The Standard Biplot

NOTE TO READERS OF THE ELECTRONIC PDF VERSION:

This article contains embedded dynamic graphics in Figures 3, 6 and 9, which can be seen by simply clicking on the figure to see it in motion. To enable this feature you should use the latest version of the free Acrobat Reader, version 9, to open the PDF file.

Michael Greenacre

Department of Economics and Business

Universitat Pompeu Fabra
Ramon Trias Fargas, 25-27
08005 Barcelona
SPAIN

E-mail: michael@upf.es

Abstract: Biplots are graphical displays of data matrices based on the decomposition of a matrix as the product of two matrices. Elements of these two matrices are used as coordinates for the rows and columns of the data matrix, with an interpretation of the joint presentation that relies on the properties of the scalar product. Because the decomposition is not unique, there are several alternative ways to scale the row and column points of the biplot, which can cause confusion amongst users, especially when software packages are not united in their approach to this issue. In an attempt to unify the scaling of the biplot we propose a new scaling of the solution, called the standard biplot, which can be applied to a wide variety of analyses such as correspondence analysis, principal component analysis, log-ratio analysis and the graphical results of a discriminant analysis/MANOVA, in fact to any method based on the singular-value decomposition. The standard biplot also handles data matrices with widely different levels of inherent variance. Two concepts taken from correspondence analysis are important to this idea: the weighting of row and column points, and the contributions made by the points to the solution. In the standard biplot one set of points, usually the rows of the data matrix, optimally represent the positions of the cases or sample units, which are weighted and usually standardized in some way unless the matrix contains values that are comparable in their raw form. The other set of points, usually the columns, is represented in accordance with their contributions to the low-dimensional solution. As for any biplot, the projections of the row points onto vectors defined by the column points approximate the centred and (optionally) standardized data. The method is illustrated with several examples to demonstrate how the standard biplot copes in different situations to give a joint map which needs only one common scale on the principal axes, thus avoiding the problem of enlarging or contracting the scale of one set of points to make the biplot readable. The proposal also solves the problem in correspondence analysis of low-frequency categories that are located on the periphery of the map, giving the false impression that they are important.

Keywords: Biplot, contributions, correspondence analysis, discriminant analysis, MANOVA, principal component analysis, scaling, singular-value decomposition, weighting.

1. Introduction

The *biplot*, a term introduced by Gabriel (1971) in the context of principal component analysis, is a graphical display of the rows and columns of a data matrix as points in a low-dimensional Euclidean space, usually of dimensionality two or three. Any matrix $\mathbf{T}$ ($I \times J$) can be decomposed in an infinite number of ways as the product of two matrices $\mathbf{A}$ ($I \times K$) and $\mathbf{B}^{\mathsf{T}}$($K \times J$):

$$\mathbf{T} = \mathbf{A}\,\mathbf{B}^{\mathsf{T}}\,. \tag{1}$$

We call $\mathbf{T}$ the *target matrix*, and $\mathbf{A}$ and $\mathbf{B}$ the *left* and *right matrices*, respectively, of the decomposition.

A convenient way of obtaining (1) is to use the singular-value decomposition (SVD) of $\mathbf{T}$:

$$\mathbf{T} = \mathbf{U}\,\mathbf{\Gamma}\,\mathbf{V}^{\mathsf{T}}, \text{ where } \quad \mathbf{U}^{\mathsf{T}}\mathbf{U} = \mathbf{V}^{\mathsf{T}}\mathbf{V} = \mathbf{I},\ \mathbf{\Gamma} = \mathrm{diag}(\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_K > 0). \tag{2}$$

Then $\mathbf{A}$ and $\mathbf{B}$ can be defined as:

$$\mathbf{A} = \mathbf{U}\,\mathbf{\Gamma}^{\alpha} \qquad \mathbf{B} = \mathbf{V}\,\mathbf{\Gamma}^{1-\alpha}\,. \tag{3}$$

for any value of α. The advantage of forming the decomposition using the SVD is the well-known Eckart-Young theorem which states that the least-squares matrix approximation of the target matrix $\mathbf{T}$ can be obtained for any given rank K^* of the approximating matrix, using the first (i.e., largest) K^* singular values in the diagonal of $\mathbf{\Gamma}$, and the corresponding first K^* left and right singular vectors in $\mathbf{U}$ and $\mathbf{V}$ (Eckart and Young 1936).

Thus a target matrix $\mathbf{T}$ can be approximated by its 'closest' (least-squares) approximation of rank 2, for example:

$$[\mathbf{u}_1 \quad \mathbf{u}_2]\begin{bmatrix}\gamma_1 & 0\\ 0 & \gamma_2\end{bmatrix}[\mathbf{v}_1 \quad \mathbf{v}_2]^{\mathsf{T}} \tag{4}$$

and then **A** and **B** could be defined, for example with $\alpha = 1$ in (3), as

$$\mathbf{A} = [\mathbf{u}_1 \quad \mathbf{u}_2]\begin{bmatrix}\gamma_1 & 0\\ 0 & \gamma_2\end{bmatrix} \qquad \mathbf{B} = [\mathbf{v}_1 \quad \mathbf{v}_2] \tag{5}$$

The rows of **A** and of **B** provide two-dimensional coordinates for points representing the rows and columns of the target matrix in a planar display, so that the scalar product between the i-th row point and the j-th column point approximates the (i,j)-th element s_{ij} of the target matrix. Since scalar products between two vectors **a** and **b** are equal to the length of the perpendicular projection of **a** onto **b** multiplied by the length of **b** (or vice versa), it follows that the direction vectors representing the columns (for example) can be calibrated so that the approximated values of the target matrix can be read off by projecting the row points perpendicularly onto each column vector. In this case we refer to the calibrated column vectors as *biplot axes*. For more details about biplots, see Gabriel and Odoroff (1990), Greenacre (1993) and Gower and Hand (1996).

In spite of the simplicity of the biplot, its use often causes confusion because of the many ways in which **A** and **B** can be formed from the SVD. While the scalar product rule and the calibration holds for all biplots, the use of different values of α produces biplots with different within-set interpretations of the clouds of row and column points. The usual choices α=1 and α=0 give biplots in which distances are approximated between rows and between columns respectively. Often the two sets of plotted points are on such different scales that they are plotted on two different scales (see, for example, the `biplot` function in the R package). To alleviate this the choice α=½ is sometimes recommended so that row and column points have

similar dispersions in the display, but this choice optimizes neither the distances between rows nor the distances between columns. This last choice, called the *symmetric biplot* by Greenacre (2007) has the least benefit for the user, since data matrices are invariably interpreted asymmetrically, with at least one of the sets of distances being important to visualize (see Greenacre (2007: Epilogue) for further discussion about this aspect). So we are interested in improving the biplot display for the choices α=1 or α=0, where either the row or column distances are optimally displayed. Our objective is to find a biplot which has the following characteristics:

(a) the row and column configurations should be able to be plotted to the same scale, independently of the level of variance in the data matrix;

(b) supposing that row distances are visualized, then the lengths of the vectors to the column points should have some advantage for interpretation;

(c) the biplot should be applicable across different types of multivariate methods that visualize data tables using the SVD, for example correspondence analysis, principal component analysis, log-ratio analysis and discriminant analysis.

The above three properties are satisfied by the proposed *standard biplot*, which we first define in the context of correspondence analysis.

2. The standard correspondence analysis biplot

In the standard biplot the weighting of the points and the contributions made by each point to the solution will play crucial roles. Since these concepts are inherent in correspondence analysis (CA), we choose to use CA to illustrate the general idea. Given a contingency table **N**, and associated correspondence table **P** = (1/n) **N** of relative frequencies, where n is the grand total of

N, let **r** and **c** be the row and column marginal totals of **P** (called row and column *masses* in CA terminology). Let $\mathbf{D}_r$ and $\mathbf{D}_c$ be the diagonal matrices of these masses. CA has as target matrix the standardized residuals of the relative frequencies:

$$\mathbf{T} = \mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{r}\mathbf{c}^\mathsf{T})\mathbf{D}_c^{-1/2} \tag{6}$$

which can be written equivalently as:

$$\mathbf{T} = \mathbf{D}_r^{1/2}(\mathbf{D}_r^{-1}\mathbf{P} - \mathbf{1}\mathbf{c}^\mathsf{T})\mathbf{D}_c^{-1/2} \tag{7}$$

