Two more things about compositional biplots: Quality of projection and inclusion of supplementary elements

J. DAUNIS-I-ESTADELLA[1], S. THIÓ-HENESTROSA[2] and G. MATEU-FIGUERAS[2]

[1] Universitat de Girona, Spain josep.daunis@udg.edu
[2] Universitat de Girona, Spain

Abstract

The biplot is a widely and powerful methodology used with multidimensional data sets to describe and display the relationships between observations and variables in an easy way. Compositional data consist of positive vectors each of which is constrained to have a constant sum; due to this property standard biplots can not be performed with compositional data, instead of a previous transformation of the data is performed. Due to this constant sum constraint, a transformation of data is needed before performing a biplot and, consequently, special interpretation rules are required. However, these rules can only be safely applied when the elements of a biplot have a good quality of projection, for which a new measure is introduced in this paper. Also, we extend the compositional biplot defined by Aitchison and Greenacre on 2002, in order to include the display supplementary elements that are not used in the definition of the compositional biplot. Different types of supplementary elements are considered: supplementary parts of the composition, supplementary continuous variables external to the composition, supplementary categorical variables and supplementary observations. The projection of supplementary parts of the composition is done by means of the equivalence of clr and lr biplots. The other supplementary projections are done by classical methodology. Both the qualities of projections and the supplementary projections are explained using real geological data: a sample of 72 observations of soil in an area about 20 km west of Kiev in the area south of Kiev Polessie.

Keywords: Log-ratio transformation; Simplex; Aitchison geometry; Singular value decomposition; Principal component analysis

1. Introduction

Biplot is a very powerful tool for describing multidimensional data sets by means of a display in a two-dimensional space. Also, biplots are useful to reduce the dimensionality of a data set by concentrating most of the statistical information contained in the data into a few new variables that are linear combinations of the original ones. Aitchison and Greenacre (2002) have described the way to proceed with biplots of compositional data due to the relationships between parts of a composition. This special feature implies a special interpretation of compositional biplots and the elements projected in them.

This paper is focused into two different directions. The first direction goes deeper into the rules of interpretation of a biplot warning when they could be applied. The second direction extends the compositional biplot by means of the introduction of new supplementary elements that could improve the description of a data set.

The first purpose of the paper is intended to work as a warning about the rules of interpretation described by Aitchison and Greenacre. Using different biplots with a good overall quality of projection, we have seen that some problems using those rules of interpretation appear. These rules could only be safely applied when the elements of a biplot have a good individual quality of projection -not only a good overall quality- that is, a good quality for every projection involved in the rule. For this purpose, an appropriate measure of individual quality of projection is introduced.

Egozcue, J.J., Tolosana-Delgado, R. and Ortego, M.I. (eds.)
ISBN: 978-84-87867-76-7

The second purpose of the paper is to show the way to include in the display supplementary elements that are not used in the definition of the compositional biplot. The compositional biplot is performed with a set of parts of a composition. These parts could be called actives in front of other parts and variables of the data set, called supplementary, that don't play any role on the definition of the biplot. However, it is possible to project these supplementary parts on the previously displayed biplot. Different types of supplementary elements are considered: supplementary parts of the composition, supplementary continuous variables external to the composition, supplementary categorical variables and supplementary observations. The projection of supplementary parts of the composition is done by means of the equivalence of clr and lr biplots (Daunis-i-Estadella et al, 2010). The other supplementary projections are done by using classical methodology of biplots.

2. Compositional data

Compositions are vectors of positive components whose sum is constant. Due to this restriction the use of standard statistical techniques is not appropriate and consequently it is not possible to make classical interpretations, as far as on nineteenth century Karl Pearson (1897) described the problem. John Aitchison (1982, 1986) established the theoretical basis of the analysis of compositional data by using log-ratios. He stated that the interest lies in the relative magnitudes and variations of components instead of in their absolute values. Since then, other authors as Barceló-Vidal et al. (2001), Daunis-i-Estadella et al. (2006), Egozcue et al. (2003), Egozcue and Pawlowsky-Glahn (2005, 2006), Martín Fernández and Thió-Henestrosa (2006) and Pawlowsky-Glahn and Egozcue (2001), among others, developed and expanded the analysis of compositional data following the main principles defined by Aitchison. Specific software like the freeware CoDaPack (Thió-Henestrosa and Martín-Fernández, 2005, 2006; Thió-Henestrosa et al., 2008), the R packages "compositions" (Van den Boogaart et al., 2008) and "robCompositions" (Templ et al., 2010) implements the specific methodology of compositional data analysis.

One of the most common ways to work with compositional data is to use centred log-ratio (clr) vectors, whose expression is

$$clr(\mathbf{x}) = \left(\ln \frac{x_1}{g(\mathbf{x})}, \ln \frac{x_2}{g(\mathbf{x})}, \dots, \ln \frac{x_D}{g(\mathbf{x})} \right),$$

where g(x) is the geometric mean of the parts.

The clr-transformation maps a D-part composition to a D-part vector with the sum of its components equal to zero. As the clr transformation is bijective, the transformed vector lies on a plane which goes through the origin of the D-dimensional real space and is orthogonal to the vector (1,1,…,1). Consequently the covariance matrices of a clr transformation are always singular. The data can then be analysed in this transformation by all classical multivariate analysis tools not relying on a full rank of the covariance. As an alternative, the ilr-transformation maps a composition in the D-part Aitchison-simplex isometrically to a D-1 dimensonal euclidian vector. The data can then be analysed in this transformation by all classical multivariate analysis tools. However the interpretation of the results is more difficult, since there is no one-to-one relation between the original parts and the transformed variables.

As at those times the ilr transformation was not described, the centred log-ratio transformation was the transformation suggested by Aitchison and Greenacre (2002) to perform a compositional biplot.

