# Power transformations in correspondence analysis

Michael Greenacre

*Department of Economics and Business*

*Universitat Pompeu Fabra*

*Ramon Trias Fargas, 25-27*

*08005 Barcelona*

*SPAIN*

*E-mail:* michael@upf.es

**Abstract:** Power transformations of positive data tables, prior to applying the correspondence analysis algorithm, are shown to open up a family of methods with direct connections to the analysis of log-ratios. Two variations of this idea are illustrated. The first approach is simply to power the original data and perform a correspondence analysis – this method is shown to converge to unweighted log-ratio analysis as the power parameter tends to zero. The second approach is to apply the power transformation to the contingency ratios, that is, the values in the table relative to expected values based on the marginals – this method converges to weighted log-ratio analysis, or the spectral map. Two applications are described: first, a matrix of population genetic data which is inherently two-dimensional, and second, a larger cross-tabulation with higher dimensionality, from a linguistic analysis of several books.

**Keywords:** Box-Cox transformation, chi-square distance, contingency ratio, correspondence analysis, log-ratio analysis, power transformation, ratio data, singular value decomposition, spectral map.

---

This research has been supported by the Fundación BBVA, Madrid, Spain. Partial support of Ministry of Education and Science grant MEC-SEJ2006-14098 is also gratefully acknowledged. This paper has been conditionally accepted for publication in the special issue of Correspondence Analysis and Related Methods, in the journal *Computational Statistics & Data Analysis.*

## 1. Introduction

*Correspondence analysis* (CA) has found acceptance and application by a wide variety of researchers in different disciplines, notably the social and environmental sciences (for an up-to-date account, see Greenacre, 2007a). The method has also appeared in the major statistical software packages, for example SPSS, Minitab, Stata, SAS, Statistica and XLSTAT, and it is freely available in several implementations in R (R Development Core Team, 2007) – see, for example, Nenadić and Greenacre, 2007. The method is routinely applied to a table of non-negative data to obtain a spatial map of the important dimensions in the data, where proximities between points and other geometric features of the map indicate associations between rows, between columns and between rows and columns.

Methods based on log-ratios have quite different origins in the physical sciences, notably chemistry and geology, (Aitchison, 1983, 1986) and lead to maps where vectors between points depict the logarithms of the ratios between data values in the corresponding pairs of rows or columns. Interestingly, this *log-ratio analysis* (LRA), with the slight but highly significant adaptation of weighting the rows and columns of the table proportional to their marginal totals (exactly as is done in CA), has been used extensively for more than three decades in the pharmaceutical industry, originated by Lewi (1976). In this context it has been called the *spectral map* because it depicts the information from biological activity spectra. The spectral map, which we otherwise call weighted LRA to distinguish it from the unweighted form, can also be used to analyse contingency tables (see Lewi, 1998), in fact any ratio-scale data, as long as there are no zero values. Unweighted LRA has been treated in detail by Kazmierczak (1985) and its biplot and model-diagnostic properties have been investigated by Aitchison and Greenacre (2002). It is known that in spatial maps produced by LRA (unweighted or weighted), points that line up approximately as straight lines suggest equilibrium models in the rows or columns corresponding to those points (for example, see Aitchison and Greenacre, 2002; Greenacre and Lewi, 2005; Greenacre, 2007b). CA does not have this property, but has the advantage that it routinely handles data with zero values, which is one of the reasons why it is so popular in ecology and archaeology, where data tables are often quite large and sparse.

So the present situation is one of two competing methods, each with its particular advantages, and no apparent direct theoretical link between them, apart from the fact that both are based on singular value decompositions. It is known that CA and weighted LRA give very similar results if the variance in the table is low (this is a result of the approximation $\log(1+x) \approx x$ when $x$ is small), but differ when the variance is high – see Greenacre and Lewi (2005) or Cuadras, Cuadras and Greenacre (2006).

In the present paper we show that there is a much closer theoretical affinity between the two methods, in fact they both belong to the same family of methods defined by power transformations of either the original data or certain ratios calculated from the data. The power transformation family, as embodied in the Box-Cox transformation (Box and Cox, 1964), is usually used in statistics to symmetrize the distribution of a response variable in a regression model to satisfy the model assumptions (Hinkley, 1975). In the analysis of frequency data, assuming the counts follow a Poisson distribution, the square root transformation is used to stabilize the variance (Bartlett, 1936). In ecology abundance data is almost always highly over-dispersed and a particular school of ecologists routinely applies a fourth-root transformation before proceeding with statistical analysis (Field, Clarke & Warwick, 1982). Here we study the family of power transformations in the context of correspondence analysis (CA). Some special cases emerge, notably the spectral map, which is a limiting case as the power transformation parameter tends to zero.

The main result in this paper is thanks to the Box-Cox transformation $f(x) = (1/\alpha)\,(x^{\alpha} - 1)$, which converges to $\log(x)$ as $\alpha$ tends to 0. Because we are making a comparison with LRA, only strictly positive data will be considered. In Section 2 we give two equivalent definitions of CA and show how power transformations can generate two respective families of methods, the first giving a direct link between CA and unweighted LRA, and the second a direct link between CA and weighted LRA. Properties of these families are illustrated in Section 3 using two examples, a data matrix from population genetics with high inherent variance, and a linguistic example with very low variance. Sections 4 and 5 treat related topics and literature, and Section 6 concludes with a discussion. R code that permits dynamic viewing of the smooth transition from a CA map to a LRA map (weighted or unweighted) is available for download from `http://www.carme-n.org` and videos showing the results of executing this code can be viewed at `http://www.econ.upf.edu/~michael/videos`. In this article we can merely show some static “snapshots” of this transition.

## 2. Power families: from correspondence analysis to log-ratio analysis

CA and LRA are two of the many multivariate methods based on the singular value decomposition (SVD) (see, for example, Greenacre 1984: chapter 3). In the geometric interpretation of the SVD, the rows and/or columns of the data matrix define points in a multidimensional space and the SVD identifies subspaces of low dimensionality which capture maximum sum-of-squares in the data. Different weights for the rows and columns can be introduced into this scheme so that weighted sum-of-squares is decomposed. The weighting can be considered either as assigning different weights to each point, or as a change of the Euclidean metric to a weighted one, or both of these at the same time, as is the case in CA. To establish notation, the following subsection contains a summary of standard theory to obtain the *principal coordinates* of the row and columns points in a so-called *symmetric CA map* (for more details, see Greenacre, 2007a).

