Mathematical Geology, Vol. 34, No. 3, April 2002 (© 2002)

Understanding Perturbation on the Simplex: A Simple Method to Better Visualize and Interpret Compositional Data in Ternary Diagrams[1]

Hilmar von Eynatten,[2] Vera Pawlowsky-Glahn,[3] and Juan José Egozcue[4]

Perturbation is an operation defined on the simplex and can be used for centering compositional data in a ternary diagram, applying objective criteria. Because a straight line in the original diagram is still a straight line in the perturbed diagram, gridlines or compositional fields defined by straight lines can easily be included in the operation. Simultaneous perturbation of data, gridlines, and/or compositional fields is shown to improve both visualization and graphical interpretation of compositions in ternary diagrams. This is illustrated by some examples using simulated as well as real data.

KEY WORDS: compositional fields, geometric mean, metric center, rescaling.

INTRODUCTION

Ternary diagrams are frequently used in geosciences to visualize compositional data characterized by three or more components, which are amalgamated to three components. Problems with both visualization and graphical interpretation of these data particularly arise when the compositions are close to the boundaries of the ternary diagram.

Two different ways are generally used to get over these problems. The first is to simply cut away a part of the diagram and magnify the residual diagram (e.g., Decelles and others, 1998); the second is to rescale the diagram by multiplying at least one of the components by a scalar (e.g., Bhatia and Crook, 1986),

[1]Received 23 June 2000; accepted 28 May 2001.

[2]Institut für Geowissenschaften, FSU Jena, Burgweg 11, D-07743 Jena, Germany; e-mail: eynatten@geo.uni-jena.de

[3]Departament d'Informàtica i Matemàtica Aplicada, Universitat de Girona, Campus Montilivi—P1, E-17071 Girona, Spain; e-mail: vera.pawlowsky@udg.es

[4]Departament de Matemàtica Aplicada III, Universitat Politècnica de Catalunya, Campus Nord—C2, E-08034 Barcelona, Spain; e-mail: egozcue@ncsa.es

whereby the scalar is chosen using subjective criteria. Mathematically, this second option is known as perturbation, a commutative group operation defined on the simplex, which is equivalent to translation in real space (Aitchison, 1986). Perturbing any three-component data vector by its inverse results in the neutral element of the group, which is graphically the baricenter of the ternary diagram (Buccianti and others, 1999; Martín-Fernández, Barceló-Vidal, and Pawlowsky-Glahn, 1999). Therefore, perturbation by the inverse serves as a tool to move a composition into the center of the ternary diagram which, in fact, results in a rescaling of the original diagram without relying upon subjective criteria.

The aim of this paper is to show the potential for graphical interpretation derived from simultaneous perturbation of data, grid lines, and/or compositional fields. To do that, we first recall the essential properties of perturbation for our purposes and we demonstrate that perturbation on the simplex transforms straight lines into straight lines. This implies that gridlines or compositional fields defined by straight lines (e.g., discrimination diagrams) can easily be included in the perturbed diagram. Examples using both simulated and real data serve to illustrate the potential of the method in order to encourage its use in geosciences.

METHOD

Compositional data are by definition vectors of which each variable (component) is positive and with all components summing to a constant c. We will use $c = 100$, since it is usual in the geosciences to denote the components in percentages. The sample space for compositional data is not the real space R^D but the simplex S^D (Aitchison, 1986). If $D = 3$, the simplex is graphically represented by a ternary diagram.

The operation of perturbation on the simplex and its properties are defined in Aitchison (1986). For the purpose of this paper, some important properties are pointed out below:

(1) perturbing a vector $\mathbf{x} = (x_1, x_2, x_3)$ in S^3 by a vector $\mathbf{p} = (p_1, p_2, p_3)$ in S^3 results in a new vector $\mathbf{p} \oplus \mathbf{x} = C(p_1x_1, p_2x_2, p_3x_3)$ in S^3, where C denotes the closure operation, that is, each component of the vector $\mathbf{p} \oplus \mathbf{x}$ is divided by the sum of all its components; and

(2) perturbing a vector $\mathbf{x}$ by its inverse $\mathbf{x}^{-1} = (1/x_1, 1/x_2, 1/x_3)$ results in the neutral element $e = C(1, 1, 1) = (c/3, c/3, c/3)$, which is exactly located in the baricenter of a ternary diagram.