3. Biplots

Let X be a data matrix of dimension n*D where most of the times the n rows represent observations and the D columns variables. Usually, depending on the type of data, X is transformed into a matrix Z. The most usual transformations are 1) to centre the variables or 2) to standardize the variables.

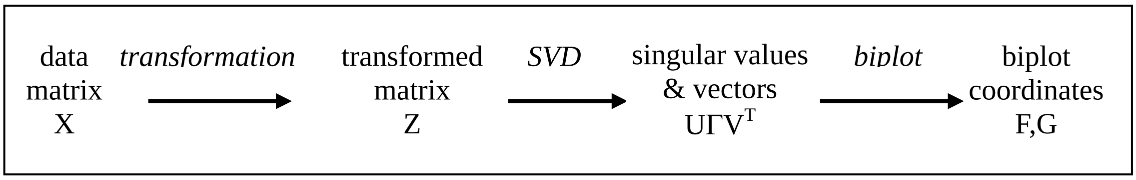

Fig. 1. General biplot scheme

Suppose that the transformed data matrix Z has rank r. The main idea of a biplot (Gabriel, 1971, 1981, Greenacre, 1984) is to find a factorization of the transformed matrix Z by means of a Singular Value Decomposition (SVD), $Z=U\Gamma V^T$. Then $Z=FG^T$, where F is a matrix of dimension is n*r and G D*r. A biplot is a graphical method to display on the same plot both the rows (as points) and columns (as rays) of the matrix Z, that is, the observations and the variables.

Usually the matrix Z is no symmetric, for this reason the left singular vectors U and the right singular vectors V are obtained by the SVD of (ZZ^T) and (Z^TZ), squared matrices:

$$ZZ^T = U\Gamma V^T V\Gamma U^T = U\Gamma^2 U^T ,$$
$$Z^T Z = V\Gamma U^T U\Gamma V^T = V\Gamma^2 V^T .$$

It is also possible to obtain the right singular vectors pre-multiplying the matrix Z by U^T, $U^TZ= U^TU\Gamma V^T=\Gamma V^T$, then $V^T = \Gamma^{-1}U^T Z$ and finally $V = \Gamma^{-1}Z^T U$. And respectively, $U = \Gamma^{-1}ZV$.

Depending on the construction of biplot coordinates, it is possible to obtain many kinds of biplots (see Greenacre and Underhill, 1982 for an exhaustive listing). The most usual are:

a) Covariance biplot. In this case $F=U$ and $G=V\Gamma$. It is called covariance biplot because the product $GG^T = V\Gamma^2 V^T$ divided by (n-1) is the covariance matrix. Thus, apart from the constant $\sqrt{n-1}$ the lengths of rays approximate the standard deviations and the cosines of angles between rays estimate their correlation. This biplot is appropriate to display the variables. This correction of $\sqrt{n-1}$ is made usually, as in Aitchison and Greenacre (2002) among others.

b) Form biplot. In this case $F=U\Gamma$ and $G=V$. The matrix of scalar products between the rows ZZ^T is now approximated by FF^T. This biplot is appropriate to display the observations or cases.

c) Symmetric scaling biplot. In this case $F = U\Gamma^{1/2}$ and $G = V\Gamma^{1/2}$. This biplot is a compromise between form and covariance biplots.

According the biplot types, the biplots are summarized in Table 1.

	Covariance	Form	Symmetric Scaling
Variable coordinates	$G = V\Gamma = Z^T U$	$G = V = Z^T U\Gamma^{-1}$	$G = V\Gamma^{1/2} = Z^T U\Gamma^{-1/2}$
Individual coordinates	$F = U = ZV\Gamma^{-1}$	$F = U\Gamma = ZV$	$F = U\Gamma^{1/2} = ZV\Gamma^{-1/2}$

Table 1: Variables and individual coordinates in the 3 biplot types

4 Geological data used for demonstration

The dataset used as an example is a sample of 72 observations of soil in an area about 20 km west of Kiev in the south of Kiev Polessie (Martin et al. 2004). There were determined the following 6 chemical elements of soil observations: Fe, Zn, Rb, Ba, Sr, and Pb. The concentrations of these elements were expressed in micrograms per gram (mg/g). Also in the database are defined columns with the name of the collected genetic sample layer, and the depth of sampling.

Elementary geochemical landscapes are grouped depending on their position in the relief, its relation with groundwater, and the way how they receive foreign chemicals (Glazovskaya 1963). Accordingly, we consider the following landscapes: eluvial-accumulative, accumulative-eluvial, transaccumulative, superaqueous and transeluvial. The first four groups also receive the qualification of dependent landscapes and the fifth one autonomous or independent landscape. This name tries to reflect the fact that these landscapes are chemical compounds, usually directly from atmospheric deposition. The rest of the elementary geochemical landscapes are called dependent because they receive the chemical components through precipitation, and also through lateral water flows -going for the slopes on the side of autonomous landscapes- and in low places of groundwater. So we have a 6-part composition, with an external continuous variable (Depth) and two external categorical variables.

5. The quality of projection

In this section we review the concept of quality of representation of a biplot, as a percentage of variance-inertia, and, based on this, we introduce the analogous concept of quality of projection related to an element in the biplot.

5.1. The overall quality of a biplot

Biplots display multidimensional clouds of data points by projecting them on a two-dimensional space whose axes represent uncorrelated derived combinations of parts that account for the greatest portion of the variation, called overall quality, among the individuals in the data cloud.

For general interpretation, the length of rays indicates the amount of variation among individuals accounted for by that variable, so a long vector indicates that values of that variable vary widely among individuals. But we must be careful with the nature of a biplot, it is just a projection and every part is projected on the plot, whenever the position of the part is in the original space. When we talk about proportion of explanation, we are talking about an overall measure of explanation, but not a specific accuracy measure.