### 2.1 Correspondence analysis

Suppose that **N** is an $I \times J$ table of non-negative data. First divide **N** by its grand total $n$ to obtain the so-called *correspondence matrix* $\mathbf{P} = (1/n)\, \mathbf{N}$. Let the row and column marginal totals of **P** be the vectors **r** and **c** respectively – these are the weights, or *masses*, associated with the rows and columns. Let $\mathbf{D}_r$ and $\mathbf{D}_c$ be the diagonal matrices of these masses. The computational algorithm to obtain coordinates of the row and column profiles with respect to principal axes, using the SVD, is as follows:

*Correspondence analysis*

1. Calculate the matrix of standardized residuals:

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

2. Calculate the SVD: $\mathbf{S} = \mathbf{U}\mathbf{D}_\sigma\mathbf{V}^{\mathsf{T}}$ where $\mathbf{U}^{\mathsf{T}}\mathbf{U} = \mathbf{V}^{\mathsf{T}}\mathbf{V} = \mathbf{I}$ (2)

3. Principal coordinates of rows: $\mathbf{F} = \mathbf{D}_r^{-1/2}\mathbf{U}\mathbf{D}_\sigma$ (3)

4. Principal coordinates of columns: $\mathbf{G} = \mathbf{D}_c^{-1/2}\mathbf{V}\mathbf{D}_\sigma$ (4)

The rows of the coordinate matrices in (3) and (4) refer to the rows or columns, as the case may be, of the original table, while the columns of these matrices refer to the principal axes, or dimensions, of the solution. The sum of squares of the decomposed matrix **S** is a quantity called the *total inertia*, or simply inertia, of the data table:

$$\text{inertia} = \phi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(p_{ij} - r_i c_j)^2}{r_i c_j} = \sum_{i=1}^{I}\sum_{j=1}^{J} r_i c_j \left(\frac{p_{ij}}{r_i c_j} - 1\right)^2 \tag{5}$$

The inertia is exactly Pearson's mean-square contingency coefficient, that is, the Pearson chi-square statistic for the table divided by the grand total $n$ of the table, and is used as a measure of total variance. The squared singular values $\sigma_k^2$ decompose the inertia, and the row and column principal coordinates are scaled in such a way that $\mathbf{F}^{\mathsf{T}}\mathbf{D}_r\mathbf{F} = \mathbf{G}^{\mathsf{T}}\mathbf{D}_c\mathbf{G} = \mathbf{D}_\sigma^2$, i.e. the weighted sum-of-squares of the coordinates on the $k$-th dimension (or their inertia in the direction of this dimension) is equal to $\sigma_k^2$, called the *principal inertia* (or eigenvalue) on dimension $k$. A two-dimensional solution, say, would use the first two columns of the coordinate matrices, and the explained inertia accounted for in the two-dimensional solution is the sum of the first two terms $\sigma_1^2+\sigma_2^2$, usually expressed as percentages of the total inertia. Standard coordinates are defined as in (3) and (4) without scaling on the right by the singular values $\mathbf{D}_\sigma$, and hence have weighted sum-of-squares equal to 1.

Notice in (5) how the inertia can be defined using either *contingency differences* between observed and expected relative frequencies, $p_{ij} - r_i c_j$, or *contingency ratios*, $p_{ij} / r_i c_j$. The matrix **S** in (1) can be written equivalently as follows, in terms of the matrix of contingency ratios $\mathbf{Q} = \mathbf{D}_r^{-1}\mathbf{P}\mathbf{D}_c^{-1}$:

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

where **1** denotes a vector of ones of appropriate order in each case. The pre- and post-multiplication of **Q** by the centring matrices $(\mathbf{I} - \mathbf{1}\mathbf{r}^\mathsf{T})$ and $(\mathbf{I} - \mathbf{1}\mathbf{c}^\mathsf{T})^\mathsf{T}$ amounts to a weighted double-centring of the contingency ratios. This second definition of CA is particularly useful for comparing with LRA.

### 2.2 Logratio analysis, weighted and unweighted

A weighted LRA (i.e., spectral map) is based on the logarithms of the elements of **N**: $\mathbf{L} = [\log(n_{ij})]$; hence we only consider strictly positive data here. Using the same masses **r** and **c** as in CA, the matrix is then double-centred, and then a weighted SVD is performed, as summarized in the following computational scheme:

*Weighted log-ratio analysis (spectral map)*

1. Calculate the weighted, double-centred matrix:

$$\mathbf{S}^* = \mathbf{D}_r^{1/2}(\mathbf{I} - \mathbf{1}\mathbf{r}^\mathsf{T})\mathbf{L}(\mathbf{I} - \mathbf{1}\mathbf{c}^\mathsf{T})^\mathsf{T}\,\mathbf{D}_c^{1/2} \tag{7}$$

2. Calculate the SVD: $\mathbf{S}^* = \mathbf{U}\mathbf{D}_\mu\mathbf{V}^\mathsf{T}$ where $\mathbf{U}^\mathsf{T}\mathbf{U} = \mathbf{V}^\mathsf{T}\mathbf{V} = \mathbf{I}$ (8)

3. Principal coordinates of rows: $\mathbf{F} = \mathbf{D}_r^{-1/2}\mathbf{U}\mathbf{D}_\mu$ (9)

4. Principal coordinates of columns: $\mathbf{G} = \mathbf{D}_c^{-1/2}\mathbf{V}\mathbf{D}_\mu$ (10)

Notice that steps (8) – (10) are identical to (2) – (4) of CA. It is just the pre-processing and first step (7) that differs. The unweighted LRA is obtained simply by setting $\mathbf{r} = (1/I)\mathbf{1}$ and $\mathbf{c} = (1/J)\mathbf{1}$ in the above scheme, so that the initial matrix **S*** is replaced by

$$\mathbf{S}^\circ = (IJ)^{-1/2}(\mathbf{I} - (1/I)\mathbf{1}\mathbf{1}^\mathsf{T})\mathbf{L}(\mathbf{I} - (1/J)\mathbf{1}\mathbf{1}^\mathsf{T}) \tag{11}$$

Since the logarithm of the contingency ratios is $\log(n_{ij}) - \log(n) - \log(r_i) - \log(c_j)$, and the double-centring removes the "constant" $\log(n)$ and "main effects" $\log(r_i)$ and $\log(c_j)$, the only difference between the initial matrices **S** and **S*** is that in (6) CA operates on the contingency ratios whereas in (7) weighted LRA operates on the log-transformed contingency ratios.

The total variance in weighted LRA (i.e., the sum of squares of matrix **S*** in (7)) can be written in terms of the logarithms of the "double-ratios":

$$\sum\sum_{i<i'}\sum\sum_{j<j'} r_i r_{i'} c_j c_{j'} \left( \log \frac{n_{ij}}{n_{i'j}} \frac{n_{i'j'}}{n_{ij'}} \right)^2 \tag{12}$$

For the unweighted LRA, again replace the row masses by $(1/I)$ and the column masses by $(1/J)$.