Therefore, perturbing a compositional data set by the inverse of its center g^{-1} results in an optimized rescaling, which leads to the centering of the data set around the baricenter of the ternary diagram (Martín-Fernández, and others, 1999; Buccianti, and others, 1999). The center of a compositional data set, obtained as the closure of the univariate geometric means of each component, is used, because it provides the most suitable measure of location of a compositional data set (Aitchison, 1989).

The third property of perturbation, which is essential for the purpose of this paper, is that it transforms a straight line in the original diagram, for example, a grid line of a ternary diagram or the boundary of a compositional field, into a straight line in the perturbed diagram. To prove this, consider a straight line L on the simplex S^3, embedded in 3D real space R^3. In R^3, the line L can be expressed as the intersection of the plane containing the simplex, whose equation is $x_1 + x_2 + x_3 = 1$, and a plane P. There are an infinity of planes containing a given line. Thus, to find the equation of P, it is necessary to choose one out of that manifold. For simplicity, it seems reasonable to choose the one passing through the origin, whose equation is of the form $ax_1 + bx_2 + cx_3 = 0$.

Consider now a perturbation $p = (p_1, p_2, p_3)$ on S^3. A perturbation is always computed in two steps: for a point $x = (x_1, x_2, x_3)$ in S^3 we compute first (p_1x_1, p_2x_2, p_3x_3), and then we apply the closure operation. The points lying on the line L satisfy simultaneously the equations $x_1 + x_2 + x_3 = 1$ and $ax_1 + bx_2 + cx_3 = 0$. Therefore, multiplication with the components of the perturbation vector implies that the new points $(y_1, y_2, y_3) = (p_1x_1, p_2x_2, p_3x_3)$ satisfy the equation of the transformed plane $ap_2p_3y_1 + bp_1p_3y_2 + cp_1p_2y_3 = 0$, which also passes through the origin. The closure operation is equivalent to intersecting this transformed plane with the plane $x_1 + x_2 + x_3 = 1$ containing the simplex, and the intersection of two planes in R^3 is always a straight line.

There remains one open question, and that is if the intersection can be empty for some perturbation. The answer is no. By definition, components of a perturbation are positive, thus forcing the transformed plane to have a nonempty intersection with the positive orthant of R^3 and, as it goes through the origin, it has to have a nonempty intersection with the simplex.

Note that this reasoning can be directly extended to straight lines embedded in a simplex of an arbitrary dimension, and therefore it holds in general that the perturbation of a straight line in the simplex is again a straight line in the simplex.

Finally, let us point out a fact concerning the labelling of vertices before and after perturbation. Mathematically, vertices are invariant under perturbation, as they have all components equal to zero except for one, which is equal to one. Thus, considered as points, their components remain the same before and after perturbation. They play the role of points at infinity in real space, which remain also unchanged by monotone increasing transformations. But, to our understanding, the labels refer both to the point and to the scale of the variables considered, and the scale changes with perturbation. Therefore, we suggest using the following notation: for a vertex V write after perturbation $p(V)$ to indicate (1) that it is a perturbed and rescaled diagram and (2) to point out the fact that perturbation of vertices remain unchanged.

EXAMPLES USING SIMULATED DATA

To illustrate the graphical effects of perturbation we have simulated three compositional data sets each represented by 25 compositions of components X, Y, and Z. Theoretical means were chosen so that the three data sets represent the following three cases: (1) compositions distributed near the baricenter of the ternary diagram; (2) compositions distributed along the binary mixing line between two components; and (3) compositions distributed near one vertex of the ternary diagram. Simulation was performed in R^2 assuming a bivariate normal distribution with variances equal to 0.5 and correlation equal to 0.6. After simulation the data were transformed back into S^3 using the agl transformation (Aitchison, 1986), and the three centers $g_i (i = 1, 2, 3)$ were computed (Figs. 1(A), 2(A), and 3(A)). Now we perturb each data set including its center and the grid (chosen as 1, 10, 33, 67, 90, 99%) with the inverse of its center g_i^{-1}. The resulting ternary diagrams are shown in the corresponding Figures 1(B), 2(B), and 3(B).