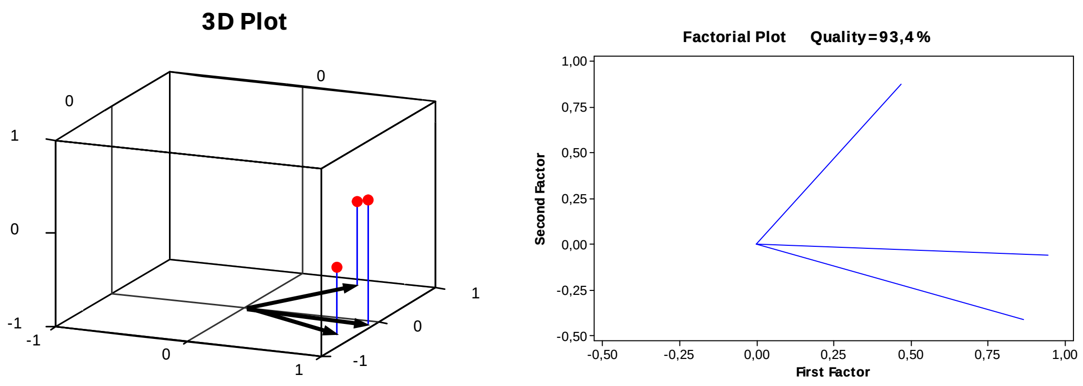

Fig. 2. Example of a very good representation, a) Original data in a three-dimensional space, b) Projection into a plane

ISBN: 978-84-87867-76-7

A good overall quality indicates that globally all the variability of the variables is collected by the biplot representation. But, more the dimensionality are, less the overall quality may be, except if there is not a strong relation among the parts

In Fig. 2, we have the complete three-dimensional representation of three-dimensional artificial variables (Fig. 2a) with a good overall quality of representation (Fig. 2b).

All the main variability (93,4%) among the three variables is represented in the biplot. Only a little variation on the vertical axis of Fig. 2a is not captured by the biplot. When we project these data into a two-dimensional space the loss of information is very small.

On the other hand, in Fig. 3a we have the three-dimensional display and its biplot (Fig. 3b) of another set of artificial data with only an overall quality of 75.2%. The vertical variability, now, is not captured in the two dimensional biplot, we have a loss of information in the biplot display corresponding to the vertical variation.

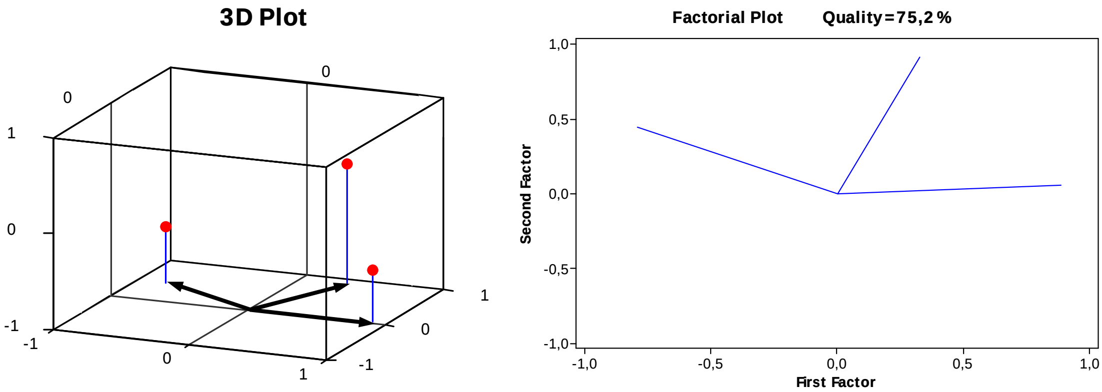

Fig. 3. Example of a not very good representation, a) Original data in a three-dimensional space, b) Projection into a plane

5.2. The specific quality of projection of elements of biplot

In Fig. 4a we can observe the biplot display of three elements, u, v and w (omitted in the figure) in a two dimensional space. The display is a projection of the original elements in the first factorial plane, which has the higher overall quality. According to Fig. 4b, the length of the projection of vectors u and w is equivalent to the original ones, whereas the length of projection of vector v is shorter than the original length. Vectors u and w are well represented but the representation of v is poor.

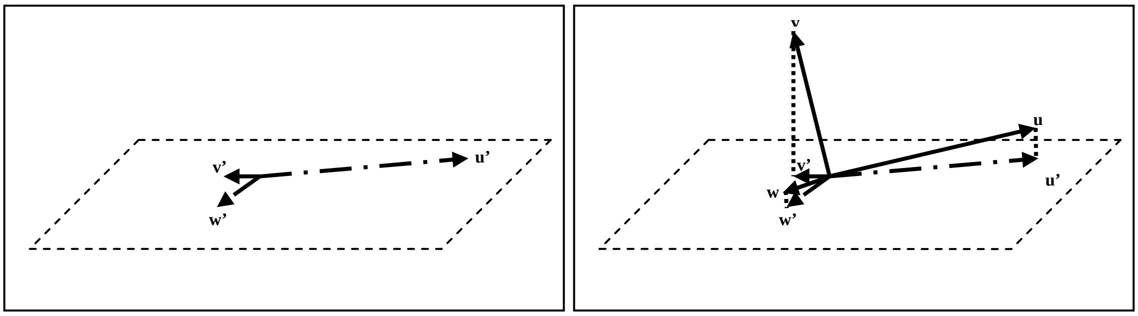

Fig. 4. Examples of good and bad projections, a) Projection of variables, b) Original and projected variables in a three-dimensional space

If we come into the representation of the vectors and its projections, we can decompose the projection of every vector in two perpendicular parts, g1 and g2, corresponding to the two factorial axes (Fig. 5).