### 2.3 Power families of analyses generated by power transformations

The two forms of CA starting from the correspondence matrix in (1) or the contingency ratios in (6) suggest two ways of introducing a power transformation.

***Power family 1:*** Pre-transform the matrix **P** (or, equivalently **N**), by the power transformation $p_{ij}(\alpha) = p_{ij}^{\ \alpha}$. After dividing out this matrix by its total to obtain the new correspondence matrix and recalculating the row and column masses, proceed as in (1) to calculate the matrix to be decomposed, denoted by $\mathbf{S}(\alpha)$, and then continue as in (2) – (4) above. To standardize the analyses with different values of the power parameter $\alpha$ the singular values $\sigma_k$ are divided by $\alpha$, so the inertia is divided by $\alpha^2$ – this is equivalent to dividing $\mathbf{S}(\alpha)$ by $\alpha$ before applying the SVD.

***Power family 2:*** Pre-transform the matrix **Q** of contingency ratios by the power transformation $q_{ij}(\alpha) = q_{ij}^{\ \alpha}$. Calculate $\mathbf{S}^*(\alpha)$ using the power-transformed contingency ratios, as in (6), followed by (2) – (4). In this case the masses $r_i$ and $c_j$ are maintained constant throughout, equal to their original values irrespective of $\alpha$. Again, to standardize the analyses with different values of the power parameter $\alpha$, the singular values $\sigma_k$ are divided by $\alpha$, so the inertia is divided by $\alpha^2$ – this is equivalent to dividing $\mathbf{S}^*(\alpha)$ by $\alpha$ before applying the SVD, or to dividing the power-transformed contingency ratios $q_{ij}(\alpha)$ by $\alpha$ before double-centring and decomposing.

In power family 2, whether we double-centre $(1/\alpha)\, q_{ij}^{\ \alpha}$ or $(1/\alpha)\,(q_{ij}^{\ \alpha}–1)$ makes no difference at all, because the constant term will be removed. Hence, the analysis in this case amounts to the Box-Cox transformation of the contingency ratios:

$$\frac{1}{\alpha}\left(q_{ij}^{\alpha} - 1\right) \tag{13}$$

which converges to $\log(q_{ij})$ as $\alpha \to 0$. This shows that power family 2 converges to weighted LRA as $\alpha \to 0$.

In power family 1, we are also analysing contingency ratios of the form $(1/\alpha)\, q_{ij}{}^{\alpha}$, or $(1/\alpha)\,(q_{ij}{}^{\alpha} - 1)$, but then the ratios as well as the weights and double-centring are all with respect to row and column masses that are changing with $\alpha$. At the limit as $\alpha \to 0$, these masses tend to constant values, i.e. $1/I$ for the rows and $1/J$ for the columns; hence this shows that the limiting case of power family 1 is the analysis of the logarithms with constant masses, or unweighted LRA.

## 3. Applications

### 3.1 Two-dimensional example: the M-N system in population genetics

If the data are inherently two-dimensional then there will be little difference in the unweighted and weighted LRA solutions, just a slight rotation of the principal axes, so this serves as a good demonstration of the difference between the CA and LRA configurations. This is the case with the data set in Table 1 from population genetics, concerning the estimated frequencies in 24 populations of three groups in the M-N genetic system. The two alleles, M and N, in this system are co-dominant, so that the three groups are MN, M (denoting MM) and N (denoting NN).

Figure 1 shows the transition in power family 2, with fixed masses, from CA ($\alpha = 1$) to weighted LRA (limit as $\alpha \to 0$, i.e. log-transformation) in three intermediate steps: $\alpha$=0.75, $\alpha$=0.50 and $\alpha$=0.25 (using the R code referred to at the end of Section 1, one can see dynamically a smooth change from CA to LRA, using smaller steps, for example $\alpha$=0.99, 0.98, …, 0.02, 0.01, $\to$0). This example is interesting because the CA solution shows the well-known arch effect, with 86.4% inertia on the first axis, and thus 15.6% on the second. As $\alpha$ descends the curve starts to straighten out until at the limit of the weighted LRA, the configuration is practically one-dimensional with 96.8% explained inertia on the first principal axis (the inset boxes show the evolution of the total inertia, depicted by the upper curve, and the two principal inertias shown by the two lower curves, as $\alpha$ descends from 1 to 0).

The linearity of M, MN and N in the final weighted LRA and the almost equal distance between the three points imply a model for the logratios: log(MN/M) = log(N/MN) + constant, which perfectly diagnoses the Hardy-Weinberg equilibrium for this genetic system: $\mathrm{MN}^2 / \mathrm{M{\cdot}N} = 4$ (see, for example, Greenacre, 2007b).

The result for power family 1, with changing masses, is almost identical in this two-dimensional case, the only noticeable difference being the way the total inertia and the parts of inertia are measured, since the limiting case as $\alpha \rightarrow 0$ is the unweighted LRA, where the percentage of inertia explained by the first axis is slightly higher, 97.2%.

### 3.2 Higher-dimensional example: the "author" data

The data set "author" consists of counts of the letters *a* to *z* in samples of texts from 12 books (or chapters of books) by six famous English authors (Table 2). This data set has an extremely low inertia, since there are very small differences in the relative frequencies of the letters, but the differences between authors is still substantively meaningful (for more detailed analyses of this data set, see Greenacre and Lewi, 2005; Greenacre, 2007a: Chapter 10). There is one zero value in this table (a count of zero occurrences of the letter *q* in the sample of text from *Farewell to Arms* by Hemingway), which we have replaced by a ½, otherwise LRA breaks down. It is already known that CA and LRA will resemble one another when the inertia is low (Greenacre and Lewi, 2005; Cuadras *et al.*, 2006). Figure 2 shows CA in the first panel, weighted LRA in the last panel and the analysis of the power-transformed contingency ratios with $\alpha = 0.50$ in the middle panel. The differences between the configurations of the books are minor, as expected, and the cumulated percentage of inertia explained by the first two axes is slightly lower in the LRA map. The benefit of the LRA approach is that letters that form straight lines indicate linear models in the corresponding log-ratios. For example, as shown by Greenacre and Lewi (2005), the straight line formed by *k*, *y* and *x* in the last panel of Figure 2 indicates an equilibrium relationship between these three letters which amounts to: $y \propto x^{0.2} k^{0.8}$. In the CA map (first panel of Figure 2) such relationships can not be diagnosed.