In each case the center of the data set is moved to the baricenter of the diagram, but its position (and also the position of the 25 compositions) with respect to the simultaneously perturbed gridlines is the same. The latter can be best shown by the first example (Fig. 1), which shows a data set with a center close to the baricenter of the ternary diagram: after perturbation the gridlines are, of course, no longer parallel to the boundaries, but the subfields defined by the gridlines are still quite similar in shape and the position of the data relative to the grid are visually pretty much the same as in the original diagram. Actually, they are exactly the same if an appropriate distance is used.

In general, the method applied allows us to better visualize the internal structure of the data and this advantage becomes more important the closer the data

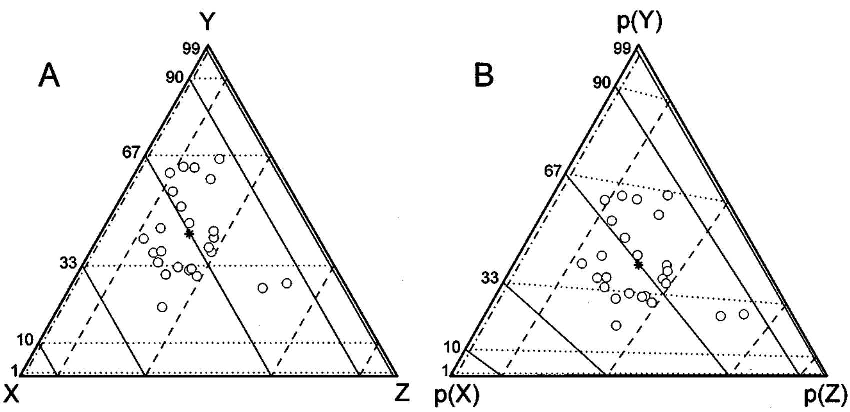

Figure 1. Original, A, and perturbed diagram, B, of the simulated data set 1 with center $g_1 = (33.6, 42.8, 23.6)$.

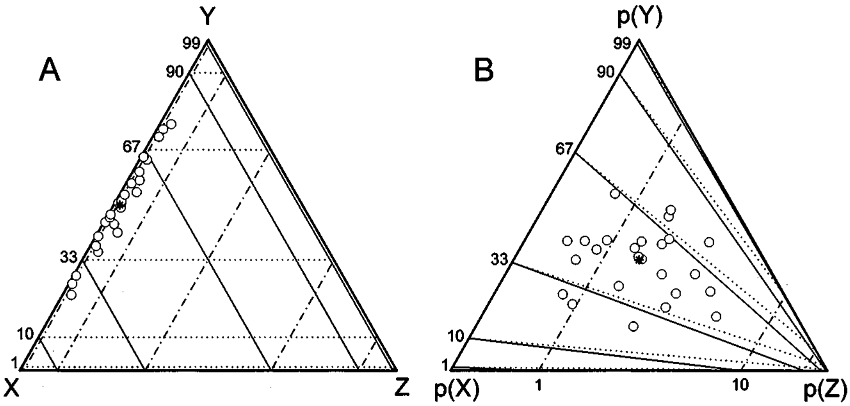

Figure 2. Original, A, and perturbed diagram, B, of the simulated data set 2 with center $g_2 = (48.4, 50.0, 1.6)$.

lie to the boundaries of the simplex (Figs. 2 and 3). A simple explanation lies in the fact that people are used to thinking in Euclidean distances, but the latter are not appropriate measures of distance in the simplex (Martín-Fernández, Barceló-Vidal, and Pawlowsky-Glahn, 1998). However, the closer the two compositions are to the baricenter of a ternary diagram the more "Euclidean" is the distance between them, or, the closer the two compositions are to the boundaries of a ternary diagram the worse is our visual assessment of distance between them.