ISBN: 978-84-87867-76-7

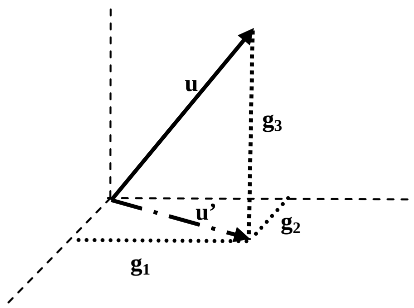

Fig. 5. Elements of a projection

Applying the Pythagoras' Theorem the square length of u', the projection of vector u, is equal to the sum of squares of the coordinates in the two first axes. But if we take this sum of squares over all the factorial coordinates, it is exactly the square length of the u vector that, in case of centered data, it is exactly the variance of this part,.

For this reason, we define a measure of the specific quality of the projection of one part, as:

$$\text{Specific quality} = \frac{\sum_{i=1,2} g_i^2}{\sum_{i=1,\ldots k} g_i^2} = \frac{\sum_{i=1,2} g_i^2}{\text{variance}}.$$

This definition could also be used to calculate the specific quality of the projection of individuals.

5.3. Specific quality and interpretation rules of compositional biplots

To use the interpretation rules of compositional biplots described by Aichison and Greenacre (2002) it is important not only a good overall quality of the biplot but also a good specific quality of projection of all parts involved in the rules of interpretation.

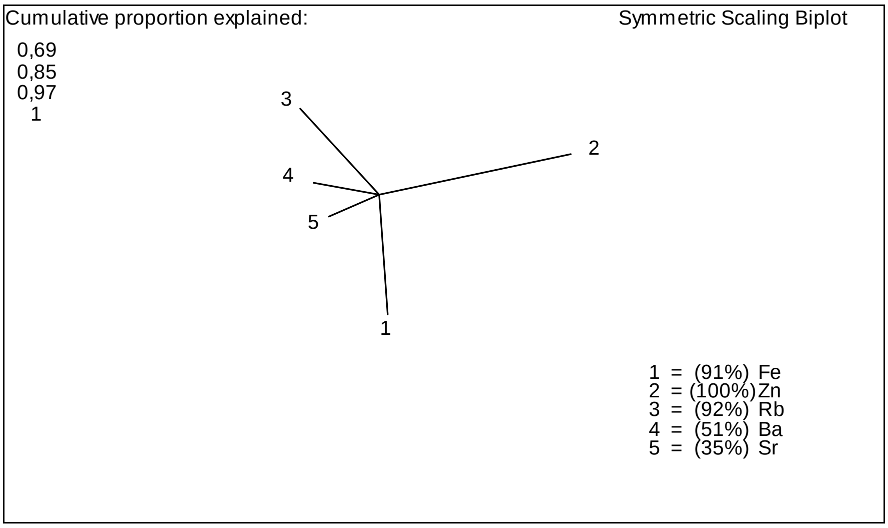

Fig. 6. Compositional Biplot of accumulative-eluvial landscapes (n=26)

In a biplot with a good overall quality, i.e. with a large proportion of variance explained, the lengths of rays with a high specific quality of projection in the plot are proportional to the clr variance of the components. Figure 6 shows the biplot of 26 observations corresponding to accumulative-eluvial landscapes. The overall quality of projection is quite good (85%) but the

ISBN: 978-84-87867-76-7

specific quality of projection of Ba and Sr is not good enough (51% and 35% respectively). Table 1 shows the clr variance of the five elements of the biplot. Zn is the element with higher variance while the other four elements have a similar one. However, the projection of these four components on the biplot only reflects well this variability in the elements well represented (Fe and Rb).

	Fe	Zn	Rb	Ba	Sr
clr variance	0,0924	0,2166	0,1073	0,0973	0,0945

Table 1. Clr variability of data of Fig. 6

Another rule of interpretation is that if three rays are aligned then the relation between these three parts is linear. In our exemple rays of the biplot (figure 7a) corresponding to parts Rb, Ba and Sr are aligned but, as it is possible to see in figure 7b, there is no lineal relation between these three parts. It can be seen if we model this relation by means of a Principal Component where the linear trend only captures a 66% of the variability of these three parts. This is due because the projection on the biplot of parts Ba and Sr has a poor specific quality of projection and the linearity shown on the biplot is not real.

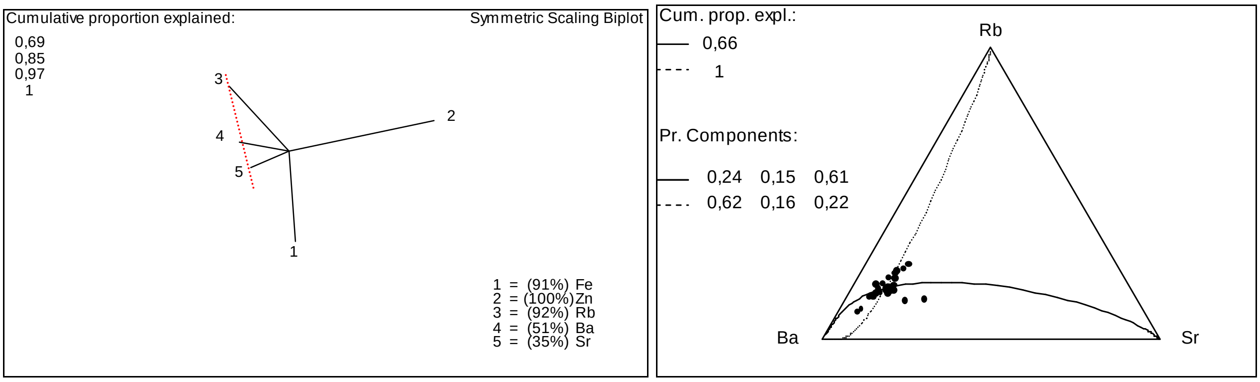

Fig. 7. a. Compositional Biplot of accumulative-eluvial landscapes (n=26). b. Ternary diagram of Na, Ba and Sr subcomposition with the corresponding principal components.