## 4. Connection with Hellinger distance and spherical analysis

Escofier (1978) studied the properties of the Euclidean distance defined on the square-root transformed profile values, called Hellinger distances (see, for example, Rao, 1995). Domenges and Volle (1979) called the principal component analysis of such transformed data *analyse sphérique* (spherical analysis), because the square-root transformation places the profile points on a hypersphere. Cuadras *et al.* (2006) have studied the connection between CA and a slightly different form called "Hellinger analysis", which differs in the way the transformed data are centred. These variants can also be thought of as a power-transformed family if we start from the following equivalent form of the matrix **S** in (1) or (6), in terms of *row profiles* (the rows of **P** divided by their row sums, i.e., the rows of $\mathbf{D}_r^{-1}\mathbf{P}$):

$$\mathbf{S} = \mathbf{D}_r^{1/2}(\mathbf{D}_r^{-1}\mathbf{P} - \mathbf{1}\mathbf{c}^\mathsf{T})\mathbf{D}_c^{-1/2} \tag{14}$$

Hellinger analysis is based on the SVD of the matrix:

$$\tilde{\mathbf{S}} = \mathbf{D}_r^{1/2}[(\mathbf{D}_r^{-1}\mathbf{P})^{1/2} - (\mathbf{1c}^\mathsf{T})^{1/2}] \tag{15}$$

which can be written as:

$$\tilde{\mathbf{S}} = \mathbf{D}_r^{1/2}[(\mathbf{D}_r^{-1}\mathbf{P}\mathbf{D}_c^{-1})^{1/2} - \mathbf{11}^\mathsf{T}]\mathbf{D}_c^{1/2}$$

This suggests another family based on the power transformation of the contingency ratios $\mathbf{D}_r^{-1}\mathbf{P}\,\mathbf{D}_c^{-1}$:

$$\tilde{\mathbf{S}}(\alpha) = \mathbf{D}_r^{1/2}[(\mathbf{D}_r^{-1}\mathbf{P}\mathbf{D}_c^{-1})^{\alpha} - \mathbf{11}^\mathsf{T}]\mathbf{D}_c^{1/2} \tag{16}$$

(again, we would multiply this matrix by $1/\alpha$ before decomposing with the SVD). This family passes smoothly from CA to Hellinger analysis as $\alpha$ changes value from 1 to 0.5 (Cuadras and Cuadras, 2007). In spherical analysis the square-root transformed profiles in (15) are centred with respect to their weighted average $\mathbf{r}^\mathsf{T}(\mathbf{D}_r^{-1}\mathbf{P})^{1/2}$, and so this variant would be a special case of weighted PCA of power-transformed profiles, centred in the usual way and weighted by the row masses. Neither Hellinger analysis nor spherical analysis seems to have any practical benefit over CA or LRA, apart from the claimed advantage that the metric between the rows does not depend on the column margins, as is the case in CA. Figure 3 shows the M-N example for this family with $\alpha$ = 1, 0.75, 0.5. There is hardly any change in the row configuration and the percentage of inertia on the first dimension, after an initial increase, is less in Hellinger analysis. This data set is two-dimensional in CA and LRA and in both power families described in Section 3, but is three-dimensional in the case of Hellinger and spherical analyses and the power families described above that lead to them (apart from the case $\alpha$ = 1, which is CA and thus two-dimensional). The introduction of a third dimension could be deemed a disadvantage because a size effect has now been mixed in with the analysis, whereas CA and LRA concentrate only on shape effects.

Bavaud (2002, 2004) looks at families of dissimilarity measures based on the contingency ratios $q_{ij}$, defined, for example, between rows as:

$$\sum_j c_j \left(f(q_{ij}) - f(q_{i'j})\right)^2 \qquad \text{where} \quad f(q) = \frac{1}{\alpha}(q^\alpha - 1) \tag{17}$$

for which $\alpha$ = 1 gives the chi-square distance, $\alpha$ = ½ gives the Hellinger distance, and the limit as $\alpha$ tends to 0 gives the following weighted distance based on the logarithms of the row profiles: $\sum_j c_j \left(\ln(p_{ij}/r_i) - \ln(p_{i'j}/r_{i'})\right)^2$. Notice that this distance function is similar but not the same as the one implicit in weighted LRA, which divides elements $p_{ij}$ in each row by their respective weighted geometric mean $p_{i1}^{c_1} p_{i2}^{c_2} \cdots p_{iJ}^{c_J}$ , not by their sum $r_i$.

## 5. More relations between methods

The same idea can be applied to many other methods related to CA, such as multidimensional scaling (MDS) and so-called "non-symmetrical correspondence analysis" (NSCA) (Lauro and D'Ambra , 1984; Kroonenberg and Lombardo, 1999).

NSCA is a principal component analysis of profile vectors, using the profile masses as weights, in other words the same as spherical analysis described in Section 4, but without the square-root transformation. As in spherical analysis, the rows and columns are treated differently, depending on whether the data are considered as predicting the rows given the columns, or the columns given the rows. For example, in the latter case:

*Non-symmetrical correspondence analysis for predicting columns, given rows*

1. Calculate the matrix:

$$\breve{\mathbf{S}} = \mathbf{D}_r^{1/2}(\mathbf{D}_r^{-1}\mathbf{P} - \mathbf{1}\mathbf{c}^\mathsf{T}) \tag{18}$$

2. Calculate the SVD: $\breve{\mathbf{S}} = \mathbf{U}\mathbf{D}_\sigma\mathbf{V}^\mathsf{T}$ where $\mathbf{U}^\mathsf{T}\mathbf{U} = \mathbf{V}^\mathsf{T}\mathbf{V} = \mathbf{I}$ (19)

3. Principal coordinates of rows: $\mathbf{F} = \mathbf{D}_r^{-1/2}\mathbf{U}\mathbf{D}_\sigma$ (20)

4. Principal coordinates of columns: $\mathbf{G} = \mathbf{V}\mathbf{D}_\sigma$ (21)

Compare (18) with (15) – the only difference is that the square-root transformation of profiles is omitted.

Various power versions can be considered, depending on what is transformed and how centring is performed:

(i) power up the original data, $p_{ij}^{\alpha}$ , in which case the row masses will change according to $\alpha$;

(ii) power up the profiles, $(p_{ij} / r_i)^{\alpha}$ , and average profile, $c_j^{\alpha}$ , keeping the row masses equal to the original ones for all $\alpha$ – this version has Hellinger analysis as a special case when $\alpha$ = ½ .

(iii) power up the profiles, $(p_{ij}/r_i)^{\alpha}$, and centre them at their weighted average, which has elements $\sum_i r_i (p_{ij}/r_i)^{\alpha}$, $j = 1,\ldots, J$ – this version has spherical analysis as a special case when $\alpha = ½$.