Furthermore, because the perturbation operation is also applied to the grid, the method still enables us to assess the original values from the perturbed diagram.

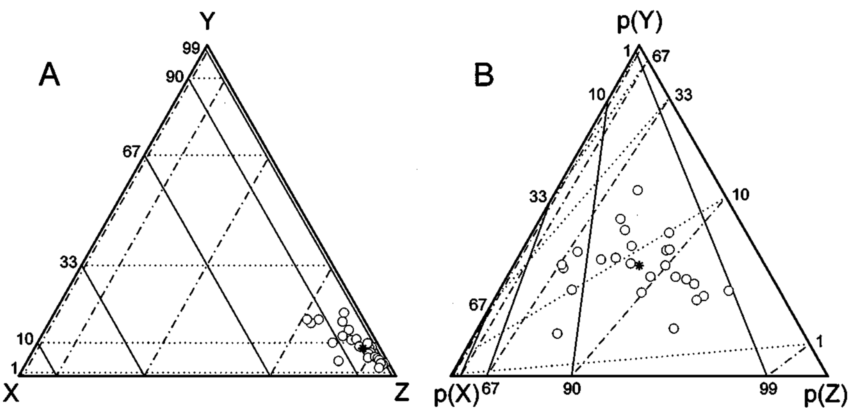

Figure 3. Original, A, and perturbed diagram, B, of the simulated data set 3 with center $g_3 = (4.6, 8.3, 87.2)$.

For example, in Figure 2(B) the absolute variation of the very small component Z as well as the relative variation of all the compositions with respect to the center is much better visualized than in Figure 2(A). We can assess directly from the perturbed diagram that, for example, 3 of 25 compositions are higher than 67% in component Y and 8 of 25 compositions are lower than 1% in component Z. In Figure 3(B), the variations of the smaller components X and Y, both absolutely as well as relative to the centers, are much better visualized than in Figure 3(A), and again, we can better assess the original values from the perturbed diagram, for example, 8 of 25 compositions are both lower than 10% in component Y and lower than 90% in component Z.

EXAMPLES USING REAL DATA

To illustrate the method using real data we have chosen an example from the literature by Critelli and Ingersoll (1994). The study reports petrographic data (sandstone framework composition) of two coeval petrofacies of the Neogene Siwalik Group, one situated in northwestern Pakistan ($n = 64$), the other in Nepal ($n = 20$). The example was not chosen to question the conclusions drawn by Critelli and Ingersoll (1994).

The data are represented graphically as detrital modes in several ternary diagrams of which we have chosen the QFL and LmLvLs diagrams for illustration purposes. The former represents proportions of quartz (Q), feldspar (F), and lithoclasts (L). The latter represents proportions of metamorphic (Lm), volcanic (Lv), and sedimentary lithoclasts (Ls). The QFL diagram is superposed on the major provenance fields defined by Dickinson (1985), which are frequently used in sedimentary petrology (Fig. 4A). The two data sets are separated on the basis of higher amounts of feldspar and (meta)volcanic lithoclasts in the Pakistan samples. To better visualize the petrographic difference between the two groups the authors draw “error polygons” based on univariate standard deviations of each variable from the arithmetic mean, although no statistical rigor supports this method (Pawlowsky-Glahn and Barceló-Vidal, 1999).

In the perturbed QFL diagram (Fig. 4(B)) the two data groups are more clearly separated from each other. The structure of the Pakistan samples (crosses) is quite similar to the original diagram (Fig. 4(A)) because the data were already relatively close to the baricenter (compare Fig. 1). The structure of the Nepal samples (open circles) becomes much clearer, because (i) the data are obviously less homogenous than suggested by the original diagram, (ii) the separation into two subgroups becomes more visible, and (iii) the dispersion of the quartz-rich subgroup ($n = 12$) turns out to be almost as high as the dispersion of all the Pakistan samples. Because of the simultaneous perturbation of the compositional fields of Dickinson (1985) the information on the recycled orogen provenance of both data sets can also be read from the perturbed diagram.