Fig. 8 is a replica of Fig. 6. Here, however, the ends of rays 3 (Rb) and 1 (Fe), and 5 (Sr) and 2 (Zn) are connected; these connecting lines are called links. When two links are orthogonal the participating pairs of rays may have a specialty: provided also that the specific quality of projection is good enough, the background variables, factors, processes, etc. associated with the participating pairs of components are probably independent.

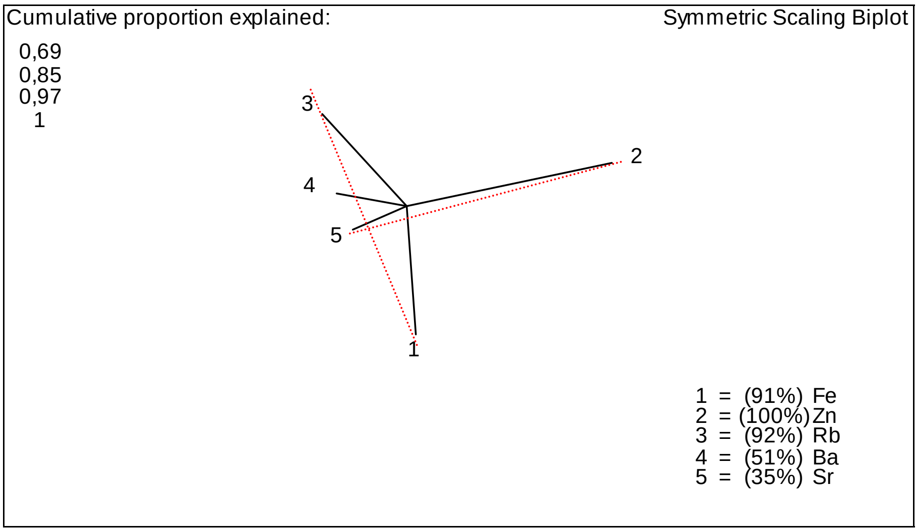

Fig. 8. Compositional Biplot of accumulative-eluvial landscapes (n=26) with orthogonal links.

ISBN: 978-84-87867-76-7

This independence may also be found when we calculate the correlation between the logratios of the two pairs of rays that generate the orthogonal links.

If the specific quality of projection is not good enough, applying this property may provide us wrong conclusions. For example, in Fig. 8, links between rays Rb and Fe, and Sr and Zn, appear perpendicular. Thus, a near zero correlation of their corresponding logratios is expected, but in this case

$$cor\left(\log\left(\frac{Rb}{Fe}\right),\log\left(\frac{Sr}{Zn}\right)\right)=0.383$$

significantly different from 0, which becomes comprehensible if one considers the bad quality of representation of one of the elements involved: specific quality of Sr is 35%.

Although we don't show an example of the other rules of interpretation, they could be applied only if the specific quality of projection is good.

6. Supplementary elements on a biplot

It is possible to project supplementary variables or supplementary observations in the different types of biplot. This new rays or this new observations markers are easily obtained by substitution of the different equations of the biplot types to obtain the F and G coordinates, but projecting over these coordinates the supplementary elements, with the scaling and centering transformations which are required according the previous transformations of the data matrices (Graffelman and Aluja, 2003).

These projections are an approximation of the original supplementary data by the scalar product with respect to all variable vectors of the coordinates, usually the first two dimensions. After factoring Z as FG^T we try to factor the matrix of supplementary information W as AB^T, where A is already given by F. From $W=FB^T$ we derive

$$B^T = \left(F^T F\right)^{-1} F^T W ,$$

and in case of orthogonal vectors, as our case, it is reduced to $B^T= F^T W$.

Likewise, a supplementary individual projection is done by

6.1 Projection of supplementary elements on a compositional biplot

Let be X the n*D matrix of compositional data which we want to perform the clr biplot. This is composed by a D-part composition of n observations. Let's call X as the active matrix to perform the biplot, that is, we have D active parts and n active observations.

Once performed the clr biplot with this active matrix, it is possible to project different kind of supplementary elements on the clr biplot, and, depending on the nature of this supplementary elements, there are different ways to proceed. Moreover, these projections of supplementary elements will be done according the different type of biplot coordinates (covariance, form and symmetrical scaling).

6.1.2 Projection of supplementary parts of the composition

Due to the special structure of the double centring transformation, on the clr components, it is not possible to include supplementary parts of the composition on a compositional biplot (clr biplot) as a simple projection of the part over the coordinates. By means of the relation of the clr and lr biplot (Daunis-i-Estadella et al, 2010), this problem can be solved easily, because it is possible to compute the log-ratio between this new supplementary part and all the active parts.

Given q parts of the composition, t_i, i=1...q, not used to define the biplot, it is possible to calculate the log-ratio of these new parts with each part of the active composition
$\log(t_i/x_j)$.

Once the log-ratio are done, let be W_i the n*D matrix, centered with respect to the column means, it is possible to plot the new log-ratio matrix W_i in the lr biplot as supplementary parts, where the supplementary coordinates F_s are given by $F_s=\Gamma_\alpha U^T W_i$ and it is possible to transfer this rays in the lr biplot to a new supplementary ray in the clr biplot by means the relation established on Daunis-i-Estadella, (2010), a clr ray is the addition of all the lr rays with the same numerator divided by de dimension p of the original composition.

6.1.3 Projection of external supplementary variables

Let be W the n*q matrix of q new external continuous variables, centered with respect to the column means, it is possible to plot the matrix W in the clr biplot as supplementary variables (Graffelman and Aluja, 2003), where the supplementary coordinates F_s are given by $F_s = F_{clr}^T W$.