Siciliano (1989) introduced a logarithmic transformation into NSCA – this corresponds to the limiting case of version (iii) above as $\alpha$ tends to 0.

In order to relate NSCA to CA, compare (18) with (14) – the only difference is that post-multiplication by $\mathbf{D}_c^{-½}$ is omitted. To make a direct comparison with CA, an equal weighting of $1/J$ should be introduced for the columns, i.e., $\mathbf{D}_c^{-½}$ in the CA formulation (14) should be replaced by $(1/J)^{-1/2} = J^{1/2}$. We can then illustrate graphically the difference between CA and NSCA by incorporating a parameter, $\beta$ say, which allows a transition from one weighting system to another. For example, let $\mathbf{D}_w = \beta\mathbf{D}_c + (1-\beta)(1/J)\mathbf{I}$ and replace steps (18) and (21) above by, respectively:

$$\text{Matrix to be decomposed} \qquad \breve{\mathbf{S}} = \mathbf{D}_r^{1/2}(\mathbf{D}_r^{-1}\mathbf{P} - \mathbf{1c}^{\mathsf{T}})\mathbf{D}_w^{-1/2} \tag{22}$$

$$\text{Principal coordinates of columns:} \quad \mathbf{G} = \mathbf{D}_w^{-1/2}\mathbf{V}\mathbf{D}_\sigma \tag{23}$$

As $\beta$ varies from 1 to 0 the resulting maps will pass smoothly from CA to NSCA respectively, where the equal column weighting of $1/J$ has been introduced into the NSCA definition. Figure 4 shows three snapshots of the transition – since the row masses are approximately equal there is very little change in the configurations and percentages of inertia, only an increase in the inertias for the non-symmetrical version.

The same idea can be used to compare CA with PCA in terms of their respective standardizations of the matrix columns, say, where CA standardizes by the square root of the mean and PCA by the standard deviation. This would make sense if the two methods were analysing comparable equal-weighted rows, for example if the rows add up to 1 for data that are proportions (or percentages adding up to 100%) so that the profiles were the original data and all rows received the same mass. As before, the standardization could be defined parametrically as post-multiplication of the data matrix by $\gamma\mathbf{D}_c^{-½} + (1-\gamma)\,\mathbf{D}_s^{-1}$, where the columns masses (means in this case) are in the diagonal of $\mathbf{D}_c$ and the column standard deviations are in the diagonal of $\mathbf{D}_s$. Hence, as $\gamma$ varies from 1 to 0, the resulting maps pass smoothly from CA to PCA.

In MDS we are trying to match observed distances $d_{ij}$ with fitted distances $\delta_{ij}$ in a map. To reduce the influence of large distances in the fitting process, a power transformation can be introduced, for example:

$$d_{ij}^*(\alpha) = \frac{1}{\alpha}\left((1+d_{ij})^{\alpha} - 1\right) \tag{24}$$

This starts with the original distances when $\alpha = 1$ and converges to a logarithmic transformation $\log(1+d_{ij})$ as $\alpha \to 0$.

Carroll, Kumbasar and Romney (1997) showed a different connection between CA and MDS that is not governed by a power transformation but is a limiting result in the same spirit as those presented here. Their result was that the CA of a suitably transformed distance matrix has as a limiting case classical scaling. We give Carroll *et al.*'s result in our present notation. Suppose $[d_{ii'}]$ is an $I \times I$ square matrix of observed distances, and define a new table as follows:

$$n_{ii'} = \frac{1}{\alpha} - d_{ii'}^2 \tag{25}$$

where $\alpha > 0$ and $1/\alpha \geq \max\{d_{ii'}^2\}$, i.e., squared distances are subtracted from a number at least as large as their maximum so that the $n_{ii'}$ are all nonnegative. Then the CA of the matrix $\mathbf{N} = [n_{ii'}]$ converges to the classical scaling solution as $\alpha \to 0$. As in all cases above, a rescaling needs to be introduced to make the solutions equivalent. In the case of CA, we perform steps (1) and (2) on the correspondence matrix $\mathbf{P}$ based on (25) and then the solution coordinates are:

$$\mathbf{H} = \mathbf{D}_r^{-1/2}\mathbf{U}\mathbf{D}_\sigma^{1/2}\frac{1}{2\alpha} \tag{26}$$

Hence $\mathbf{H}$ consists of the standard coordinates $\mathbf{D}_r^{-1/2}\mathbf{U}$ scaled by the *square roots* of the singular values (i.e., the fourth roots of the inertias[†] in the CA of $\mathbf{N}$), then rescaled by dividing by $2\alpha$. The eigenvalues of the classical scaling can be recovered from $(I/2\alpha)\sigma_k$, remembering that all these results apply in the limit – in practice, an $\alpha$ about one thousandth of the maximum of the $d_{ii'}^2$, i.e. $1/\alpha$ about 1000 times this maximum, gives a solution very close to the classical scaling one.

## 6. Discussion and conclusion

We have shown that CA and both unweighted and weighted LRA can be connected by considering the power transformation of the original data matrix or the matrix of contingency ratios respectively. When the power parameter $\alpha$ is equal to 1 we have simple CA in both cases, and as $\alpha$ tends to 0 we

[†] The distinction between singular values, eigenvalues and inertias becomes a bit confusing in this case where $\mathbf{N}$ is a square matrix. The singular values of $\mathbf{N}$ are actually eigenvalues (at least those corresponding to positive eigenvalues), and the inertias in the CA of $\mathbf{N}$ (often themselves referred to as eigenvalues) are the squares of the singular values of $\mathbf{N}$. Hence these inertias are fourth powers of $\mathbf{N}$'s eigenvalues.

obtain the unweighted or weighted cases respectively. This shows that LRA is theoretically part of the same family as CA, and not as different as one might have thought. The connection is especially surprising for CA and the spectral map (weighted LRA) because the two methods have been developed and applied extensively for over 30 years as completely separate methodologies.

The idea of linking methods by a parameter and especially the dynamic visualization of smooth changes from one method to another can be highly enlightening as to the properties of these methods. Various other methods can be linked to CA in this way, as we have shown: CA to spherical analysis and Hellinger analysis, CA to NSCA, CA to PCA and CA to MDS. Unfortunately, in these pages we can only show "snapshots" of some steps between the methods for selected values of the power parameter, but the R code given on the site `www.carme-n.org` can be used to get an idea of the dynamic graphics possibilities, and is easily adapted to the other cases described above.

