

Reply to “On the Harker Variation Diagrams; . . .” by J.A. Cortés

Juan José Egozcue

Received: 30 August 2008 / Accepted: 26 July 2009 / Published online: 3 September 2009
© International Association for Mathematical Geosciences 2009

Keywords Compositional data principles · Olivine · Fractionation · Stoichiometry · Lever rule · Conservation of mass

1 Introduction

In the letter “On the Harker Variation Diagrams; A Comment on “The Statistical Analysis of Compositional Data. Where Are We and Where Should We Be Heading?” by Aitchison and Egozcue (2005)” by J.A. Cortés in *Mathematical Geosciences* (Cortés 2009), the author starts with some criticisms of those he refers to as compositional statisticians (Aitchison and Egozcue 2005). In this latter paper, the authors discourage the use of Harker Variation Diagrams (HVD). The main argument used is that such diagrams do not display meaningful features from the compositional point of view and, particularly, that any correlation can be misleading and spurious. An apparent confrontation between the point of view of the so-called petrologists and compositional statisticians is presented in Cortés (2009). I would like to clarify that these collective entities, if any, are not represented by either the present author or by Cortés. Both authors are looking for scientific principles when analysing data. Therefore, it does not matter from which field the data come from, or the authors are affiliated to. To this end, a brief discussion on the principles of compositional data analysis may clarify what can be seen in a HVD and what is not relevant in the analysis.

J.J. Egozcue (✉)

Dept. Matemàtica Aplicada III, Universitat Politècnica de Catalunya, Jordi Girona Salgado 1-3, C2,
08034 Barcelona, Spain
e-mail: juan.jose.egozcue@upc.edu

2 Principles of Compositional Data Analysis

The principles of compositional data analysis were established by Aitchison (1986) and they were inspired more by the experience in different scientific fields, like geology, than by previous purely statistical contributions. We can reduce these principles to the definition of compositional data and three principles summarising their essence. Based on a number of contributions (Aitchison and Egozcue 2005; Aitchison 1986), as well as some additional key contributions (Barceló-Vidal et al. 2001; Martín-Fernández et al. 2003), I propose the following formulation.

Definition A compositional vector, or simply a composition, of D parts is a positive real vector of D components, describing quantitatively the parts of some whole, which carry exclusively relative information between the parts.

The following principles give a more detailed insight into the previous definition.

- A. *Scale Invariance*. The information in a composition does not depend on the particular units in which the composition is expressed. Proportional positive vectors represent the same composition. Any sensible characteristic of a composition should be invariant under a change of scale.
- B. *Permutation Invariance*. Permutation of components of a composition does not alter information conveyed by the compositional vector.
- C. *Subcompositional Coherence*. Information conveyed by a composition of D parts should not be in contradiction with that coming from a sub-composition containing d parts, $d \leq D$. This generic principle can be formulated more precisely as

Sub-compositional Dominance: If $\Delta_n(\cdot, \cdot)$ is any distance between compositions of n parts then

$$\Delta_D(\mathbf{x}, \mathbf{y}) \geq \Delta_d(\mathbf{x}_d, \mathbf{y}_d),$$

where \mathbf{x}, \mathbf{y} are compositions of D parts and $\mathbf{x}_d, \mathbf{y}_d$ are sub-compositions of the previous ones with d parts, $d \leq D$.

Ratio preserving: Any relevant characteristic expressed as a function of the parts of a composition is exclusively a function of the ratios of its parts. In a sub-composition these characteristics depend only on the ratios of the selected parts and not on the discarded parts of the parent composition. Principle A applies to the sub-composition.

These three principles are not specifically statistical. They are based on scientific practice, particularly on geochemistry. Principle A comes from the fact that the mass of geological samples is considered irrelevant. Even using units like g/litre or molar units, principle A should be applied. To use these units, the absolute masses obtained experimentally should be multiplied by constants to reduce them to per litre or per unit of mass (Buccianti and Pawlowsky-Glahn 2005). Note that principle A applies independently of the fact that the composition is closed or not. Principle B is quite trivial and appears naturally in geochemistry, when the elements are ordered alphabetically. Principle C is the most subtle of the three principles. The spurious correlation when using closed data is a demonstration of the violation of such a principle and has

been extensively studied in geological contexts (Chayes 1960). It is remarkable that these principles are not based on mathematical or statistical considerations but on a generic scientific practice in search of rigorous general methods.