It is this latter form (7) that is useful here, because it shows that CA is the analysis of the row profiles (rows of frequencies relative to their totals, in the matrix $\mathbf{D}_r^{-1}\mathbf{P}$), centred with respect to the average profile $\mathbf{c}^\mathsf{T}$, with rows weighted by the masses in **r** and the distances between row profiles defined by chi-square distances using the metric $\mathbf{D}_c^{-1}$. After decomposing **T** by the SVD, as in (2): $\mathbf{T} = \mathbf{U}\,\boldsymbol{\Gamma}\,\mathbf{V}^\mathsf{T}$, where $\mathbf{U}^\mathsf{T}\mathbf{U} = \mathbf{V}^\mathsf{T}\mathbf{V} = \mathbf{I}$, the rows and columns are often plotted according to the so-called *principal coordinates* of the rows: $\mathbf{F} = \mathbf{D}_r^{-1/2}\mathbf{U}\boldsymbol{\Gamma}$ and the *standard coordinates* of the columns: $\mathbf{Y} = \mathbf{D}_c^{-1/2}\mathbf{V}$. The rationale for this choice of coordinates is as follows: the row points defined by **F** visualize approximate chi-square distances between row profiles that are weighted by the row masses; the column points defined by **Y** are the projections of the unit profiles, or vertex points of the simplex space of the profiles; and row points are weighted averages of the column vertex points, using the elements of the row profiles as weights: $\mathbf{F} = \mathbf{D}_r^{-1}\mathbf{P}\mathbf{Y}$. The joint display of **F** and **Y** is a biplot, which from their definitions and from (7) is seen to be:

$$\mathbf{F}\mathbf{Y}^\mathsf{T} = \mathbf{D}_r^{-1/2}\mathbf{U}\boldsymbol{\Gamma}\mathbf{V}^\mathsf{T}\mathbf{D}_c^{-1/2} = (\mathbf{D}_r^{-1}\mathbf{P} - \mathbf{1}\mathbf{c}^\mathsf{T})\mathbf{D}_c^{-1} \tag{8}$$

Hence scalar products between row and column points in a low-dimensional display (using the first columns of **F** and **Y**) are approximating the centred profiles relative to their average values: $(p_{ij}/r_i - c_j)/c_j$. This joint plot is also called the (row principal) asymmetric map – see Greenacre (1993, 2007).

The fact that the row profiles and the extreme unit profiles (vertices) lie in the same multidimensional space is an attractive property of the asymmetric map, but in practice it is problematic when the variance in the data, called *inertia* in CA, is low. For example, the 'author' data set, available in the **ca** package of Nenadić and Greenacre (2007), consists of the counts of the 26 letters of the alphabet in 12 samples of text from books by famous English authors (the data are reproduced in Greenacre and Lewi (2009)). The profiles of the books across the letters are very similar, as expected, and the inertia is consequently very low. The asymmetric map of this table, reproduced from Greenacre (2007) as Figure 1, shows the 12 books clustered so tightly at the centre of the map that they cannot be labelled. This is clearly not a good biplot and the scale of one of the sets of points needs to be changed to make the joint plot legible. The directions defined by the column points are the correct ones, but they are too long. Rather than simply reduce their overall scale, the proposed standard biplot reduces their length by multiplying each column point (representing a letter in the example) by the square root of the mass, $c_j^{½}$, associated with the column point. Thus all letters are pulled in, but low frequency letters such as z and q are pulled in a lot more than high frequency letters such as a and e. This is a premultiplication of **Y** by $\mathbf{D}_c^{1/2}$ and converts (8) into the following biplot:

$$\mathbf{F}(\mathbf{D}_c^{1/2}\mathbf{Y})^{\mathsf{T}} = (\mathbf{D}_r^{-1}\mathbf{P} - \mathbf{1}\mathbf{c}^{\mathsf{T}})\mathbf{D}_c^{-1/2} \qquad (9)$$

The elements being approximated (shown in the target matrix on the right of (9)) are now the standardized differences $(p_{ij}/r_i - c_j)/c_j^{½}$ between the profile and its average, standardized by

dividing by the square root of the average, as in the chi-square metric. The right matrix $\mathbf{D}_c^{1/2}\mathbf{Y}$ of the biplot in (9), representing the columns, is the original matrix $\mathbf{V}$ of right singular vectors. The fact that standardized values are being biplotted has suggested the term 'standard biplot'. The standard CA biplot for the same data is given in Figure 2, and clearly the problem of the difference in scales has been cleared up. But, in addition, the lengths of each vector joining the letters to the origin are related to the contributions that the letters make to the principal axes. Since it is these letters that are determining the solution, it is correct that they have more prominent positions in the joint display.

To show the connection with the contributions, the parts of inertia due to each column along the principal axes are, in scalar form for column j along axis k: $c_j y_{jk}^2$, relative to the sum of these quantities, which is equal to 1 from the definition of $\mathbf{Y}$: $\mathbf{Y}^\mathsf{T}\mathbf{D}_c\mathbf{Y} = \mathbf{V}^\mathsf{T}\mathbf{V} = \mathbf{I}$. Thus by 'shrinking' each y_{jk} by $c_j^{\frac{1}{2}}$, the square of the new coordinate is exactly the contribution of the j-th point to the inertia along the k-th axis.

Thus the standard biplot of the 'author'data in Figure 2 shows that the letters d, h, w and c have high contributions to the first axis, and y is especially dominant on the second – the closer letters are to the centre the less important they are to the interpretation. Table 1 lists the contributions for each letter to the two axes. Thus, d, for example, has contributions of 0.1704 to axis 1 and 0.0593 to axis 2, and the square roots of these values are $0.1704^{\frac{1}{2}} = 0.413$ and $0.0593^{\frac{1}{2}} = 0.244$, respectively. The coordinates of d in Figure 2 are (again from Table 1, the square roots of the masses multiplied by the standard CA coordinates) $0.046^{\frac{1}{2}} \times (-1.9256) = -0.413$ and $0.046^{\frac{1}{2}} \times 1.1354 = 0.244$ – this is an empirical check that the squares of the coordinates are the contributions. Notice, however, that the squared lengths of the vectors, equal to the sum of their

squared lengths along the axes, are not equal to their contributions to the two-dimensional map, since the inertias along the axes are not the same.

Figure 3 is a dynamic graphic which shows the transition from Figure 1 to Figure 2, as the biplot vectors for the letters are gradually shrunk proportional to the square roots of their respective masses. The size of the triangular symbol for each letter is proportional to the mass of the letter, so it is the points with smaller triangles that are pulled in the most by the standard biplot scaling. During the transition the row (book) points are in fixed positions and the scale of the map is adjusted as the letter points shrink differentially towards the centre – the dashed box shown throughout the animation actually delimits the area of the final frame of the animation in which the standard CA biplot resides.

The standard CA biplot functions equally well for a table with very low inertia, such as the data set 'author', as for a table with high inertia. The data set 'benthos' (reported in Greenacre and Lewi (2009) and also analyzed in Greenacre (2007: chapter 10)) consists of counts of 92 marine species at 13 different locations on the North Sea bed, 11 of which lie close to an oil-drilling platform that causes sea-bed pollution, while the other two (R40 and R42) are reference stations far away from the pollution source. These data have high inertia, which is characteristic of ecological abundance data of this type. Thus, the station points are much more spread out in the profile space with respect to the extreme vertex points, as shown in the asymmetric CA map of Figure 4. Figure 5 shows the standard CA biplot of the same data and Figure 6 shows the dynamic transition from the asymmetric map (Figure 4) to the standard CA biplot (Figure 5). Again each species point is shown with a symbol proportional to its overall abundance. Only the top 10 contributing species have been labelled – the remaining 82 unlabelled species are all closer to the centre of the map in the CA biplot and contribute only 15% to the solution. Showing the species' vectors in terms of their contributions assists and clarifies the

interpretation by concentrating attention on the most influential species. In Figure 5 it can be clearly seen that there is a group of species chiefly responsible for the separation to the right of the unpolluted reference stations R40 and R42 (these would be the species most affected and reduced by the pollution). There is mainly one species (*Chaetezona setosa*) that dominates the polluted stations, especially station S15 which is one of the closest to the oil-drilling platform and most contaminated. Station S24 is a special case because there is an unusually high occurrence of *Myriochele oculata*, which is unrelated to the pollution gradient.

Notice that in Figure 2 the book points show less dispersion than the station points in Figure 5, relative to the spread of the 'variables' in each biplot (letters in Figure 2, species in Figure 5, respectively). This partly testifies to the lower inertia in the 'author' data set, compared to the 'benthos' data set, although the impression of the spread of the rows relative to the columns is also governed by how many variables there are. The more variables there are, the more the numerical values of the contributions are reduced as a group (because the contributions sum to 1 for each axis).