## 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.

Bartlett, M. S. (1936). The square root transformation in analysis of variance. S*upplement to the Journal of the Royal Statistical Society*, **3**, 68–78.

Bavaud, F. (2002). Quotient dissimilarities, Euclidean embeddability, and Huygens' weak principle. In H. A. L. Kiers *et al.*(eds), *Data Analysis, Classification and Related Methods*, pp. 194–195. Heidelberg: Springer.

Bavaud, F. (2004). Generalized factor analyses for contingency tables. In D. Banks *et al.*(eds), *Classification, Clustering and Data Mining Applications*, pp. 597–606. Heidelberg: Springer.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of Royal Statistical Society, Series B*, **35**, 473–479.

Carroll, J. D., Kumbasar, E. and Romney, A. K. (1997). An equivalence relation between correspondence analysis and classical metric multidimensional scaling for the recovery of Euclidean distances. *British Journal of Mathematical and Statistical Psychology*, **50**, 81–92.

Cuadras, C. and Cuadras, D. (2006). A parametric approach to correspondence analysis. *Linear Algebra and its Applications*, **417**, 64–74.

Cuadras, C, Cuadras, D. and Greenacre, M. J. (2005). A comparison of methods for analyzing contingency tables. *Communications in Statistics – Simulation and Computation*, **35**, 447–459.

Domenges, D. and Volle, M. (1979). Analyse factorielle sphérique: une exploration. *Annales de l'INSEE*, **35**, 3-84.

Escofier, B. (1978). Analyse factorielle et distances répondant au principe d'équivalence distributionnelle. *Revue de Statistique Appliquée*, **26**, p. 29-37

Field, J. G., Clarke, K. R. and Warwick, R. M. (1982). A practical strategy for analysing multispecies distribution patterns. *Marine Ecology Progress Series*, **8**, 37–52.

Greenacre, M. J. (2007a). *Correspondence Analysis in Practice. Second Edition.* London: Chapman & Hall / CRC Press.

Greenacre, M. J. (2007b). Diagnosing models from maps based on weighted logratio analysis. In J. del Castillo, A.Espinal and P. Puig (eds), *Proceedings of the 22nd International Workshop on Statistical Modelling.*. Barcelona: Institut d'Estadistica de Catalunya.
URL `http://XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX`

Greenacre, M. J. and Lewi, P. J. (2005). Distributional equivalence and subcompositional coherence in the analysis of contingency tables, ratio scale measurements and compositional data. Economics Working Paper 908, Universitat Pompeu Fabra.
URL `http://www.econ.upf.edu/en/research/onepaper.php?id=908`
Revised version accepted for publication in *Journal of Classification* (2008).

Hinkley, D. (1975). On power transformations to symmetry. *Biometrika*, **62**, 101–111.

Kazmierczak J.-B. (1985). Analyse logarithmique : deux exemples d'application. *Revue de Statistique Appliquée*, **33**, pp 13-24.

Kroonenberg, P. M. and Lombardo, R. (1999). Nonsymmetric correspondence analysis: a tool for analysing contingency tables with a dependence structure. *Multivariate Behavioral Research*, **34**, 367–396.

Lauro N.C. and D'Ambra L. (1984). Non-symmetrical correspondence analysis. In E. Diday et al. (eds), *Data Analysis and Informatics – III*, pp. 433-446. Amsterdam: North Holland.

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

Lewi, P. J. (1998). Analysis of contingency tables. In B. G. M. Vandeginste, D. L. Massart, L. M. C. Buydens, S. de Jong, P. J. Lewi, J. Smeyers-Verbeke (eds), *Handbook of Chemometrics and Qualimetrics: Part B*, Chapter 32, pp. 161–206. Amsterdam: Elsevier.

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/`

R Development Core Team (2005). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org`

Rao, C. R. (1995). A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance. *Qüestiió*, **19**, 23–63.

Siciliano R. (1989). Non-symmetrical logarithmic analysis for contingency tables. In A. Decarli *et al.* (eds), *Statistical Modelling, Proceedings of GLIM 89 and the 4th International Workshop on Statistical Modelling*, pp. 278-285. Berlin: Springer-Verlag.

***Table 1*** Data set “M-N”: estimated proportions of three genetic groups of the M-N system, with two co-dominant alleles M and N.

| Population | MN | M | N |
|---|---|---|---|
| 1 | 0.12 | 0.01 | 0.87 |
| 2 | 0.19 | 0.02 | 0.79 |
| 3 | 0.37 | 0.05 | 0.58 |
| 4 | 0.39 | 0.08 | 0.53 |
| 5 | 0.41 | 0.02 | 0.57 |
| 6 | 0.50 | 0.25 | 0.25 |
| 7 | 0.52 | 0.25 | 0.23 |
| 8 | 0.51 | 0.31 | 0.18 |
| 9 | 0.50 | 0.31 | 0.19 |
| 10 | 0.49 | 0.27 | 0.24 |
| 11 | 0.50 | 0.28 | 0.22 |
| 12 | 0.49 | 0.35 | 0.16 |
| 13 | 0.47 | 0.43 | 0.10 |
| 14 | 0.44 | 0.47 | 0.09 |
| 15 | 0.40 | 0.51 | 0.09 |
| 16 | 0.42 | 0.51 | 0.07 |
| 17 | 0.39 | 0.53 | 0.08 |
| 18 | 0.39 | 0.59 | 0.02 |
| 19 | 0.15 | 0.79 | 0.06 |
| 20 | 0.15 | 0.83 | 0.02 |
| 21 | 0.36 | 0.61 | 0.03 |
| 22 | 0.34 | 0.61 | 0.05 |
| 23 | 0.30 | 0.68 | 0.02 |
| 24 | 0.28 | 0.67 | 0.05 |

***Table 2*** Books from which text is sampled for the “author” data, and abbreviations used in Figure 2.

| | | | |
|---|---|---|---|
| **TD-Bu** | Three Daughters (Buck) | **FA-He** | Farewell to Arms (Hemingway) |
| **EW-Bu** | East Wind (Buck) | **Is-He** | Islands (Hemingway) |
| **Dr-Mi** | The Drifters (Michener) | **SF6-Fa** | Sound and Fury, ch.6 (Faulkner) |
| **As-Mi** | Asia (Michener) | **SF7-Fa** | Sound and Fury, ch.7 (Faulkner) |
| **LW-Cl** | Lost World (Clark) | **Pe2-Ho** | Pendorric, ch.2 (Holt) |
| **PF-Cl** | Profiles of the Future (Clark) | **Pe3-Ho** | Pendorric, ch.3 (Holt) |

*Figure 1*: From correspondence analysis ($\alpha = 1$) to weighted log-ratio analysis ($\alpha \rightarrow 0$), with three intermediate steps, for the "M-N" data, showing the symmetric maps (both rows and columns in principal coordinates). The box shows the numerical value of $\alpha$ and the percentage of inertia explained on the first dimension, as well as a graph of the values of the total inertia and two principal inertias as $\alpha$ decreases.

