



Compositional data analysis in geochemistry: Are we sure to see what really occurs during natural processes?



A. Buccianti^{a,*}, E. Grunsky^b

^a Department of Earth Sciences, University of Florence, Italy

^b Geological Survey of Canada, Ottawa, Ontario K1A 0E8, Canada

ARTICLE INFO

Article history:

Received 11 March 2014

Accepted 17 March 2014

Available online 26 March 2014

Keywords:

Compositional data analysis

Geochemical data

Simplex geometry

Log-ratio approach

Environmental modelling

ABSTRACT

Geochemical data are typically reported as compositions, in the form of some proportions such as weight percents, parts per million, etc., subject to a constant sum (e.g. 100%, 1,000,000 ppm). This latter implies that such data are “closed”; that is, for a composition of D -components, only $D - 1$ components are required. The statistical analysis of compositional data has been a major issue for more than 100 years. The problem of spurious correlation, introduced by Karl Pearson in 1897, affects all data measuring parts of some whole, which are by definition, constrained; and such type of measurements are present in all fields of geochemical research. The use of the log-ratio transform was introduced by John Aitchison to overcome these constraints by opening the data into the real number space, within which standard statistical methods can be applied. However, many statisticians and users of statistics in the field of geochemistry are unaware of the problems affecting compositional data, as well as solutions that overcome these problems. A look into the ISI Web of Science and Scopus databases shows that most papers where compositional data are the core of a geochemical research continue to ignore methods to correctly manage constrained data. A key question is how we can demonstrate that the interpretation of the behaviour of chemical species in natural environment and in geochemical processes is improved when the compositional constraint of geochemical data is taken into account through the use of new methods. In order to achieve this aim, this special issue of the Journal of Geochemical Exploration focuses on the correct statistical analysis of compositional data. Applications in exploration, monitoring and environments by considering several geological matrices are presented and discussed illustrating that several paths can be followed to understand how geochemical processes work.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

More than 100 years have elapsed since Karl Pearson wrote his paper on spurious correlation in 1897 (Pearson, 1897; Fig. 1) and more than 30 years from a first solution based on log-ratios proposed by John Aitchison (Aitchison, 1982; Fig. 2). Since then, the approach has been characterised by many studies on the natural geometry of the sample space where compositional data are positioned. In order to understand the meaning of these words, it is necessary to link *geochemistry* to *geometry*, two fields of research apparently distant but, in reality, closely linked.

The *geometry* of a composition is the metric of the *sample space*. For example, when we measure concentrations of some geological material in laboratory we do not expect to find negative values, only positive values from which an interpretation is based. The sample space is

where compositional values are located. Compositions are compared by measuring their distance and translations along linear or non-linear trends. It is in this sample space where *random variables*, the mathematical rules to attribute probability to the occurrences of events, are defined for statistical inference.

Some sample spaces may be better than others to exploit the information contained in the data. This is the case of compositional data where the elements of the composition are non-negative and sum up to a constant, e.g. to unity since they have been scaled by the total of the components as a standardization practice. A consequence of this is that a composition of D parts, $[x_1, x_2, \dots, x_D]$, can be identified with a closed vector

$$\mathbf{x} = C[x_1, x_2, \dots, x_D] = \left[\frac{x_1 \cdot k}{\sum_{i=1}^D x_i}, \frac{x_2 \cdot k}{\sum_{i=1}^D x_i}, \dots, \frac{x_D \cdot k}{\sum_{i=1}^D x_i} \right], \quad (1)$$

where C is called the closure operation to the constant k (Aitchison, 1986).

* Corresponding author. Tel.: +39 0552757493; fax: +39 055284571.
E-mail address: antonella.buccianti@unifi.it (A. Buccianti).



Fig. 1. Karl Pearson (27 March 1857–27 April 1936) the scientist who founded the discipline of mathematical statistics.