3. The standard principal component biplot

In the case of principal component analysis (PCA), where data are assumed on interval scales, there are two distinct situations: variables that need to be standardized, and variables that are all on the same scale and that need no standardization. The first situation is the most common, when variables are on different measurement scales. Table 2 shows the 'environ' data set consisting of $J = 10$ variables measured at the same $I = 13$ locations as the previous 'benthos' data set. The usual PCA approach is to standardize the variables, in which case the total variance is equal to 10, the number of variables. In calculating the variances of each variable with a view to standardization, we use I, not $I–1$, to average the squared deviations from the mean. In other words, we weight each sample by a constant $1/I$, which mimics the row

weighting idea in CA, except in the CA case there are different weights for each row. We could, however, weight the samples using the CA masses, which is what is done if we used the environmental variables as constraints in a canonical correspondence analysis (CCA).

Distances between samples are Euclidean on the standardized scales and these distances will increase as the number of variables in the analysis increases. Since we wish to limit the variation in the samples so that we can eventually plot sample distances and variable contributions on the same biplot, a convenient rescaling is to calculate Euclidean distances using average squared differences between (standardized) variables, not sum of squared differences. For a table **X** of standardized data, the distance between samples *i* and *i′* is:

$$\sqrt{(1/J)\sum_{j=1}^{J}(x_{ij} - x_{i'j})^2} \tag{10}$$

This again mimics the idea in CA of weighting the columns of the matrix – in PCA all the columns now receive an equal weight of 1/*J* compared to the differential weighting in CA. The total variance will then be averaged as well, equal to 1 rather than 10 in this example. The SVD for PCA then multiplies the data by the square roots of the row and column masses (assuming **X** is already column-centred, so that all variables have mean 0 and variance 1):

$$\sqrt{(1/I)}\sqrt{(1/J)}\mathbf{X} = \sqrt{(1/IJ)}\mathbf{X} = \mathbf{U\Gamma V}^{\mathsf{T}} \tag{11}$$

Hence the sum of the eigenvalues (that is, the sum of squared singular values) is equal to 1, and not *J* (the number of variables) as in the usual definition of PCA. The biplot coordinates are $\mathbf{F} = \sqrt{I}\mathbf{U\Gamma}$ (for the rows, where $\sqrt{I}$ is the inverse of the row masses, as in CA) and $\mathbf{Y} = \mathbf{V}$ (for the columns, where there is no rescaling by the masses, as in the standard CA biplot). Again the

squared lengths of the column coordinates are equal to the contributions of the respective variables to the respective principal axes.

Figure 7 shows the standard PCA biplot of the 'environ' data set of Table 2. All the variables measuring pollution (THC=Total Hydrocarbon Content, and the metals) point to the left and contribute the most to the horizontal axis, which separates the most polluted station S15 on the left (one of the closest to the oil rig) from the unpolluted reference stations R40 and R42 on the right. Distance versus depth contribute strongly to the second axis, with the TOM (=Total Organic Material) biplot axis positive – both the highly polluted station S15 and the two reference stations are high on TOM.

The reference stations are at shallower depths than the others, but only by a few metres, and this difference is accentuated by the standardization. Also because the variables are highly skew, a logarithmic transformation might be considered appropriate – the log-transformed data are then not standardized, so that the multiplicative differences are visualized. Figure 8 shows the standard PCA biplot of the log-transformed data, and the role of the variable depth is considerably reduced, with most of the variance now concentrated on the first axis which is the pollution gradient. This also shows that the standard PCA biplot functions well for unstandardized log-transformed data as well. The only slightly problematic case would be when all the original variables are on the same scale, and standardization is not required, but the common scale has a very small or very large range (for example, all data are ratings from 0 to 100). Then the scale of the row points will be quite different to that of the column points, which can be rectified by some appropriate overall rescaling of the original data, for example to rescale the ratings from 0 to 10.

It is important to realize that the standard PCA biplot described above is not the same as the biplot in which the variables are depicted by their correlations with the principal axes, and thus lie within a unit circle in the two-dimensional solution, for example. This latter biplot associates the singular values with the column coordinates, depicting rows (cases) with coordinates in $\sqrt{I}\mathbf{U}$ (see the SVD in (11)) and the columns (variables) with coordinates in $\mathbf{V}\boldsymbol{\Gamma}$. The squares of the correlations are the parts of the (unit) variance of that variable explained by the axes, and can be summed for the two axes, hence the drawing of a unit circle which represents a perfect explanation of the variable. The average sum of squared column coordinates on an axis is then the part of the total variance on that axis, while the cases are represented by coordinates that have average sum of squares equal to 1 on every principal axis.

4. Log-ratio analysis

Log-ratio analysis is the visualization of a table of strictly positive data, log-transformed and double-centred, again using the SVD. The method is popular in the analysis of compositional data (Aitchison 1986). There are two forms, an unweighted form (see Aitchison 1983, Aitchison and Greenacre 2002) and a weighted form, the latter also known as 'spectral mapping' (Lewi 1976, Greenacre and Lewi 2009). The row and column weights commonly used in weighted log-ratio analysis are the same as the masses in CA, that is the relative values of the column margins of the original table. Using the same notation as in CA for the row and column masses $\mathbf{r}$ and $\mathbf{c}$, and $\log(\mathbf{N})$ to denote the matrix of log-transformed data, the target matrix and its SVD are:

$$\mathbf{T} = \mathbf{D}_r^{1/2}(\mathbf{I} - \mathbf{1}\mathbf{r}^\mathsf{T})\log(\mathbf{N})(\mathbf{I} - \mathbf{1}\mathbf{c}^\mathsf{T})^\mathsf{T}\mathbf{D}_c^{1/2} = \mathbf{U}\boldsymbol{\Omega}\mathbf{V}^\mathsf{T} \tag{12}$$

(just substitute $\mathbf{r} = (1/I)\mathbf{1}$ and $\mathbf{c} = (1/J)\mathbf{1}$ for the unweighted form). Assuming rows were cases/samples and the columns variables/components, the standard log-ratio biplot would then

use the principal coordinates $\mathbf{D}_r^{-1/2}\mathbf{U}\boldsymbol{\Omega}$ for the rows and $\mathbf{V}$ for the columns. Some care is needed here, since the analysis is often performed to detect equilibrium in a subset of components, which can be diagnosed by some components lining up in straight lines (see, for example, Aitchison and Greenacre 2002). But the standard biplot normalization, while having the advantage of showing which components are important for the interpretation of the biplot, has the disadvantage of destroying this property. A dynamic graphic, which alternates between the usual map where the columns are shown in principal or standard coordinates and the standard biplot, can be useful to diagnose models in subcompositions as well as see which components are determinant in the solution – Figure 9 shows an example of this for the 'author' data, starting with a LRA solution that allows model diagnosis and ending with the standard biplot that represents component contributions.

5. The standard discriminant analysis / MANOVA biplot

Linear discriminant analysis (LDA) and multivariate analysis of variance (MANOVA) rely on a common theory, which reduces to a SVD of the matrix of group centroids, weighted by their respective group sizes, in a space structured by the Mahalanobis metric – see, for example, Mardia, Kent and Bibby (1979). The resultant principal axes are often called canonical axes in this context, and the terminology canonical variate analysis (CVA) is a common synonym as well. This is another case where the standard biplot can be used to represent the centroids by their projections onto the optimal plane along with the points for the variables such that their squared lengths along a canonical axis are the contributions to the group discrimination along the axis. The idea of a MANOVA biplot originates in the work of Gabriel (1972).

Suppose that there are G groups of cases (rows) in a cases-by-variables data matrix, and that the G vectors of variable averages are contained in the $G\times J$ matrix $\mathbf{G}$. The averages are centred with respect to the global averages of the variables across all the cases – these global averages

are identical to the centroid of the group averages, where each group is weighted proportionally to the number of cases in it. A $J \times J$ covariance matrix $\mathbf{S}_g$ is computed within each group – as in PCA the averaging of the squares and cross-products in the computation is performed by dividing by the number of cases n_g in each group, not $n_g - 1$. The estimated common covariance matrix $\mathbf{S}$ is then computed as $\Sigma_g (n_g/n)\mathbf{S}_g$. Weights of n_g/n are assigned to each group average, and these weights form the diagonal of the diagonal matrix $\mathbf{D}_w$.

The weighted PCA of the group averages, in the Mahalanobis metric $\mathbf{S}^{-1}$, equivalent to LDA/MANOVA/CVA is achieved by the following SVD, imitating the procedure followed for CA and PCA:

$$\mathbf{T} = \mathbf{D}_w^{1/2} \mathbf{G} \mathbf{S}^{-1/2} \sqrt{1/J} = \mathbf{U} \mathbf{\Gamma} \mathbf{V}^{\mathsf{T}} \tag{13}$$

Thus the standard biplot would visualize the rows by the principal coordinates of the group centroids in $\mathbf{D}_w^{-1/2} \mathbf{U} \mathbf{\Gamma}$ and the coordinates $\mathbf{V}$ of the columns (variables) that depict the contributions to the principal axes. Figure 10 shows the standard biplot in the two-dimensional plane of the three centroids of the famous Fisher 'iris' data.