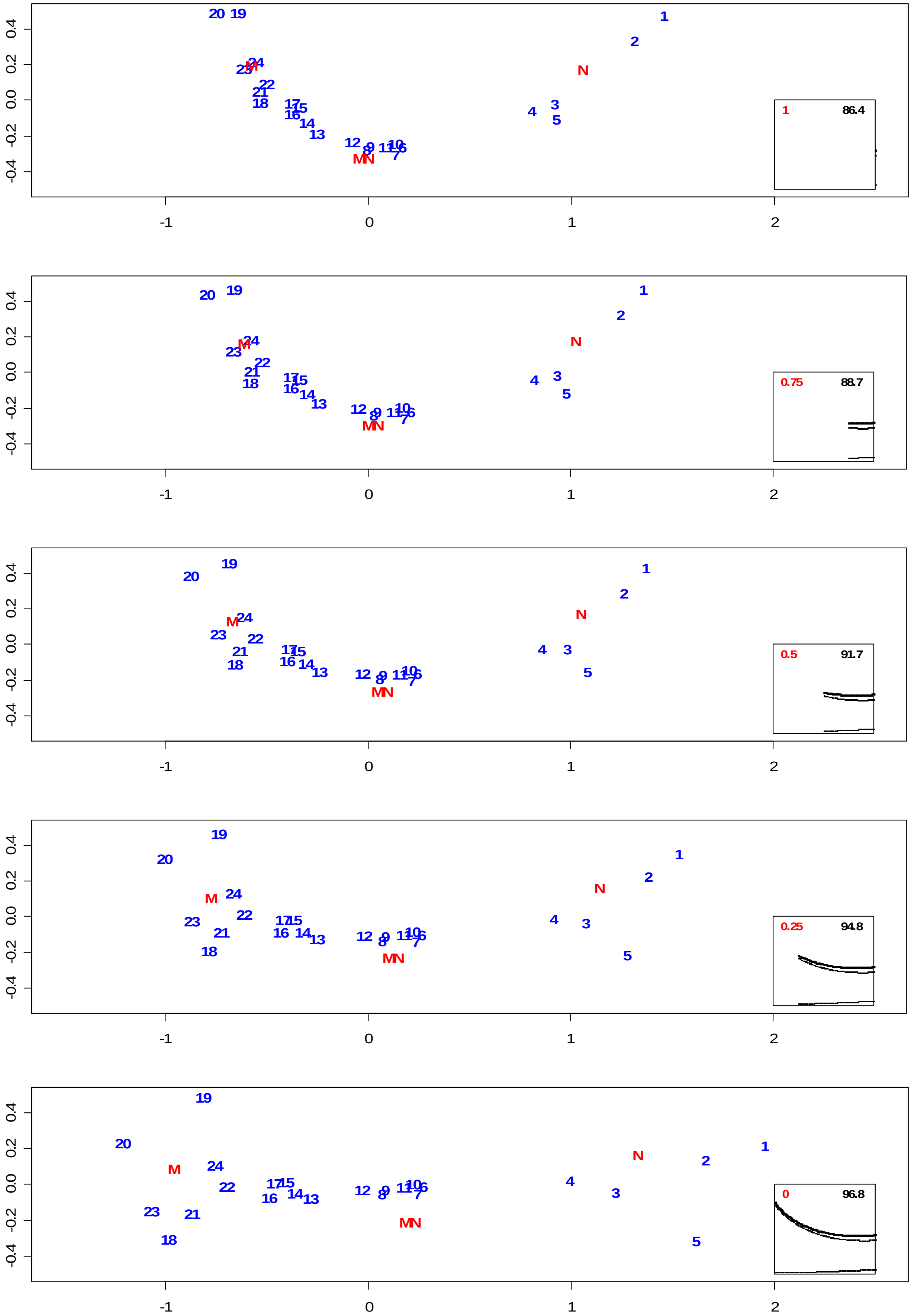

*Figure 2*: From correspondence analysis ($\alpha = 1$) to weighted log-ratio analysis ($\alpha \to 0$), with one intermediate "hybrid" analysis ($\alpha = ½$) for the "author" data, showing the symmetric maps. The box shows the numerical value of $\alpha$ and the percentage of inertia explained .in the two-dimensional map, as well as a graph of the values of the total inertia and two principal inertias as $\alpha$ decreases.