The set of real positive vectors closed to a constant k constitutes the constrained sample space called simplex of D parts, denoted by S^D and defined as

$$S^D = \{(x_1, x_2, \dots, x_D) : x_1 > 0, x_2 > 0, \dots, x_D > 0; x_1 + x_2 + \dots + x_D = k\}. \quad (2)$$

Note that geochemical data are always non-negative and are restricted to the positive part of the real sample space, R_+^D . To be noticed here is that the previous approach gives importance to the sample space (sum constraint) but compositional data cannot be a priori closed. In most situations is the analyst who decides that the total of each sample is not relevant and then normalise the data to proportions. All the sets of data are equivalence classes from a mathematical point of view (Buccianti and Pawlowsky-Glahn, 2005).

The key in understanding compositional data relies on defining a correspondence between the simplex S^D and R^D , the real space governed by Euclidean geometry, through the use of a metric where classical statistics can be applied for an unbiased interpretation of the relationships and patterns of geochemical data.

The equivalence between S^D and R^D is obtained by defining equivalent operations in S^D . The definition of the operations of sum (difference) and product, called *perturbation* and *powering*, together with the definition of other properties (*norm*, *distance*, *inner product*) allow us to consider S^D as a space with a structure governed by the Euclidean geometry completely equivalent to the geometry of the correspondent

unrestricted real space with $D - 1$ dimensions (Billheimer et al., 2001; Egozcue et al., 2003; Egozcue and Pawlowsky-Glahn, 2006).

2. The algebraic–geometric structure of the simplex

Two fundamental operations of change govern the algebraic–geometric structure (the metric) of the simplex of D -parts, **perturbation**, denoted by the symbol \oplus , and **powering** denoted by the symbol \odot (Aitchison, 1986). Consider two compositions \mathbf{x} and \mathbf{y} . To perturb \mathbf{x} by \mathbf{y} , first calculate the component-wise product and then close the result to 100 (C closure operator) to produce \mathbf{z} :

$$\mathbf{x} \oplus \mathbf{y} = C[x_1 \cdot y_1, x_2 \cdot y_2, \dots, x_D y_D] = \mathbf{z}. \quad (3)$$

It is then clear that

$$\mathbf{z} \odot \mathbf{y} = C\left[\frac{z_1}{y_1}, \frac{z_2}{y_2}, \dots, \frac{z_D}{y_D}\right] = \mathbf{x}. \quad (4)$$

Consider now a real number a ; the power-transformed composition is given by

$$a \odot \mathbf{x} = C[x_1^a, x_2^a, \dots, x_D^a]. \quad (5)$$

The used symbols \oplus and \odot emphasize the analogy with the operations of displacement or translation and scalar multiplication of vectors in R^D . Perturbation clearly corresponds to addition in R^D , while powering is the multiplication. The internal (\oplus) and external (\odot) operations define a vector or linear space structure on S^D , a structure that can be extended by the introduction of the simplicial metric or Aitchison distance:

$$d_a(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^D \left\{ \log \frac{x_i}{g(\mathbf{x})} - \log \frac{y_i}{g(\mathbf{y})} \right\}^2}, \quad (6)$$

where $g(\cdot)$ is the geometric mean of the parts of the composition, $g(\mathbf{x}) = (\prod_{i=1}^D x_i)^{1/D}$.

It is possible to demonstrate that this metric satisfies the usual metric axioms and other desirable properties. The definition of distance, together with that of norm (the distance of \mathbf{x} from the origin of a linear space), the inner product (the cosine of the angle between compositional vectors) and the operations of perturbation and powering, provide a Euclidean structure to the simplex, now called Aitchison simplicial geometry (Pawlowsky-Glahn and Egozcue, 2001).

Within this framework it is possible to work inside the simplex to perform an analysis of compositional data free of inconsistencies.

An important consequence of this approach is how the concept of distance changes from S^D , with the Aitchison distance d_a reported in Eq. (6) to the Euclidean distance d_e in R^{D-1} defined as

$$d_e(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_{D-1} - y_{D-1})^2}. \quad (7)$$

Eqs. (6) and (7) measure the same magnitude but in different sample spaces. Applying Eq. (7) in a sample space that supports, for the same concept of distance, Eq. (6) is an error that can lead to misleading results. Similar scenario appears for spherical data, where people agree to use the angular distance.