6. Discussion and conclusions

The subject of the scaling in biplots, and in particular in CA, has been the subject of much discussion and controversy. The most popular way of reporting the results of CA has been the so-called *symmetric map*, sometimes called the 'French plot' or 'Benzécri plot'. This solution, which allocates the singular values to both the left and right matrices of the singular-value decomposition (that is, both rows and columns in principal coordinates) is not a biplot, but displays inter-row and inter-column distances optimally, with the aesthetic advantage of spreading out the row and column points similarly over the map area. The asymmetric map, on

the other hand, only works well in high-inertia cases, with the scaling problem demonstrated in Figure 1 when the inherent inertia is low.

Amongst all biplot options for a typical cases × variables data set, what we are really interested in is the direction made by each 'variable' of the data set, so the proposal of this paper is to maintain these directions while making the lengths of the biplot vectors both meaningful and also on a scale which allows joint representation of the 'case' points without the need for two sets of scales for the row and column points. Both these objectives are attained by the standard biplot. The case points are always in their principal coordinate positions, so that distances between case points are interpretable. The variables are indicated by lines coinciding with the biplot axes, which could be calibrated so that the projections of the case points on a biplot axis approximate the values for that variable, but the lengths of the lines along each principal axis of the biplot are related to the contribution made by that variable to the solution. It is called the standard biplot because it generally reconstructs data on a standardized scale and also because the sum of squared lengths of these lines on any principal axis is equal to 1.

Although motivated principally in correspondence analysis because of the problem of differential masses of the row and column points, it has also been shown that the standard biplot functions well in standardized and unstandardized principal component analysis, log-ratio analysis and discriminant analysis. We believe that the standard biplot can become, as its name doubly implies, a standard way of showing the results of methods based on the singular-vaue decomposition, and may remove most of the controversy and doubt surrounding the scaling of the solutions of these methods.

Computing note

The **ca** package (Nenadić and Greenacre 2007) in R (R development core team, 2008) was used to perform the CA calculations in this article, while all others were computed using matrix functions in R function, notably `svd`. The standard biplot option for CA already exists in the **ca** package, for example `map="rowgreen"` for the standard (row principal) biplot. The option `map="rowgab"` is the one proposed by Gabriel and Odoroff (1990) where the standard coordinates of 'variable' points are multiplied by their masses, not the square roots of the masses as in the standard biplot.

Acknowledgments

This research has been supported by the Fundación BBVA, Madrid, Spain, and the author expresses his gratitude to the Foundation's director, Prof. Rafael Pardo. Partial support of Spanish Ministry of Education and Science grant MEC-SEJ2006-14098 is also acknowledged. This paper was mostly written while the author was on leave at Stanford University: thanks to Trevor Hastie and colleagues of the Department of Statistics.

References

Aitchison, J. (1983). Principal component analysis of compositional data. *Biometrika*, **70**, 57–65.

Aitchison, J. (1986), *The Statistical Analysis of Compositional Data*. London: Chapman & Hall, reprinted in 2003 by Blackburn Press.

Aitchison, J. and Greenacre, M.J. (2002). Biplots of compositional data. *Applied Statistics*, **51**, 375–392.

Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, **1**, 211–218.

Gabriel, K.R. (1972). Analysis of meteorological data by means of canonical decomposition and biplots. *Journal of Applied Meteorology*, **11**, 1071–1077.

Gabriel, K.R. and Odoroff, C.L. (1990). Biplots in biomedical research. *Statistics in Medicine*, **9**, 469–485.

Gower, J.C. and Hand, D.J. (1996). *Biplots*. London: Chapman and Hall.

Greenacre, M. J. (1993). Biplots in correspondence analysis. *Journal of Applied Statistics*, **20**, 251–269.

Greenacre, M. J. (2007). *Correspondence Analysis in Practice. Second Edition.* London: Chapman & Hall / CRC Press. Published in Spanish translation by the Fundación BBVA, Madrid, 2008, freely downloadable from http://www.fbbva.es/TLFU/tlfu/ing/publicaciones/fichalibro/index.jsp?codigo=300

Greenacre, M. J. and Lewi, P. (2009). Distributional equivalence and subcompositional coherence in the analysis of compositional data, contingency tables and ratio-scale measurements. *Journal of Classification*, **26**, 29–54.

Lewi, P.J. (1976). Spectral mapping, a technique for classifying biological activity profiles of chemical compounds. *Arzneimittel Forschung*, **26**, 1295–1300.

Mardia, K.V. , Kent, J.T. and Bibby, J.M. (1979. *Multivariate Analysis*. London: Academic Press.

Nenadić, O. and Greenacre, M. J. (2007). Correspondence analysis in R, with two- and three-dimensional graphics: The **ca** package. *Journal of Statistical Software*, **20** (1). URL http://www.jstatsoft.org/v20/i03/

Oksanen J., Kindt R., Legendre P. and O'Hara R.B. (2006). **vegan**: Community Ecology Package version 1.8-3. URL http://cran.r-project.org

R Development Core Team (2008). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org

Table 1: Masses (relative frequencies), contributions (relative to total), and standard coordinates of the 26 letters, for the two principal axes of Figures 1 and 2. The coordinates of the letters in the standard biplot of Figure 2 are the square roots of the masses multiplied by the standard coordinates. The squares of these coordinates are equal to the contributions.

	Masses	***Contributions***		***Standard coordinates***	
		Axis 1	*Axis 2*	*Axis 1*	*Axis 2*
a	0.0798	0.0000	0.0082	0.0176	0.3203
b	0.0157	0.0152	0.0025	0.9845	0.3980
c	0.0228	0.1020	0.0430	2.1150	1.3734
d	0.0460	0.1704	0.0593	-1.9256	1.1354
e	0.1271	0.0010	0.0596	0.0867	0.6848
f	0.0194	0.0317	0.0104	1.2765	0.7330
g	0.0200	0.0209	0.0025	-1.0207	-0.3530
h	0.0649	0.1463	0.0059	-1.5013	0.3024
i	0.0701	0.0050	0.0555	0.2675	-0.8895
j	0.0008	0.0002	0.0007	0.4533	-0.9160
k	0.0092	0.0697	0.0139	-2.7552	-1.2316
l	0.0427	0.0442	0.0012	1.0183	0.1650
m	0.0255	0.0130	0.0500	0.7127	-1.4010
n	0.0690	0.0028	0.0120	-0.2004	0.4173
o	0.0766	0.0009	0.0312	-0.1085	-0.6380
p	0.0152	0.0393	0.0512	1.6108	1.8379
q	0.0007	0.0111	0.0025	4.0798	1.9148
r	0.0519	0.0181	0.0280	0.5914	0.7342
s	0.0607	0.0449	0.0100	0.8602	-0.4056
t	0.0930	0.0009	0.0038	-0.1005	-0.2031
u	0.0298	0.0008	0.0308	0.1633	-1.0171
v	0.0096	0.0500	0.0003	2.2813	-0.1770
w	0.0258	0.1614	0.0021	-2.4992	0.2847
x	0.0012	0.0129	0.0206	3.3405	-4.2154
y	0.0219	0.0000	0.4851	0.0015	-4.7061
z	0.0008	0.0371	0.0099	6.8081	3.5092

Table 2: Data set 'environ' of 10 variables measured at 13 locations in the North Sea. Distance is from the oil drilling platform (the source of the pollution). The two reference stations R40 and R42 are 10 kms away and reflect an unpolluted environment.

	DEPTH	Ba	Cd	Cu	Fe	Pb	Zn	THC	TOM	DISTANCE
S4	76	1656	0.02	1.3	2022	11.7	16.1	8	0.6	750
S8	74	1373	0.02	1.1	2398	8.9	13.6	9	0.7	1800
S9	76	3680	0.07	3.9	2985	34.4	45.7	103	0.8	800
S12	72	2094	0.04	1.2	2535	21.3	15.1	6	0.7	2500
S13	72	2813	0.04	1.6	2612	17.3	18	27	1	1300
S14	76	4493	0.06	3.3	2515	28.8	26.7	546	1	850
S15	76	6466	0.14	6.2	3421	61.4	72.5	952	1.1	500
S18	72	1661	0.02	1.3	2381	19.8	13.8	10	0.9	2500
S19	72	3580	0.05	2.4	3452	33.7	28.9	32	1	1200
S23	76	2247	0.02	1.5	3457	21.4	14.9	16	0.7	1000
S24	76	2034	0.05	2.7	2311	15.5	16.8	11	0.7	450
R40	67	40	0.01	0	1804	5.9	5.9	3	0.8	10000
R42	67	85	0.01	0.17	1815	6.7	5.9	3	1	10000

Figure 1: Correspondence analysis of data set 'author', with rows in principal coordinates and columns in standard coordinates (row principal asymmetric map), showing row profiles bunched up near the origin of the biplot.