As variables W usually are not measured with the same range of values of the parts, the rays F_s usually appear out of scale. For this reason it is better to standardize the data matrix W before to be projected in the biplot. The standardization of the variables implies that the projection F_s is only a direction of growing of these variables and the length of these supplementary rays are meaningless.

6.1.4 Projection of supplementary individuals

To project m new supplementary observations, $k_i=(k_{i1},...,k_{iD})$, i = 1...m, the first step is to center with respect to the center of original data. Once this centered Q_{m*D} matrix is obtained, the supplementary individual projection in the clr biplot is done by $G_s=G_{clr}{}^T Q$.

As alternative it is possible to compute the vector of log-ratios $\log(k_{ij}/k_{ij'})$, then to center with respect to the center of original data and finally project in the lr biplot. The projection over the clr biplot is obtained again as the addition of all the lr-rays divided by de dimension of the original composition.

6.1.5 Projection of supplementary categorical variables

A supplementary categorical variable can be represented in biplots using the previous results of section 5.3. In order to do so, the supplementary categorical variable should be coded by as many indicator variables as categories it has. Each category thus leads to a different vector or point in the biplot. This element can be calculated by the averages of the components for those cases that pertain to this particular category, and then it is possible to plot this element in the biplot to represent this category.

6.2. Example of application

Let the active dataset the 53 observations corresponding to dependent landscapes with a subcomposition of five parts (Fe, Rb, Ba, Sr, and Pb). The role of active dataset is to perform the biplot of this data and then project over this biplot, as supplementary elements, other components.

So, first of all a compositional biplot of the 5-parts is performed in order to describe all together observations and parts of the active dataset. On this biplot it is possible to see that the main factor that differentiates this population is the opposition between the variable clr(Pb) among the other ones. That is, an observation of high value of clr(Pb) has a low value on the others and vice-versa. This biplot contains a 90% of the variability of the data set (Figure 9).

As a second step a new part, Zn, is included as supplementary. Figure 10 shows the same clr biplot as Figure 9 with the exception of a new ray, in red colour, that is the projection of the supplementary part over the original axis of the clr biplot. This new ray goes to the direction of the variables clr(Rb) and clr(Sr). That means that as more clr(Rb) -or clr(Sr)- is in a sample, more clr(Zn) this observation has, and also the contrary.

ISBN: 978-84-87867-76-7

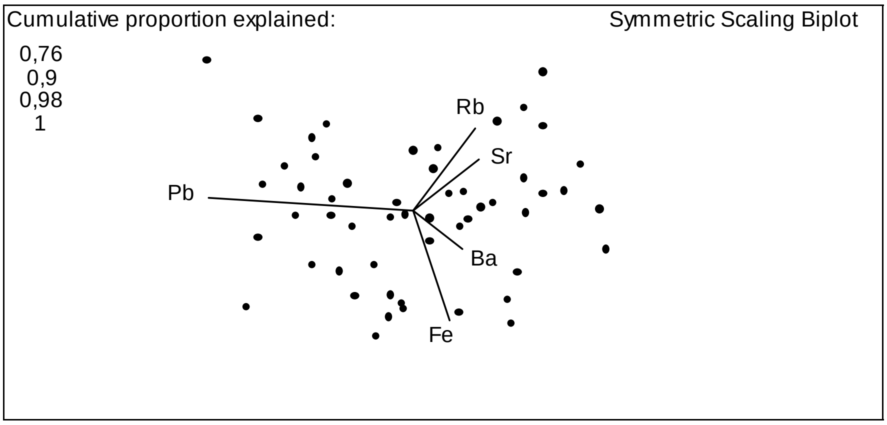

Figure 9. Clr biplot of the active data set with projection of parts (rays) and observations (dots).

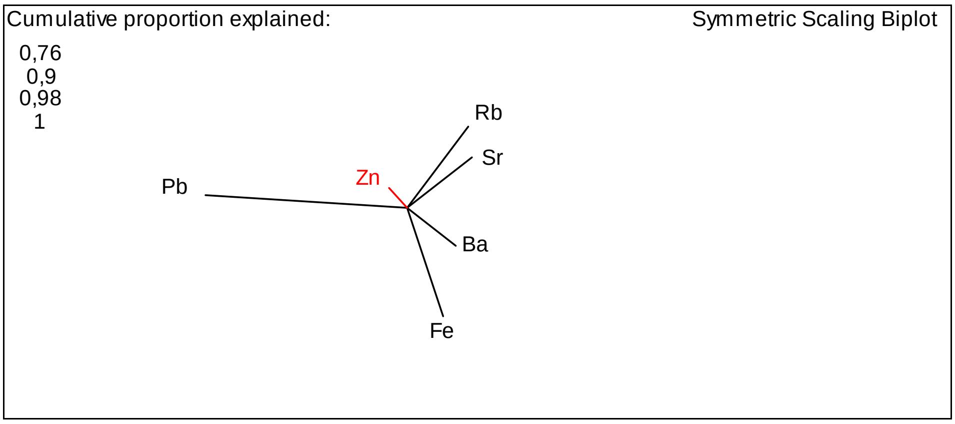

Figure 10. Clr biplot of the active data set with projection of active parts (black ray) and a supplementary part (red ray).

Also a new continuous variable, Depth, is included as supplementary, that is, the depth where the sample is done. Figure 11 shows the same clr biplot as Figure 9 with the exception of a new ray, a green arrow, that is the projection of the supplementary variable over the original axis of the clr biplot. This new ray goes opposite to the direction of the variable clr(Fe). That means that as more clr(Fe) is in a sample, less depth this observation is, and, in the contrary, as less clr(Fe) more depth.