However, compositional data can also be transformed to go out of the simplex S^D and into the unconstrained real space R^{D-1} by applying one of the proposed log-ratio transformations (Aitchison, 1982).

3. The log-ratio transformation

As explained in Aitchison et al. (2000) the log-ratio transformations allow us to make meaningful statements on compositional data only involving ratios of components. The first principle of compositional data



Fig. 2. John Aitchison in occasion of the Codawork 2005 in Girona (E).

analysis is scale invariance (vectors with proportional positive components represent the same composition, i.e. compositional data are unit-free). The second key principle met by using log-ratios is subcompositional coherence, which states that inferences about subcompositions or the full composition should be consistent, giving the same result about the relationships between the parts.

Aitchison (1982) introduced proper and complete representations of a composition in order to have all the relevant information contained in a set of log-ratios. Applications in several fields of sciences can be found in Buccianti et al. (2006) and Pawlowsky-Glahn and Buccianti (2011).

A first choice was the *additive log-ratio* (alr). For a composition \mathbf{x} in the D -part simplex S^D , it is defined as

$$alr(\mathbf{x}) = \log\left(\frac{x_1}{x_D}, \frac{x_2}{x_D}, \dots, \frac{x_{D-1}}{x_D}\right). \quad (8)$$

Although useful for parametric modelling, the alr has the inconvenience of not being invariant under permutation of the components, one part is chosen as the common denominator of all the ratios, and the application of some statistical procedures can be compromised if their results depend on such part. On the other hand, the alr transform is not isometric between the simplex and the real space, which implies that e.g. distances are not equivalent in both spaces. Consequently, Aitchison (1986) introduced the *centered log-ratio transformation* (clr), which represents a D -part composition using D clr-coefficients. It is defined as

$$clr(\mathbf{x}) = \log\left[\frac{x_1}{g(\mathbf{x})}, \frac{x_2}{g(\mathbf{x})}, \dots, \frac{x_D}{g(\mathbf{x})}\right], \quad (10)$$

with $g(\mathbf{x})$ is the geometric mean of the parts of the composition.

The clr, however, does not allow for parametric modelling. Egozcue et al. (2003) proposed the *isometric log-ratio transformation* (ilr) developing the principle of working with orthonormal bases and their corresponding coordinates. To understand this point, considers that S^D is a vector space of dimension $D - 1$. Thus, $D - 1$ perturbation-independent vectors in S^D constitute a basis. If these vectors are unitary and mutually orthogonal they form an orthonormal basis of S^D . In real vector space an example of an orthonormal basis is easily found. For example in R^3 the vectors (1,0,0), (0,1,0) and (0,0,1) form an orthonormal basis, known as the canonical basis of the real space. Although it is not so simple in S^D , some special orthonormal bases can be derived from a *Sequential Binary Partition* (SBP) of a compositional vector (Egozcue and Pawlowsky-Glahn, 2005). This represents a practical way of defining orthonormal basis and associated coordinates in the simplex. Each step of the partition, out of $D - 1$, gives rise to an ilr coordinate called a *balance*, which may be easier to interpret in practical terms. In a first step, a SBP consists of dividing the composition into two groups of parts marked with +1 and -1 (for example anions and cations in water chemistry). In subsequent steps, each previously obtained group of parts is again subdivided into two groups, always marked by +1 and -1, until all groups are made of a single part (Table 1). These ilr coordinates, or balances are given by

$$b_j = \sqrt{\frac{rs}{r+s}} \log \frac{g(\mathbf{x}_+)}{g(\mathbf{x}_-)} \quad (11)$$

where $g(\cdot)$ is the geometric mean of the argument, \mathbf{x}_+ is the group with r parts marked with +1 and \mathbf{x}_- the group of s parts marked with a -1. Balances may have a relatively easy interpretation, as they are log-ratios of geometric means of groups of parts.