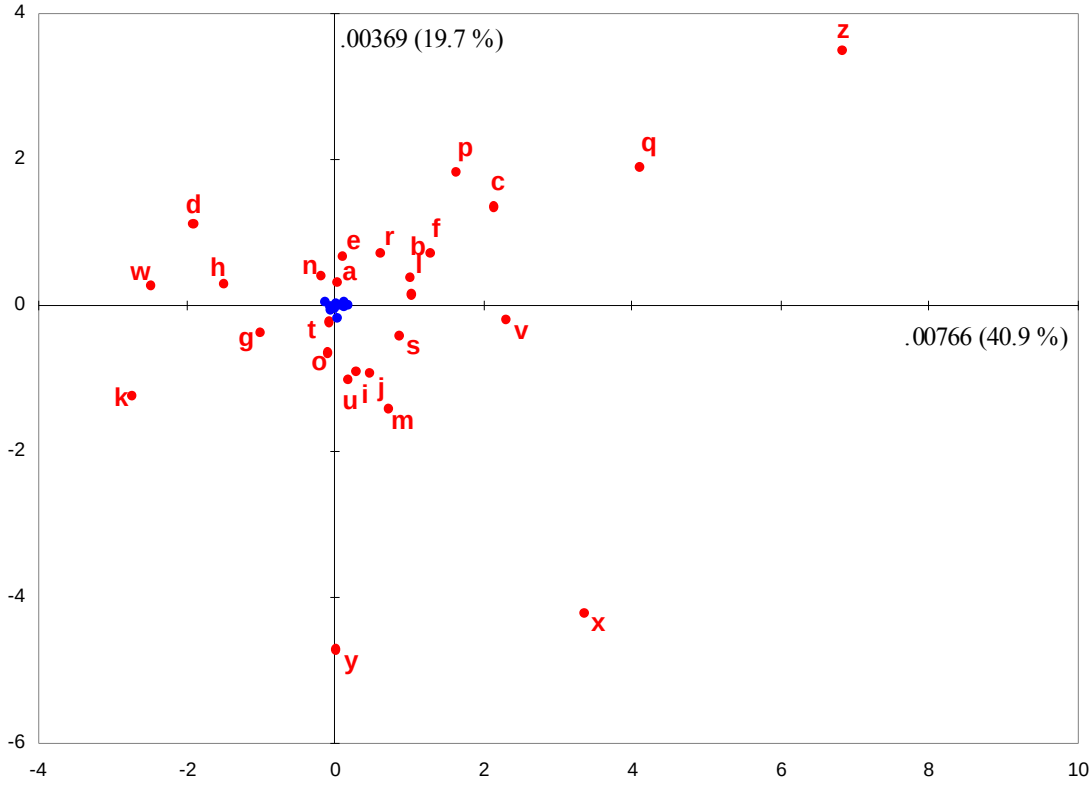

Figure 2: Standard correspondence analysis biplot of 'author' data, showing the letters as the same directions as in Figure 1, but multiplied by the square roots of their masses: the squared coordinates of each letter on an axis is equal to its contribution to that axis.

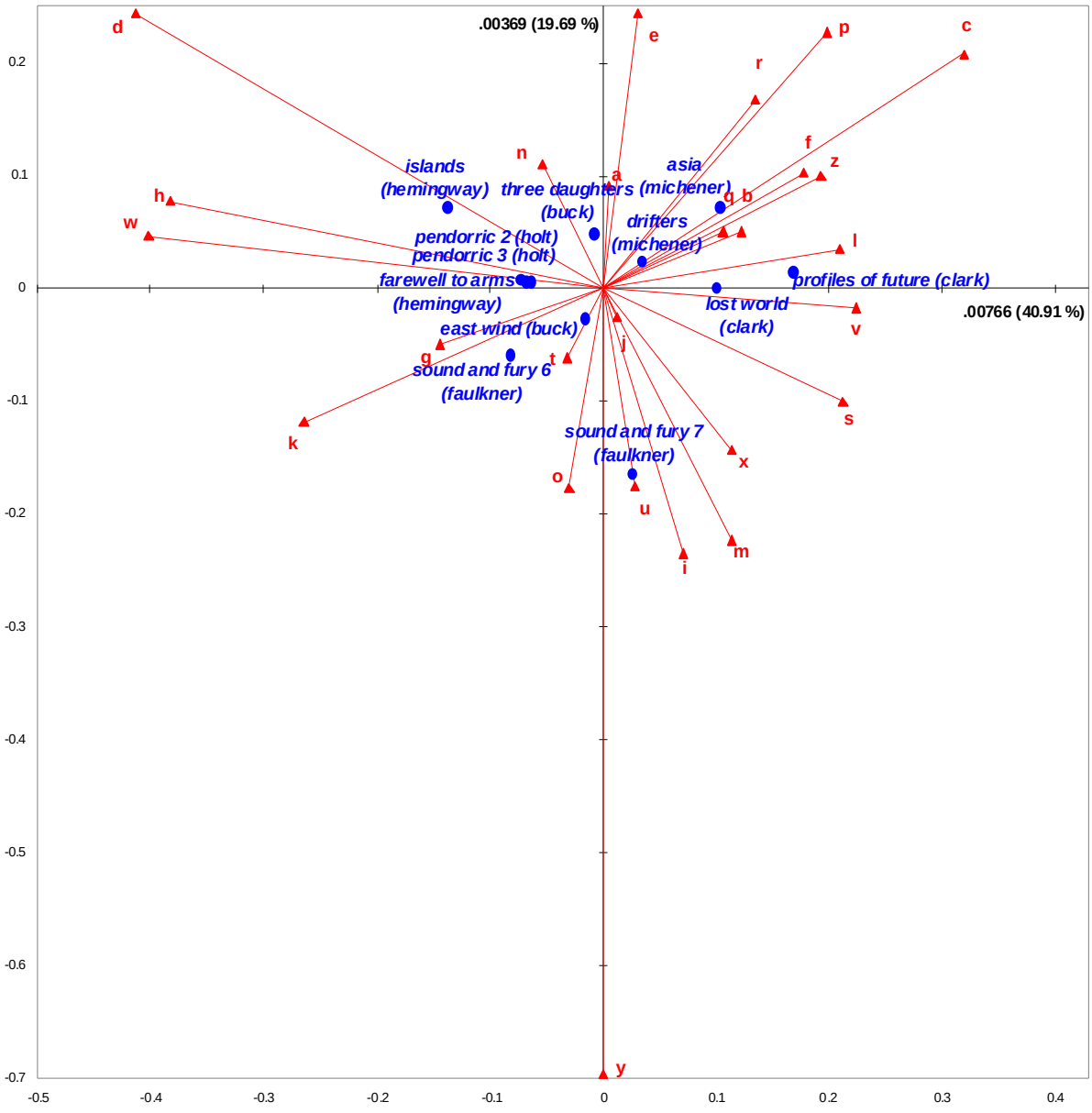

Figure 3: Dynamic transition between the asymmetric CA map (Figure 1) and the standard CA biplot (Figure 2) of the data set 'author'. Abbreviations for the books are used – see Figure 2 for full titles. ***Click on the figure to see the animation.***