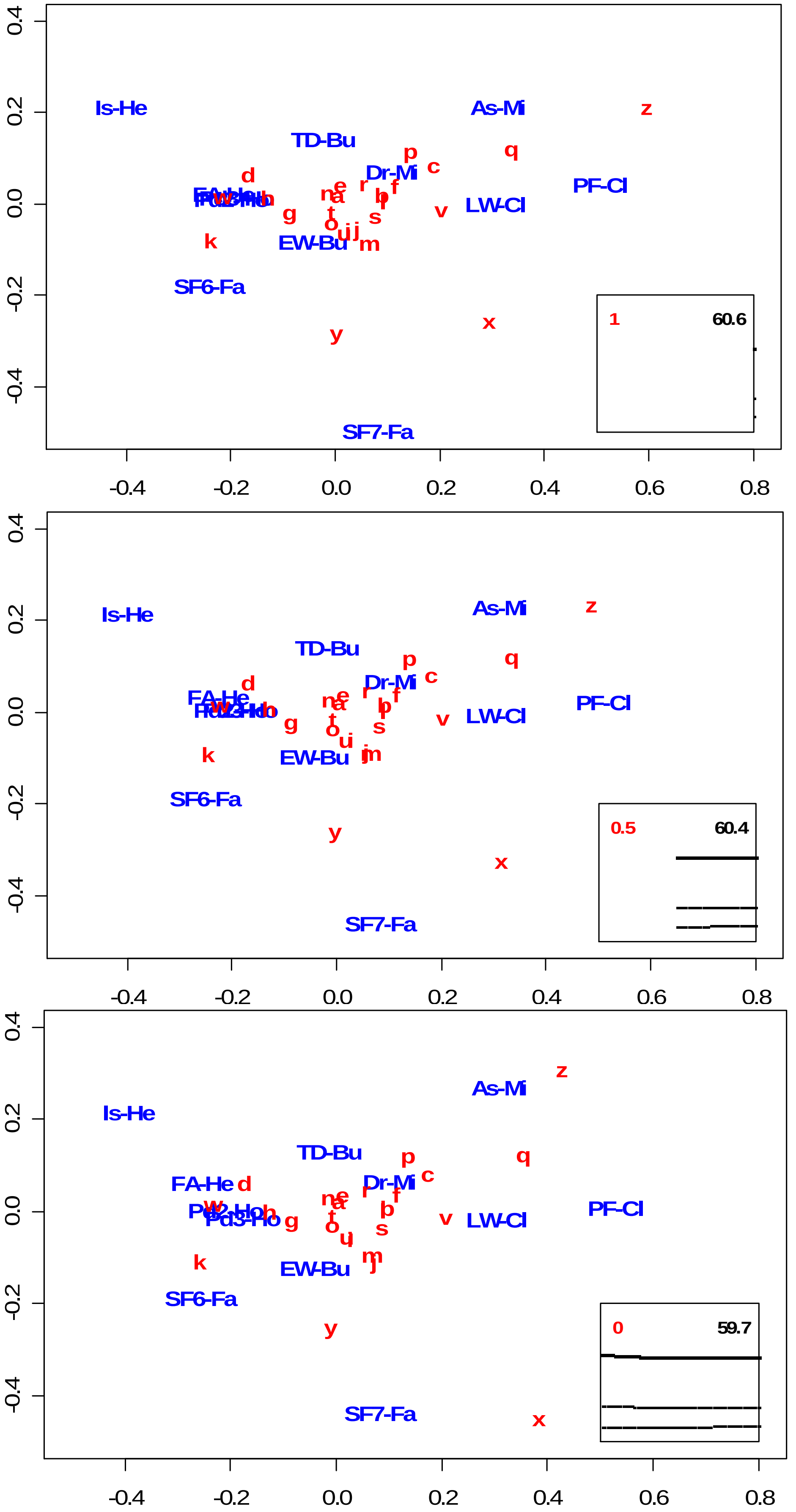

*Figure 3*: From correspondence analysis ($\alpha = 1$) to Hellinger analysis ($\alpha = 0.5$) for the "M-N" data. The box shows the numerical value of $\alpha$ and the percentage of inertia explained on the first dimension, and a graph of the values of the total inertia and two principal inertias as $\alpha$ decreases.

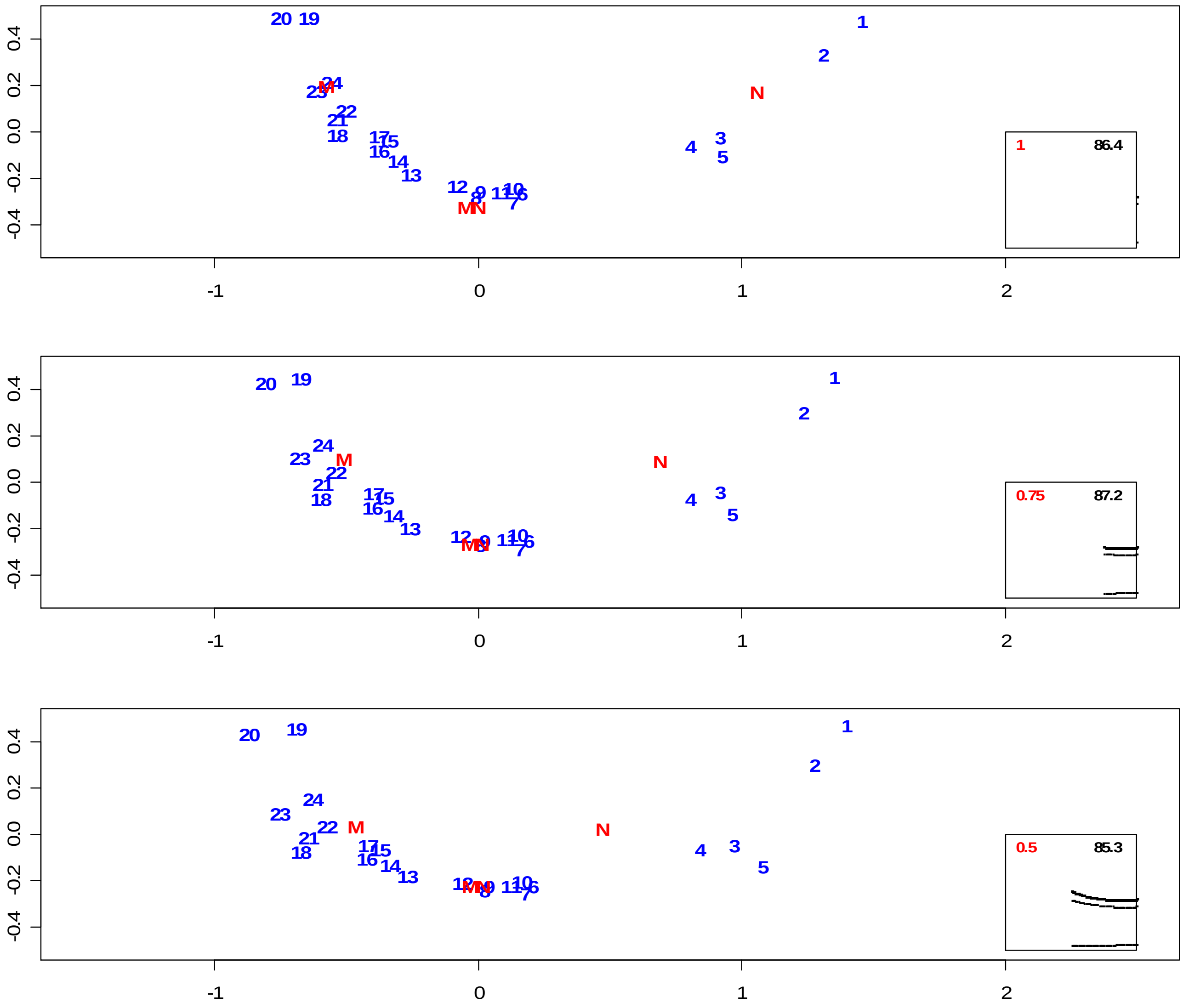

*Figure 4*: From correspondence analysis ($\beta = 1$) to non-symmetrical correspondence analysis ($\beta = 0$) for the “author” data, showing one intermediate “hybrid” step ($\beta = ½$ ). The asymmetric map is shown with columns in principal and rows in standard coordinates, where the column (letter) principal coordinates have been multiplied by 4 for better legibility. The box shows the numerical value of $\beta$ and the percentage of inertia explained in the two-dimensional map, as well as a graph of the values of the total inertia and two principal inertias as $\beta$ decreases.