3 Variation Diagrams from the Compositional Point of View

Now one can try to apply the compositional principles to HVD, assuming that the data represented in such diagrams are really compositional. In the HVD the concentration in weight-percent of two elements (typically $x = \text{SiO}_2$ and $y = \text{MgO}$) are plotted. Therefore, the implied composition is (x, y, z) , where z represents the weight-percent of the remaining elements. A first point is about the definition of z . For instance, if the total is referred to the mass of major oxides, one can ask about whether MnO has been included or not. If there are doubts on what is z , we would have to admit that the information available in the HVD is only one-dimensional, i.e. the ratio x/y . Assuming that z is well-defined, the information from the data used in HVDs is 2-dimensional. Assuming that (x, y, z) is a composition, principle A states that the used units (even molar concentrations) should be irrelevant. HVD contains all the compositional information concerning (x, y, z) . Therefore, all what can be seen in a variation diagram can be obtained from alternative plots (ternary diagrams, simplex-coordinate plots, log-scale plots, etc.). The question is which features of the data are enhanced in each representation. The use of HVD, according to Cortés (2009), is based on two points: the large experience evaluating them (mainly qualitative); and the possible application of the lever rule (quantitative). Provided that the HVD's are normalised (z is well-defined, units are completely specified), the experience might be obtained using alternative equally informative plots. Therefore, what the variation diagram really enhances is the application of the lever rule as stated in Cortés (2009) through the two lemmas in that contribution.

4 On the Lever Rule

The letter by Cortés (2009) is valuable, as it clearly states the so-called lever rule. This clarity then prompts critical comments in some points which are addressed below.

Conservation of mass Cortés (2009) mentions a relationship of the lever rule with the principle of conservation of mass. In fact, the lever rule relays on the assumption that total mass, melt plus crystal fractions, is an invariant, i.e. the system is closed. The data, being compositional (Principle A), cannot support this assumption, and the conservation of mass should be considered an external information. This means that what in the HVD appears as collinear data may represent a process that does not preserve the total mass; and conversely, non-aligned data can come from processes with constant mass.

Lever rule on a variation diagram As stated in Cortés (2009), assume that E is the point in the HVD that represents the extracted crystallised composition (x_c, y_c, z_c) , A represents the original magma composition, and B the resulting magma composition after extracting the crystallised materials. The observable composition E gives us the information related to the stoichiometry of x and y in the process of fractionation. Assuming that the total mass of the initial magma is M (invariant), and according to the lever rule, A , B , E are co-linear in the HVD (Cortés 2009, Lemma 1). The lever rule can be stated as (Cortés 2009, Lemma 2)

$$\frac{d(E, A)}{d(E, B)} = \frac{M - N}{M}, \quad (1)$$

where N is the mass of the crystallised fraction and d a distance. The rule itself does not give any information about mass but about a ratio of masses. Moreover, Cortés (2009) states that d represents the distance between points of the diagram in the sense of the Manhattan, taxicab or 1-distance given by

$$d^k(A_1, A_2) = |x_1 - x_2|^k + |y_1 - y_2|^k, \quad (2)$$

with $k = 1$. However, the proof of this result also holds for any positive with $k \geq 1$, following the method proposed by Cortés (2009, Lemma 2). The conclusion is that the lever rule is not inherent to the 1-distance ($k = 1$ in (2)) or the Euclidean distance ($k = 2$) but to a larger class of distances, called Minkowsky distances ($k \geq 1$).

Relevance of the distances The distance in (2) for any $k \geq 1$, are distances between two part sub-compositions with components (x, y) . These distances do not depend only on the ratio x/y but also on z , thus violating principle C. These distances also violate the principle A of scale invariance, i.e. when the parts are multiplied by a positive constant they change accordingly. However, the ratio of distances in (1) is scale invariant: The relevant quantity is not the distance itself (Euclidean or not) but their ratio which evaluates a balance (ratio) of masses. These facts put in question the claim that Euclidean geometry in the HVD is relevant for the analysis of balance (ratio) of masses.