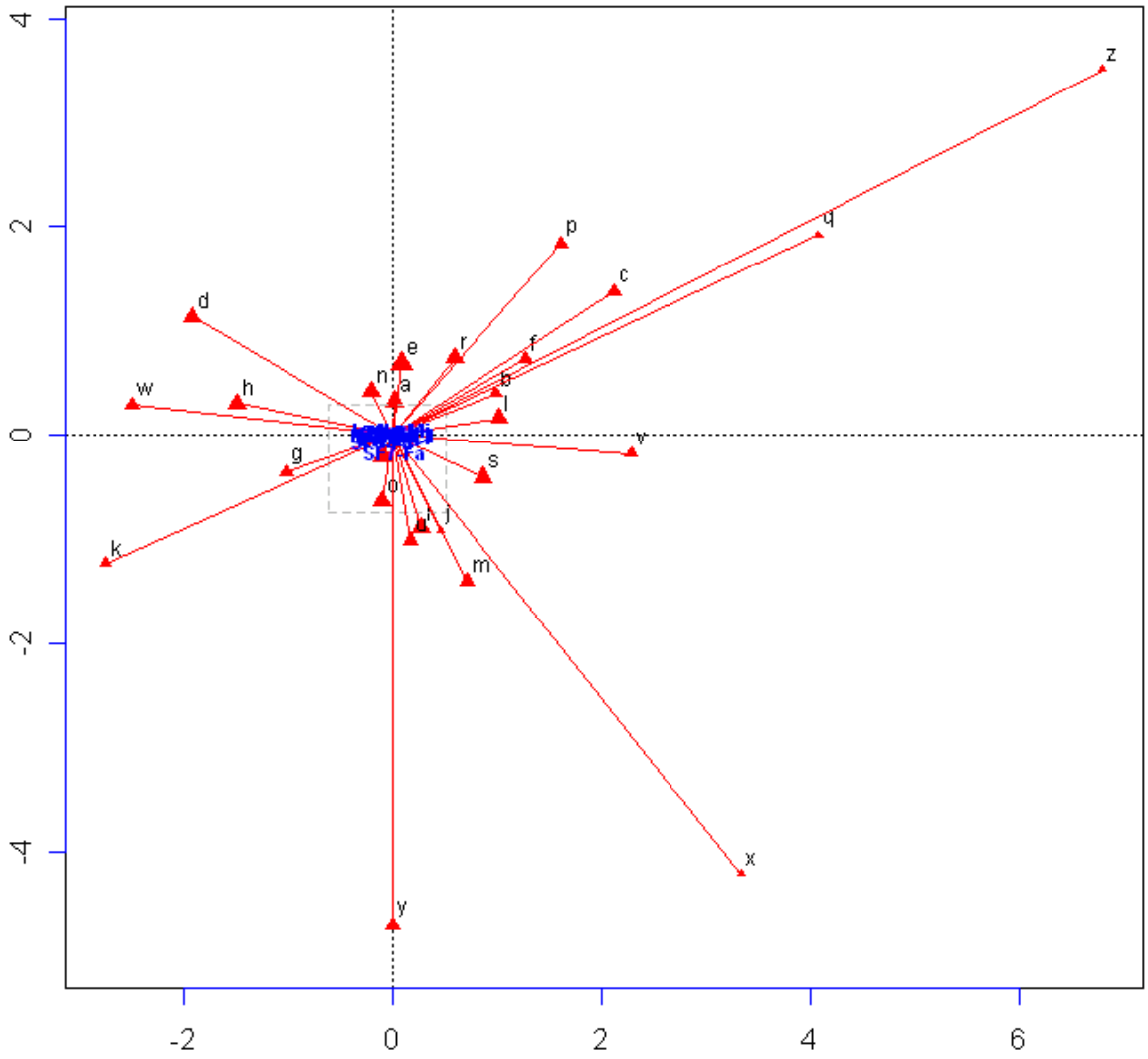

Figure 4: Asymmetric CA map of 'benthos' data set, with stations in principal coordinates, and species in standard coordinates (this is the first frame of the animation in Figure 6). The top 10 contributing species are labelled, accounting for 85% of the inertia of the two-dimensional solution. These are all highly abundant species while the remaining 82 lower abundance species collectively contribute only 15% to the solution. Notice that many of these low-contributing species are outlying in this map. The triangular symbols for the species are proportional to the overall species abundance. Inertias on axes 1 and 2 are 0.2457 (31.4%) and 0.2043 (26.1%) respectively.

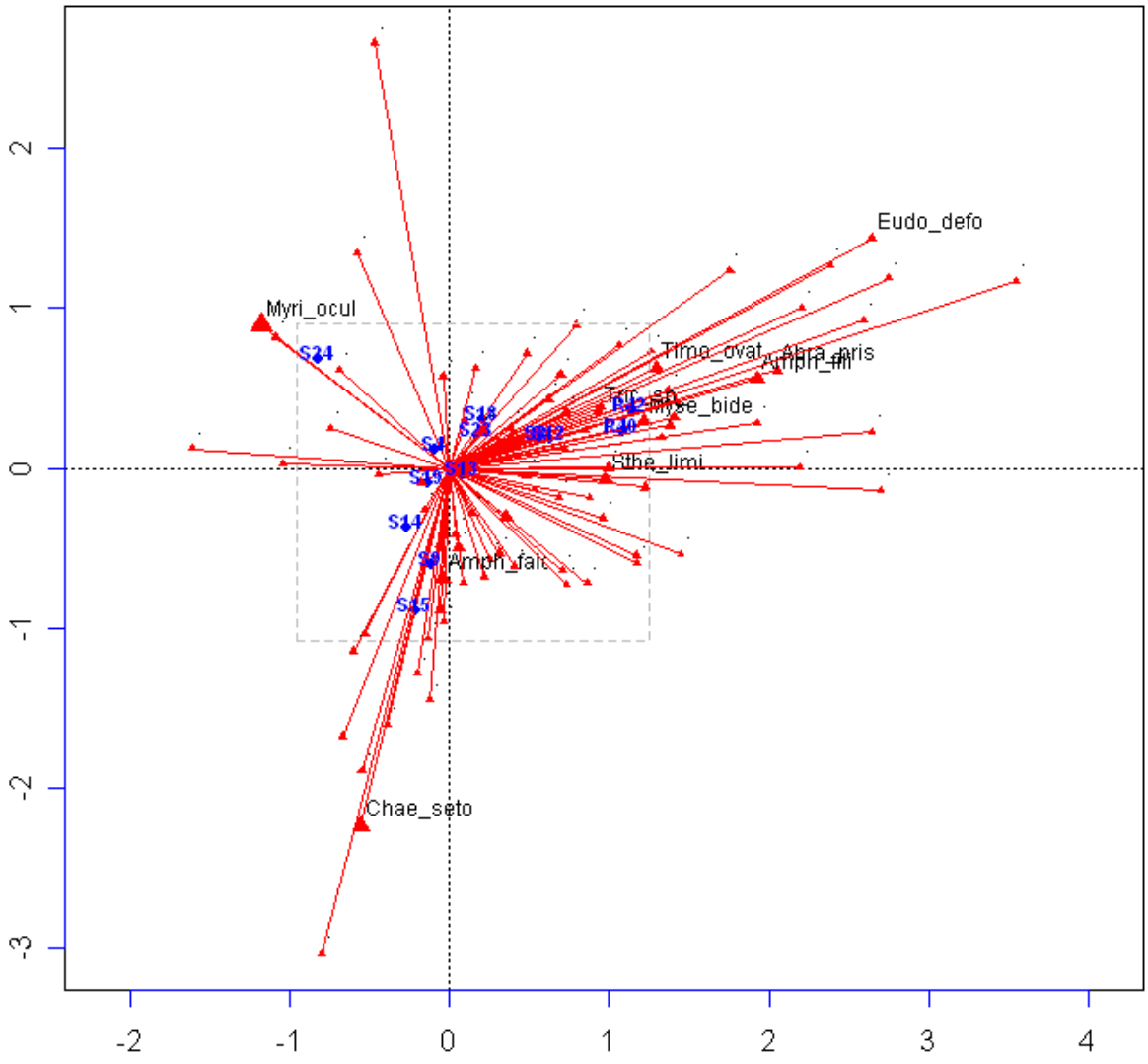

Figure 5: Standard CA biplot of 'benthos' data set. The top 10 contributing species are now the most outlying while the remaining 82 lower abundance species that collectively contribute very little all lie towards the centre of the biplot. The station points are in the same positions as in Figure 4 and have inertias 0.2457 (31.4%) and 0.2043 (26.1%) on axes 1 and 2 respectively. This is the last frame of the animation of Figure 6, and its boundaries are given by the dashed box of Figure 4, showing how much we have "zoomed in" on the configuration of station points.