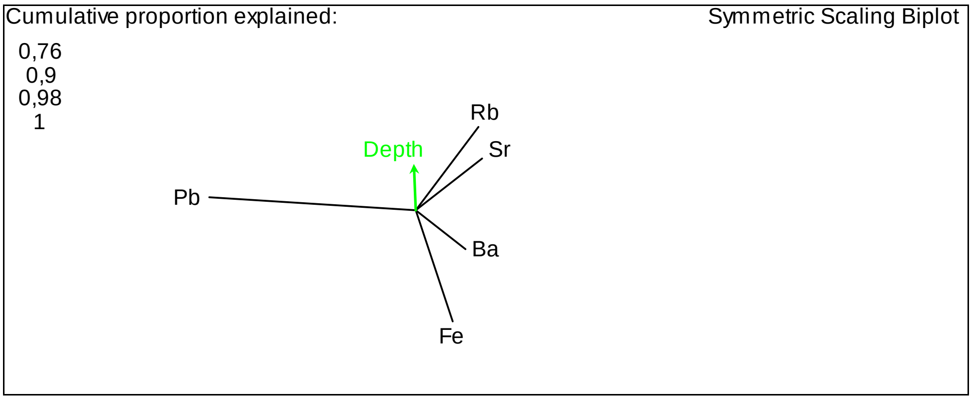

Figure 11. Clr biplot of the active data set with projection of parts (rays) and the continuous variable depth as supplementary (green arrow).

ISBN: 978-84-87867-76-7

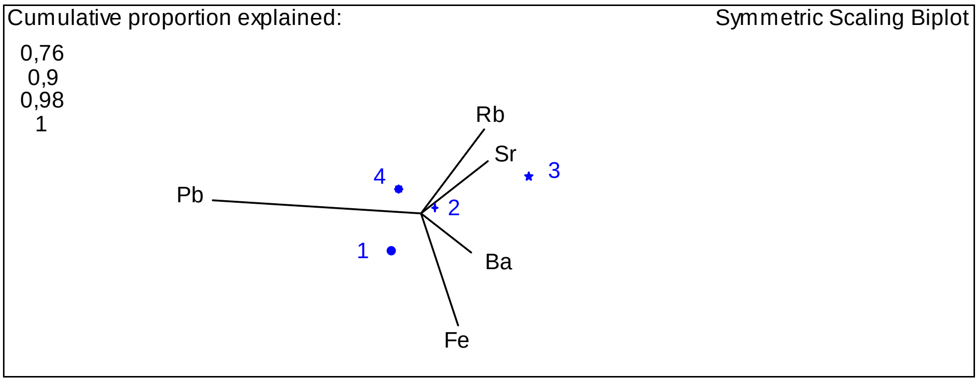

Figure 12. Clr biplot of the active data set with projection of parts (rays) and the categorical variable Landscape as supplementary (blue dots).

Also it is possible to project categorical variables as supplementary components of the original clr biplot. Figure 12 shows the projection of variable Landscape. This figure shows that landscape 3 (transaccumulative) is on the opposite side of clr(Pb) so the mean of observations corresponding to landscape 3 has a low value on clr(Pb) comparing with the other ones.

Again this supplementary projection could be done with all kind of categorical variables, for example some classification of the observations in groups.

Finally, a projection of supplementary observations is performed. The active observations correspond to dependent landscape while the 19 supplementary observations correspond to autonomous landscapes. Figure 13 show that the projection of supplementary observations (red triangles) is distributed all over the biplot following the same pattern than active observations (black dots) so there is no differences between both groups of observations.

This supplementary projection could be performed with any kind of observation with the same active parts. For example other samples obtained from other regions.

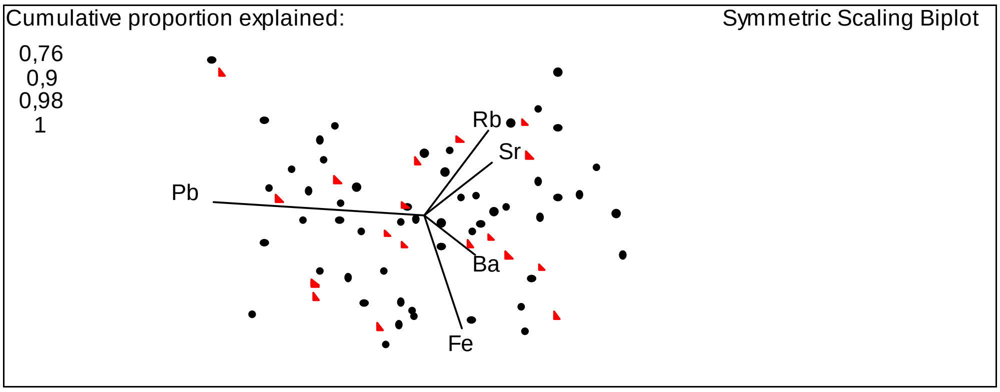

Figure 13. Clr biplot of the active data set with projection of other 19 observations as supplementary (red triangles).

7. Conclusions

Biplots are a very useful tool to represent multivariate data in a two dimensional space and they are also useful to reduce the dimensionality of a data set.

Due to the characteristics of compositional data analysis, biplots of compositional data analysis have their own rules of interpretation. These rules can only be applied with a biplot of high proportion of variance explained.

Also, some of the elements of a biplot can have a bad representation, and, consequently, they may be difficult to interpret. In order to discover these elements, we have presented a new measure, called specific quality of projection, which can be calculated for every element of a biplot (variables or observations). Only the elements with a good specific quality of projection can be used with the given rules of interpretation.

One of the usual tools of a biplot is to calculate the different components by means of a subset of variables, called active variables, and to project supplementary elements into the axis of the biplot. On a compositional biplot the active variables are parts of a composition and the supplementary elements could be other parts of the composition, external continuous variables, categorical variables or observations.