Hypothesis underlying the lever rule In order to use the lever rule, the underlying hypotheses should be examined. As mentioned previously, the main hypothesis is that the total mass M is constant along the process of fractionation. It should be assumed that the composition of the magma is actually observed, both the original magma (point A) and the final magma after fractionation (point B). Cortés (2009), in section “Example”, assumes that the whole-rock chemistry represents the composition of the magma, however, other fractionation processes may have been involved and they can severely distort the appreciation of the magma composition. Moreover, what the masses M and N or $M - N$ (initial mass of magma, extracted mass of fractionated crystals, remaining mass of magma) are is unclear. To be rigorous, the total mass of the components under consideration of the initial magma M should be determined, i.e. not mixing with the surrounding materials; and this mass, itself, should be perfectly homogeneous. Certainly, it is difficult to know M . To find approximately

N , one needs many spatially distributed samples and an extensive statistical analysis. Consequently, N is seldom known. Therefore, the proportion $(M - N)/M$ as predicted by the lever rule (1) is just a rough approximation. One must be aware that the only information conveyed by points A , B or E refers to the composition considered. This does not invalidate the use of the lever rule and may give some insight into the conditions in which the fractionation occurred. However, this is far from the claim that HVDs provide vital information (Cortés 2009).

5 Conclusions

From this author's perspective, Harker variation diagrams are simply a way of representing a 3-part composition in a 2-dimensional plot. The assumed usefulness of such diagrams comes more from the experience of geologists in its use than in the way of presenting the available information. The information is just that conveyed by a 3-part composition, and the particular standardisation of the diagram. When stoichiometry of a fractionation is well-known and the composition of the original and the final melt material is measured, the lever rule can be applied to estimate the proportion of the masses of melt material prior to and resulting from a fractionation. However, the hypotheses underlying the lever rule are seldom checked and the validity of the resulting estimated proportion of masses may be doubtful. This is the main value of the variation diagram. The lever rule is not exclusively linked to the 1-distance, or taxicab distance, in the variation diagram. The ordinary, Euclidean distance and other distances, as shown in (2), satisfy the same relationship.

The conservation of mass is not related to the distance used in the variation diagram but to the hypothesis necessary to establish the lever rule. The initial mass of the considered components in the magma is assumed invariant throughout the fractionation process so that the crystallised material, plus the mass of the remaining magma, is constant. This invariance may be seriously distorted if there is not an active mixing in the magma, preserving the compositional homogeneity during the fractionation process. Local heterogeneities result in a change of the equivalent total mass. These points make the application of the lever principle subjective and difficult and, consequently, doubtful. The analysis of the variability of the observations (compositions) is thus relevant and requires tools appropriate for a statistical analysis. Taking into account the mentioned limitations, I agree with Cortés (2009) with respect to the possible use of HVDs as an exploratory tool, and for figuring out features of geochemical data thus profiting from the long experience of their use in geology.

References

- Aitchison J (1986) The statistical analysis of compositional data. Monographs on statistics and applied probability. Chapman & Hall, London. Reprinted in 2003 with additional material by The Blackburn Press, 416 p
- Aitchison J, Egozcue JJ (2005) Compositional data analysis: where are we and where should we be heading? *Math Geol* 37(7):829–850
- Barceló-Vidal C, Martín-Fernández JA, Pawlowsky-Glahn V (2001) Mathematical foundations of compositional data analysis. In Ross G (ed) Proceedings of IAMG'01—The sixth annual conference of the international association for mathematical geology, 20 p, CD-ROM

- Buccianti A, Pawlowsky-Glahn V (2005) New perspectives on water chemistry and compositional data analysis. *Math Geol* 37(7):703–727
- Chayes F (1960) On correlation between variables of constant sum. *J Geophys Res* 65(12):4185–4193
- Cortés JA (2009) On the Harker Variation Diagrams; A Comment on “The Statistical Analysis of Compositional Data. Where Are We and Where Should We Be Heading?” by Aitchison and Egozcue (2005). *Math Geosci*. doi:[10.1007/s11004-009-9222-8](https://doi.org/10.1007/s11004-009-9222-8)
- Martín-Fernández JA, Barceló-Vidal C, Pawlowsky-Glahn V (2003) Dealing with zeros and missing values in compositional data sets using nonparametric imputation. *Math Geol* 35(3):253–278