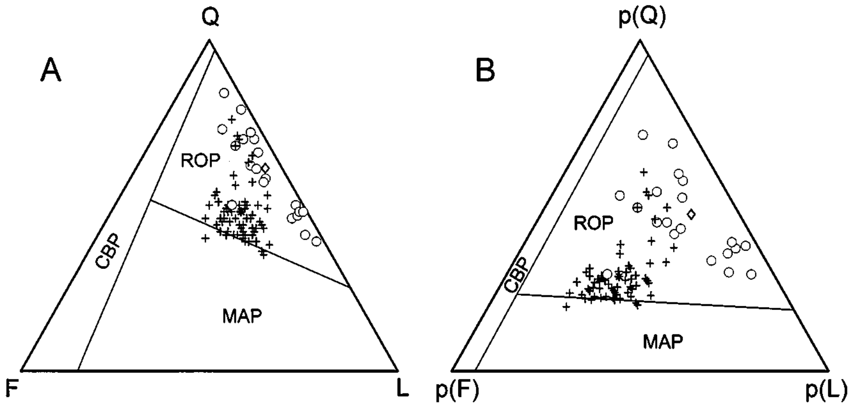

Figure 4. QFL diagram of petrographic data from Critelli and Ingersoll (1994). Crosses indicate Pakistan samples, open circles indicate Nepal samples, star and open diamond indicate centers of Pakistan and Nepal samples, respectively. Compositional fields defined by Dickinson (1985): ROP = recycled orogen provenance, MAP = magmatic arc provenance, CBP = continental block provenance. A: original diagram. B: perturbed diagram.

The LmLvLs diagram of the same samples closely resembles Case 2 of the simulated data (Fig. 2). The structure of both sample groups as well as the separation between them is obscured due to their composition near to the binary mixing line between two components Lm and Ls (Fig. 5(A)). In the perturbed diagram (Fig. 5(B)) both the structure of the data and the separation between the

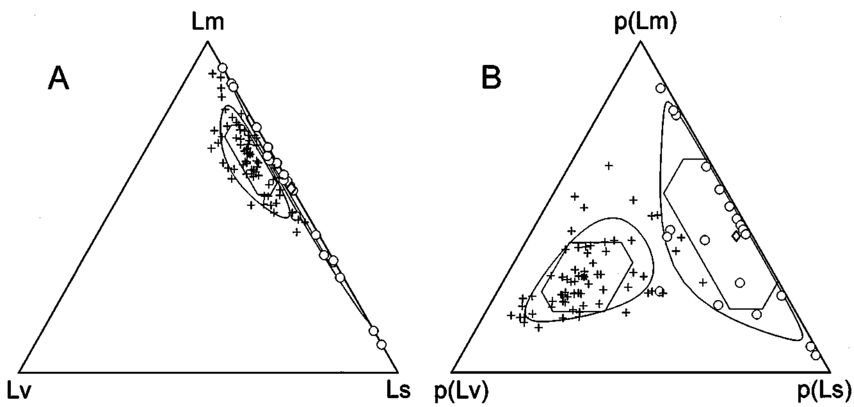

Figure 5. LmLvLs diagram of petrographic data from Critelli and Ingersoll (1994). For key see Figure 4. Both original, A, and perturbed, B, diagrams show one standard deviation "error polygons" as well as 67% predictive regions for both data groups. To calculate centers and predictive regions zero values are replaced following the method of Martín-Fernández and others (2000).

two groups is again better visualized. As a statistically sound alternative to one standard deviation "error polygons" used by Critelli and Ingersoll (1994) we show 67% predictive regions for the two data groups which were constructed assuming the additive logistic normal distribution corresponding to a multivariate normal distribution in real space (Aitchison, 1986, p. 175). The 67% predictive regions cover a considerably larger area than the "error polygons." Although the sample groups can be seperated at the 67% level, care should be placed on the assertion that the two data groups can be strictly separated in the LmLvLs simplex.

