic invertebrate communities: the implication of choosing particular data standardizations, measures of association, and ordination methods. Hydrobiologia, v. 268, p.9-26.

Jongman, R.H.G., ter Braak, C.J.F., and van Tongeren, O.F.R., 1995, Data Analysis in Community and Landscape Ecology. Cambridge University Press.

Jöreskog, K.G., Klovan, J.E., and Reyment, R.A., 1976, Geological Factor Analysis. Elsevier, Amsterdam.

Kenkel, N.C. and Orloci, L., 1986, Applying metric and non-metric multidimensional scaling to ecological studies: some new results. Ecology, v. 67, p. 919-928.

Klovan, J.E., and Billings, G.K., 1967, Classification of geological samples by discriminant-function analysis. Bulletin of Canadian Petroleum Geology, v. 15, p. 313-330.

Legendre, P. and Legendre, L., 1998, Numerical Ecology, 2nd English Edition. Developments in Environmental Modelling 20, Elsevier.

Makarenkov, V. and Legendre, P., 2002, Nonlinear redundancy analysis and canonical correspondence analysis based on polynomial regression. Ecology, V. 83, p. 1146-1161.

McCune, B. and Grace, J.B., 2002, Analysis of Ecological Communities. MjM Software Design.

Minchin, P.R., 1987, An evaluation of the relative robustness of techniques for ecological ordination. Vegetatio, v. 69, p. 89-107.

Palmer, M.W., 1993, Putting things in even better order: The advantages of canonical correspondence analysis. Ecology, v. 74, p. 2215-2230.

Palmer, M.W., 1993, Putting things in even better order: The advantages of canonical correspondence analysis. Ecology, v. 74, p. 2215-2230.

Peet, R.K., Knox, R.G., Case, J.S., and Allen, R.B., 1988, Putting things in order: The advantages of detrended correspondence analysis. The American Naturalist, v. 137, p. 704-712.

Podani, J. and Miklós, I., 2002, Resemblance coefficients and the horseshoe effect in principal coordinates analysis. Ecology, v. 83, p. 3331-3343.

Rempe, U. and Webber, E.E., 1972, An illustration of the principal ideas of MANOVA. Biometrics, v. 28, p. 235-38.

Ter Braak, C.J.F. and Prentice, I.C., 1988, A theory of gradient analysis. Advances in Ecological Research, v. 18, p. 271-317.

Ter Braak, C.J.F. and Schaffers, A.P., 2004, Co-correspondence analysis: A new ordination method to relate two community compositions. Ecology, v. 85, p. 834-846.

Ter Braak, C.J.F. and Verdonschot, P.F.M., 1995, Canonical correspondence analysis and related multivariate methods in aquatic ecology. Aquatic Sciences, v. 57, p. 255-289.

Ter Braak, C.J.F., 1985, Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. Biometrics, v. 41, p. 859-873. (Describes some important properties of CA relative to weighted averaging of ecological gradient data.)

Ter Braak, C.J.F., 1986, Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. Ecology, v. 67, p. 1167-1179.

Ter Braak, C.J.F., 1987, The analysis of vegetation-environment relationships by canonical correspondence analysis. Vegetation, v. 69, p. 69-77.

Wagner, H.H., 2004, Direct multi-scale ordination with canonical correspondence analysis. Ecology, v. 85, p. 342-351.

Wartenberg, D., Ferson, S., and Rohlf, F.J., 1987, Putting things in order: A critique of detrended correspondence analysis. The American Naturalist, v. 129, p. 434-448.

Whittaker, R.H. and Gauch, H.G., Jr., 1973, Chap. 11 Evaluation of ordination techniques. *In* Whittaker, R.H., ed., Ordination and Classification of Communities, Part V of Handbook of Vegetation Science (Tüxen, R., ed.), p. 289-321.

APPENDIX: MATRIX ALGEBRA OF PARAMETRIC ORDINATIONS

PCA

$$X_{N\times p}$$

N = number of rows
p = number of columns

$$Y_{N\times p} = X - \frac{\vec{1}'\cdot X}{\vec{1}'\cdot\vec{1}}$$

$$y_{ij} = x_{ij} - \bar{x}_j$$

$$\bar{x}_j = \sum_{i=1}^{N}\frac{x_{ij}}{N}$$

$$Z_{N\times p} = Y\cdot diag(S^{-1/2})$$

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}$$

$$S^{1/2} = [\sigma_j];\sigma_j = \sum_{i=1}^{N}\sqrt{\frac{(x_{ij} - \bar{x}_j)^2}{N}}$$

$$R_{p\times p} = \frac{Z'\cdot Z}{N}$$

$$r_{jk} = \sum_{i=1}^{N}\frac{(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\Sigma(x_{ij} - \bar{x}_j)^2}\sqrt{\Sigma(x_{ik} - \bar{x}_k)^2}}$$

$$R = A\cdot A' = U\Lambda U'$$

$$A = U\Lambda^{1/2}$$

$$F = ZA\Lambda^{-1} = ZU\Lambda^{-1/2}$$

$$Z = FA'$$

$$sum(\Lambda) = sum(diag(R))$$

CA

$$X_{N\times p}$$

N = number of rows
p = number of columns

$$B_{N\times p} = P - \overline{P} = \frac{X}{\vec{1}'\cdot X\cdot\vec{1}} - (P\cdot\vec{1})\cdot(\vec{1}\cdot P)$$

$$b_{ij} = p_{ij} - p_{i+}p_{+j}$$

p_{i+} = sum of all terms in row i
p_{+j} = sum of all terms in column j

$$Q_{N\times p} = diag(P\cdot\vec{1})^{-1/2}\cdot B\cdot diag(\vec{1}'\cdot P)^{-1/2}$$

$$q_{ij} = \frac{p_{ij} - p_{i+}p_{+j}}{\sqrt{p_{i+}p_{+j}}} = \frac{p_{ij}}{\sqrt{p_{i+}p_{+j}}} - \sqrt{p_{i+}p_{+j}}$$

$$C_{p\times p} = Q'\cdot Q$$

$$c_{jk} = \sum_{i=1}^{N}\frac{(p_{ij} - p_{i+}p_{+j})(p_{ik} - p_{i+}p_{+k})}{\sqrt{p_{i+}p_{+j}}\sqrt{p_{i+}p_{+k}}}$$

$$C = A\cdot A' = U\Lambda U'$$

$$A = U\Lambda^{1/2}$$

$$A^* = QU$$

$$sum(\Lambda) = \vec{1}'\cdot Q^2\cdot\vec{1} = sum(diag(C))$$

$$A_a = diag(\vec{1}'\cdot P)^{-1/2}A$$

$$A_a^* = diag(P\cdot\vec{1})^{-1/2}A^*$$