Table 1

Sequential binary partition. Application example for the main anions and cations in water chemistry, obtaining 7 balances.

| | HCO ₃ ⁻ | SO ₄ ²⁻ | Cl ⁻ | Ca ²⁺ | Mg ²⁺ | Na ⁺ | K ⁺ |
|-------|-------------------------------|-------------------------------|-----------------|------------------|------------------|-----------------|----------------|
| b_1 | +1 | +1 | +1 | -1 | -1 | -1 | -1 |
| b_2 | +1 | -1 | -1 | 0 | 0 | 0 | 0 |
| b_3 | 0 | +1 | -1 | 0 | 0 | 0 | 0 |
| b_4 | 0 | 0 | 0 | +1 | +1 | -1 | -1 |
| b_5 | 0 | 0 | 0 | +1 | -1 | 0 | 0 |
| b_6 | 0 | 0 | 0 | +1 | -1 | 0 | 0 |
| b_7 | 0 | 0 | 0 | 0 | 0 | +1 | -1 |

4. Centre and variability

An important consequence of the nature of compositional data is that standard descriptive statistics, in particular the arithmetic mean and the variance or standard deviation of individual components are not coherent with the Aitchison geometry as measures of central tendency and dispersion respectively. Alternatives have to be used and a fundamental role is played by the concepts of *centre* (Aitchison, 1997), *variation array* and *total variance* (Aitchison, 1986, 1997).

A measure of central tendency for a compositional data set \mathbf{x} , called *centre*, is the closed geometric mean defined as

$$g = C[g_1, g_2, \dots, g_D] \quad (12)$$

with

$$g_i = \left(\prod_{j=1}^n x_{ij}\right)^{1/n}, i = 1, 2, \dots, D \quad (13)$$

and C the closure operator to constant k and n number of cases. It is important to note that the geometric mean is considered column-wise (i.e. by parts).

A way to describe an association between parts in a compositional data set \mathbf{X} is by means of the *variation array*, a symmetrical matrix where the upper diagonal contains the log-ratio variances and the lower diagonal contains the log-ratio means. That is, the ij -th component of the upper diagonal is $\text{var}[\log(x_i/x_j)]$ and the ij -component of the lower diagonal is the expectation of $\log(x_i/x_j)$, where $i, j = 1, 2, \dots, D$. A measure of the total relative variability of a compositional matrix \mathbf{X} is given by the *total variance* of the data set, i.e. the sum of all the elements of the variation matrix \mathbf{T} divided by $2D$ (Daunis-i-Estadella et al., 2006):

$$\mathbf{T} = \begin{bmatrix} \text{var}\left(\log \frac{x_1}{x_1}\right) & \dots & \text{var}\left(\log \frac{x_1}{x_D}\right) \\ \vdots & \ddots & \vdots \\ \text{var}\left(\log \frac{x_D}{x_1}\right) & \dots & \text{var}\left(\log \frac{x_D}{x_D}\right) \end{bmatrix} \quad (14)$$

$$\text{totvar}[\mathbf{X}] = \frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \text{var}\left[\log \frac{x_i}{x_j}\right] \quad (15)$$

Free software is available to carry out compositional statistical analysis in both Windows and Mac platforms, or through specialised packages for the R statistical environment (Thió-Henestrosa and Daunis-i-Estadella, 2011; Templ et al., 2011). These CoDA tools enable anyone to process their data in a way that provides insight into the variability of geochemical processes without geometric distortion or numerical bias.

5. Why it is necessary to reconsider how geochemical data are evaluated

The basic idea to work correctly with compositional data is to consider that the sample space has a natural geometry coherent with the

intuitive concept of difference associated to this particular type of data, a concept related to perturbation difference operation. Consider for example that for compositional data it turns out that 5% is half of 10%, while 45% is 0.9 of 50%. If we perform our analysis in the Euclidean real space, by adopting univariate or multivariate descriptive or inferential methods, the difference is considered the same as described by Eq. 7.