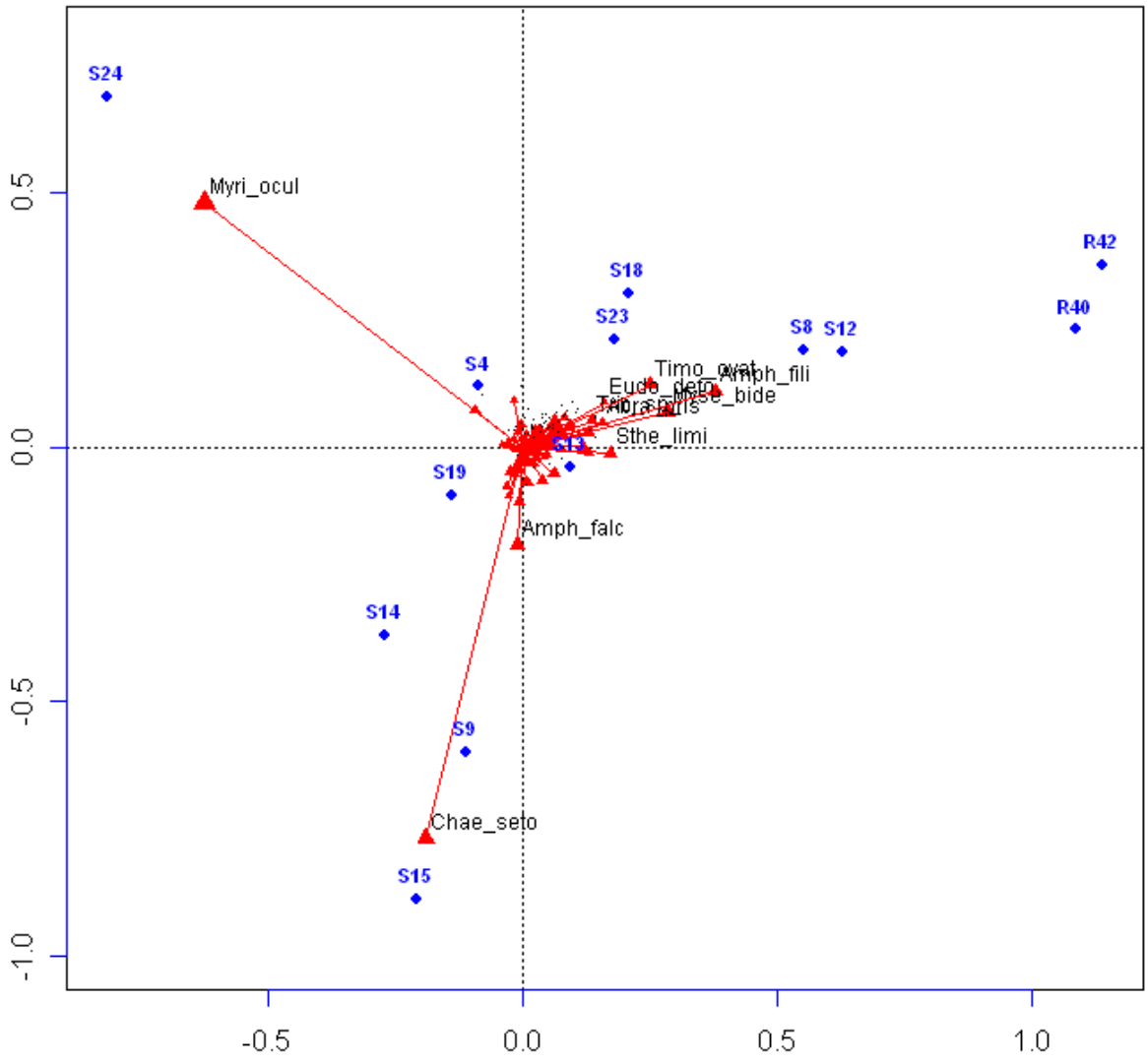

Figure 6: Dynamic transition between the asymmetric CA map (Figure 4) and the standard CA biplot (Figure 5) of the data set 'benthos'. ***Click on the figure to see the animation.***

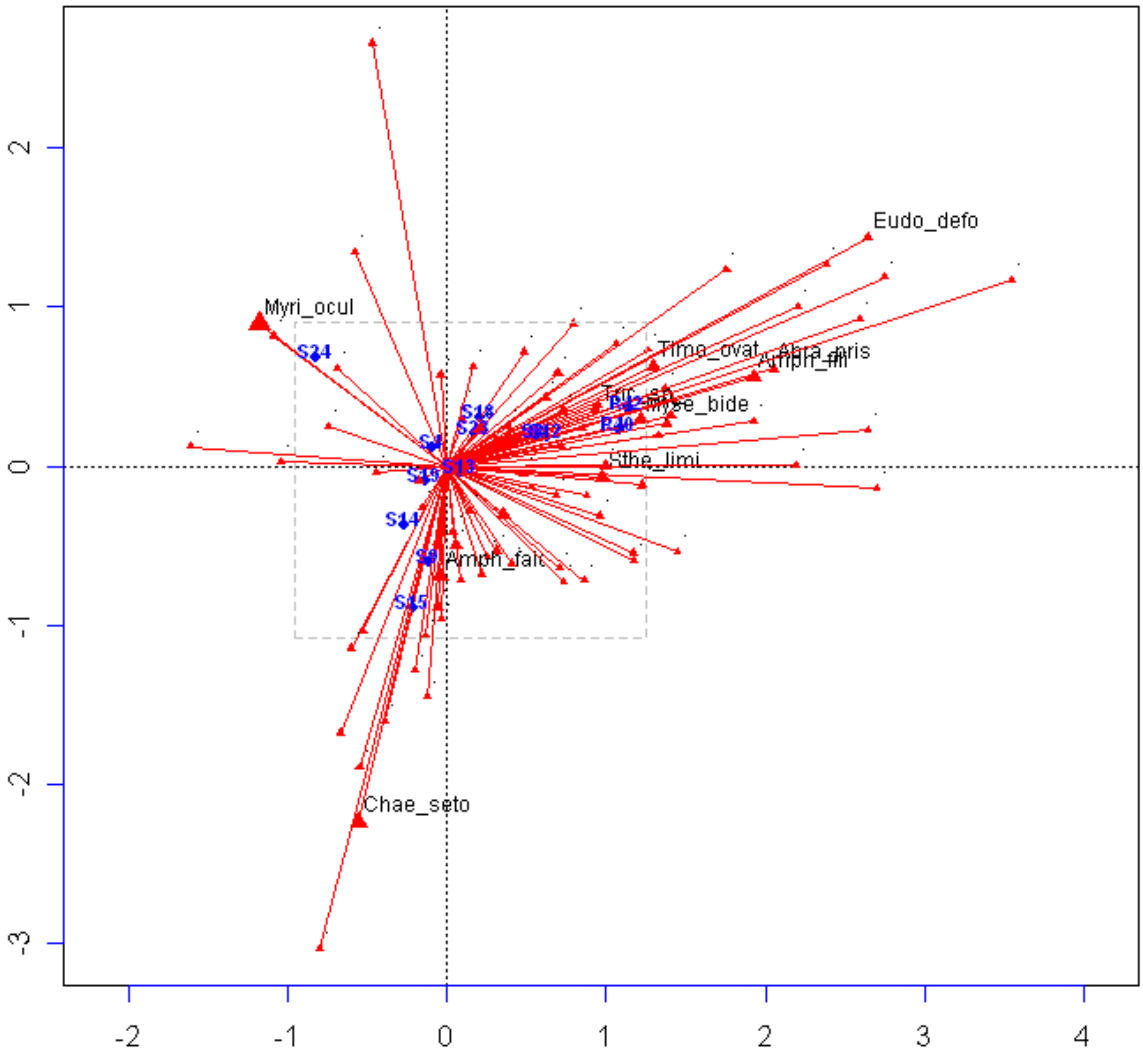

Figure 7: Standard PCA biplot of 'environ' data set of Table 2. Average sum-of-squares on the axes are 0.7008 and 0.1791 respectively (these are also proportions of variance accounted for, since total variance is scaled to be 1.