As the biplot of compositional data defined by Aitchison and Greenacre (2002) is performed by means of a clr transformation, only external continuous variables, categorical variables and observations could be projected directly into the components of a compositional biplot. It is not possible to project directly other parts of the composition because they aren't used into the clr transformation. The way to project these supplementary parts is by means of the lr biplot.

Acknowledgements

This work has received financial support from the Spanish Ministry for Education and Science through the project MTM2006-03040.

References

Aitchison, J. (1982). The statistical analysis of compositional data (with discussion): *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, v. 44 (2), p. 139-177.

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*: Chapman and Hall Ltd., London, (Reprinted in 2003 with additional material by The Blackburn Press), 416 p.

Aitchison, J., and Greenacre, M. (2002). Biplots for compositional data: *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, v. 51 (4), p. 375-392.

Barceló-Vidal, C., Martín-Fernández, J. A. and Pawlowsky-Glahn, V. (2001). Mathematical foundations of compositional data analysis. In: G. ROSS (ed.) *Proceedings of IAMG'01* – The sixth annual conference of the International Association for Mathematical Geology. 20, Kansas Geological Survey, Laurence, KS, (CD-ROM).

Daunis-i-Estadella, J., Barceló-Vidal, C., and Buccianti, A. (2006). Exploratory compositional data analysis, *in* Buccianti, A., Mateu-Figueras, G., and Pawlowsky-Glahn, V., eds, *Compositional Data Analysis in the Geosciences: From Theory to Practice*: Geological Society, London, Special Publications, v. 264, p. 161-174.

Daunis-i-Estadella, J., Thio-Henestrosa, S. and Mateu-Figueras,G. (2010). Including supplementary elements in a compositional biplot. *Computers & Geosciences*, In Press, Corrected Proof, Available online 10 November 2010, ISSN 0098-3004, DOI: 10.1016/j.cageo.2010.11.003.

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G. and Barceló-Vidal, C. (2003). Isometric log-ratio transformations for compositional data analysis. *Mathematical Geology*, 35 (3), 279–300.

Egozcue, J. J. and Pawlowsky-Glahn, V. (2005). Groups of parts and their balances in compositional data analysis. *Mathematical Geology*, 37 (7), 795–828.

Egozcue, J. J., and Pawlowsky-Glahn, V. (2006). Simplicial geometry for compositional data, *in* Buccianti, A., Mateu-Figueras, G., and Pawlowsky-Glahn, V., eds, *Compositional Data Analysis* *in the Geosciences: From Theory to Practice*: Geological Society, London, Special Publications, v. 264, p. 145-158.

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis: *Biometrika*, v. 58, p. 453-467.

Gabriel, K. R. (1981). Biplot display of multivariate matrices for inspection of data and diagnosis, *in* Barnett, V., ed., *Interpreting Multivariate Data*: John Wiley & Sons, London, p. 104-173.

Glazovskaya, M.A. (1963). On geochemical principles of the classification of natural landscapes, *Intern. Geol. Rev*. 5 , p. 1403–1431.

Graffelman, J.; Aluja-Banet, T. (2003). Optimal representation of supplementary variables in biplots from principal component analysis and correspondence analysis. *Biometrical Journal*, 45 (4) : 491-509. ISSN: 0323-3847

Greenacre, M. J., and Underhill, L. G. (1982). Scaling a data matrix in low-dimensional Euclidean space, *in* Hawkins, D.M., ed., *Topics in Applied Multivariate Analysis*: Cambridge University Press, Cambridge, p. 183-268.

Greenacre, M. J., 1984, *Theory and Applications of Correspondence Analysis*: Academic Press, London, 364 p.

Martín-Fernández, J. A., Daunis-i-Estadella, J., and Tyutyunnik, Y. G. (2004). Experiencia del estudio geoestadístico de composición química de suelos, de los indicadores de factores y de las condiciones geoquímicas. *Report de investigación IMA 04-01-RR*, Dept. d'Informàtica i Matemàtica Aplicada, Univ. de Girona, 50 p.

Martín-Fernández, J. A. and Thió-Henestrosa, S. (2006). Rounded zeros: some practical aspects for compositional data. In: Buccianti, A., Mateu-Figueras, G. and Pawlowsky-Glahn, V. (eds) *Compositional Data Analysis in the Geosciences: From Theory to Practice*. Geological Society, London, Special Publications, 264, 191–201.

Pawlowsky-Glahn, V. and Egozcue, J. J. (2001). Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment (SERRA)*, 15 (5), 384–398.

Pearson, K., (1897). Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurements of organs: *Proceedings of the Royal Society of London*, LX, p. 489-502.

Templ, M., Hron, K., Filzmoser, P. (2010). robCompositions: Robust Estimation for Compositional Data. *Manual and package, version 1.4.3. http://cran.r-project.org/package=robCompositions*

Thió-Henestrosa, S., Martín-Fernández, J.A. (2005) Dealing with compositional data: the freeware CoDaPack. *Mathematical Geology* 37 (7), 773-793.

Thió-Henestrosa, S., Martín-Fernández, J.A. (2006). Detailed guide to CoDaPack: a freeware compositional software. In: Buccianti, A., Mateu-Figueras, G., Pawlowsky-Glahn, V. (Eds.), *Compositional Data Analysis in the Geosciences: From Theory to Practice*, vol. 264. Geological Society, London, pp. 101-118 (Special Publications).

Thió-Henestrosa, S., Egozcue, J.J., Pawlowsky-Glahn, V., Ó. Kovács, L., Kovács, G.P. (2008).

Balance-dendrogram. A new routine of CoDaPack. *Computers and Geosciencies*, vol. 34(12), 1682-1696.

Van den Boogaart, G., Tolosana, R., Bren, M. (2008). compositions: Compositional Data Analysis. *Manual and package*, version 1.01-1. http://www.stat.boogaart.de/compositions

ISBN: 978-84-87867-76-7