CONCLUSIONS

Centering of compositional data in ternary diagrams by perturbation allows better visualization and graphical interpretation of the structure of the data. The grid may be perturbed simultaneously, thus preserving information on the original values in the perturbed diagram. Graphical assessment of the original values may be even better in cases where compositions are close to the boundaries of the simplex. Instead of or in addition to grids, formerly defined compositional fields (not only those defined by straight lines) may be included into the perturbation, so that the potential of graphical interpretations based on such compositional fields is fully preserved.

The method is capable of finding applications in several subdisciplines of geosciences. It should be clear that only graphical improvements are suggested which do not replace rigorous statistical analysis of compositional data.

ACKNOWLEDGMENTS

The idea of illustrating the potential of perturbation on the simplex for graphical assessments in ternary diagrams was worked out during a stay of HvE at UPC Barcelona supported by the Deutsche Forschungsgemeinschaft (Grant EY 23/2). Research on compositional data at UPC is supported by the Dirección General de Enseñanza Superior (DGES) of the Spanish Ministry for Education and Culture through the project PB96-0501-C02-01. Many thanks to Salvatore Critelli for supplying petrographic raw data from the Siwalik Group.

REFERENCES

Aitchison, J., 1986, The statistical analysis of compositional data: Chapman and Hall, London, 416 p.

Aitchison, J., 1989, Measures of location of compositional data sets: Math. Geol., v. 21, no. 7, p. 787–790.

Bhatia, M. R., and Crook, K. A. W., 1986, Trace element characteristics of graywackes and tectonic discrimination of sedimentary basins: Contrib. Miner. Petrol., v. 92, p. 181–193.

Buccianti, A., Pawlowsky-Glahn, V., Barceló-Vidal, C., and Jarauta-Bragulat, E., 1999, Visualization and modeling of natural trends in ternary diagrams: A geochemical case study, *in* Lippard, S. J., Naess, A., and Sinding-Larsen, R., eds., Proceedings of IAMG'99—The fifth annual conference of the International Association for Mathematical Geology: Trondheim, p. 139–144.

Critelli, S., and Ingersoll, R. V., 1994, Sandstone petrology and provenance of the Siwalik group (Northwestern Pakistan and Western-Southeastern Nepal): J. Sediment. Res., v. 64, p. 815–823.

DeCelles, P. G., Gehrels, G. E., Quade, J., Ojha, T. P., Kapp, P. A., and Upreti, B. N., 1998, Neogene foreland basin deposits, erosional unroofing, and the kinematic history of the Himalayan fold-thrust belt, western Nepal: Geol. Soc. Am. Bull., v. 110, p. 2–21.

Dickinson, W. R., 1985, Interpreting provenance relations from detrital modes of sandstones, *in* Zuffa, G. G., ed., Provenance of arenites: Reidel, Dordrecht, p. 333–361.

Martín-Fernández, J. A., Barceló-Vidal, C., and Pawlowsky-Glahn, V., 1998, Measures of difference for compositional data and hierarchical clustering methods, *in* Buccianti, A., Nardi, G., and Potenza, R., eds., Proceedings of IAMG'98—The fourth annual conference of the International Association for Mathematical Geology: Napoli, p. 526–531.

Martín-Fernández, J. A., Barceló-Vidal, C., Bren, M., and Pawlowsky-Glahn, V., 1999, A Measure of difference for compositional data based on measures of divergence, *in* Lippard, S. J., Naess, A., and Sinding-Larsen, R., eds., Proceedings of IAMG'99—The fifth annual conference of the International Association for Mathematical Geology: Trondheim, p. 211–215.

Martín-Fernández, J. A., Barceló-Vidal, C., and Pawlowsky-Glahn, V., 2000, Zero replacement in compositional data sets, *in* Kiers, H. A. L., Rasson, J.-P., Groenen, P. J. F., and Schader, M., eds., Advances in Data Science and Classification, Proceedings of the 7th Conference of the International Federation of Classification, IFCS'2000, Namur: Springer, Berlin, p. 155–160.

Pawlowsky-Glahn, V., and Barceló-Vidal, C., 1999, Confidence regions in ternary diagrams: 89th Annual meeting of the Geologische Vereinigung, Terra Nostra (Schriften der Alfred-Wegener-Stiftung), Vol. 99, No. 1, p. 37–47.