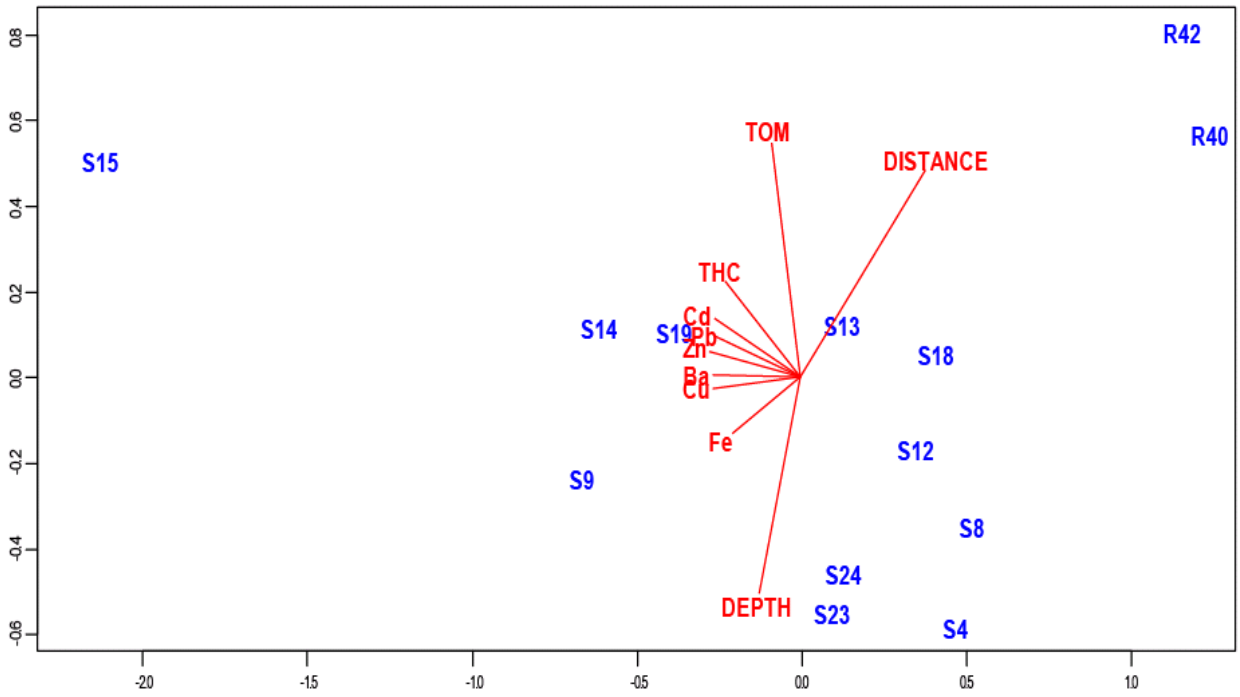

Figure 8: The standard PCA biplot of the log-transformed (and unstandardized) data of Table 2. Total variance (which is again the average of the individual variances) is 0.8692, and the percentages of variance explained on axes 1 and 2 are 84.5% and 11.0% respectively.

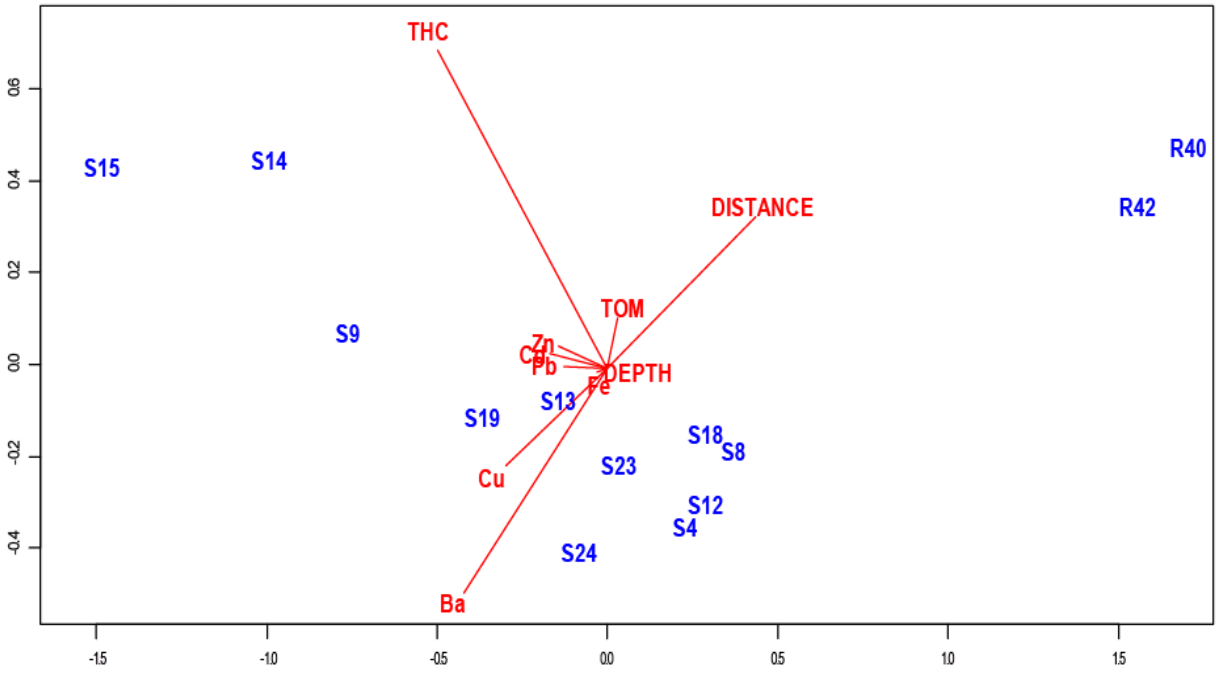

Figure 9: Dynamic transition between the asymmetric LRA map (weighted) and its standard biplot version, for the data set 'author'. In the first frame one can see the lining up of three letters, k, y and x, which indicates an equilibrium model for these letters (given in Greenacre and Lewi, 2009), but which cannot be diagnosed in the standard biplot. ***Click on the figure to see the animation.***

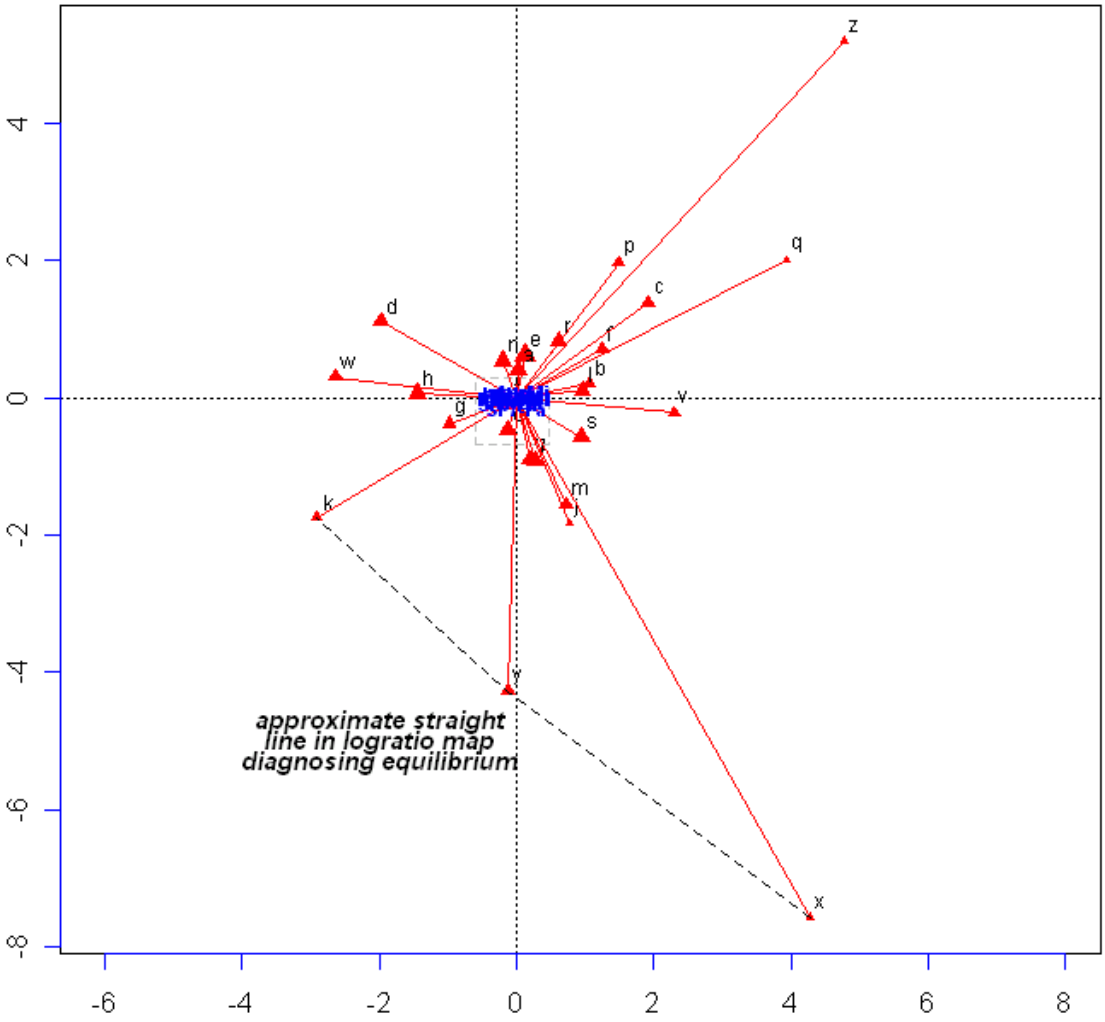

Figure 10: The standard MANOVA biplot of the three groups in the classic 'iris' data set. Total variance of the three group centroids in the Mahalanobis space is 8.119, of which 99.1% is accounted for by the first canonical axis. The variable Sepal Length has the lowest contribution to the discriminant space.