However, geochemists know that natural processes are not well described in this way. The error is to consider this condition a “data problem”, when it is actually a problem with the statistical methods used. In fact, it is expected of statistics to be able to capture our perception of the natural scale of the data. Now this possibility is offered by the developments of the Compositional Data Analysis theory (CoDA), and these developments provide an opportunity for a better understanding of natural phenomena (Buccianti, 2011). However the path of new ideas through a consolidated thought is always difficult. Alain Connes, Fields award in 1982 in “A view of Mathematics” wrote: “It often begins by an act of rebellion with respect to the existing dogmatic description of that reality that one will find in existing books” (Connes, 1982).

6. The structure of the Special Issue

This special issue provides an opportunity to illustrate how several scientists have investigated geochemical systems or processes using compositional data analysis methods.

The eight papers that constitute the issue cover different fields of investigation but all of the data are compositional in nature.

The first one, by Wang et al. (2014) deals with problems related to the mapping of Fe mineralization, a typical problem of geochemical exploration. The geochemical signatures of stream sediment data from eastern Tianshan (China) were analysed by using the log-ratio approach. The paper by Grunsky et al. (2014) investigates the lake sediment geochemistry of the Melville Peninsula (Canada) by multivariate methods thus developing tools for predictive geological mapping. In fact, the treatment of lake sediment geochemical data within the compositional framework of log-ratio analysis, and the use of principal component analysis, analysis of variance, linear discriminant analysis and spatial analysis with ordinary kriging, describes both qualitative and quantitative mechanisms for predictive lithological mapping.

The paper by Tolosana-Delgado and van den Boogaart (2014) is also focused on the analysis of spatial data. The contribution explores a compositionally compliant approach to potential mapping using geochemical evidence, in which a log-ratio transformation is applied to the whole set of components before any potential mapping technique is used.

The paper by Parent et al. (2014) examines the biogeochemistry of inorganic and organic phosphorus in soil, an interesting item by considering the perturbation imposed on the natural cycle of the element from the anthropogenic activities. The compositional approach used by the authors is based on the ilr transformation obtained from compositional balances between the groups of components. The methodology develops independent coordinates in the real space from balances defined according to sound knowledge of the contributions made by each component to the biogeochemical processes.

Linking balances to chemical reactions and, in particular, compositional data analysis with thermodynamic geochemical modelling, is proposed by Engle and Blondes (2014) in their investigation of sedimentary basin brines, a geological material useful to reconstruct the hydrogeochemical history of hydrocarbon producing formations. Results based on the use of the isometric log-ratio (ilr) transformation allow for the construction of a hydrogeologic model in which meteoric inflow occurs primarily by vertical transport from overlying units, rather than within unit recharge.

The paper of Palarea-Albaladejo et al. (2014) focuses on compositional data matrices characterised by the presence of data reported as less than the lower limit of detection. The features of different methods of imputation are discussed, coherently with the properties of compositional

data, with application examples illustrated by the use of computer routines for the R open source platform.

Water chemistry data characterised by the presence of values below the lower limit of detection for nitrogen species was also investigated by Buccianti et al. (2014). The authors present the effects of different methods of imputation, both on the relationships among the species of nitrogen and on the full composition, given by the main anions and cations of the aqueous solutions. The paper clearly demonstrates that the elimination of data below the lower limit of detection changes the perception of the background value of an area, an important item when nitrate has to be monitored for health purposes.

The special issue concludes with the work of Zeng et al. (2014), where a typical petrological problem concerning the Lange's plagioclase-liquid hygrometer/thermometer is revised in the light of the compositional approach, with implications for models of experiments with mixtures.

The papers presented in this special issue of compositional data analysis in geochemistry cover several fields of investigation and highlight the need to address some unresolved problems. In this respect, the future potential of the CoDA approach has been examined with a broad overview of compositional data analysis methods and applications.

We hope that you will enjoy and find benefit in the papers presented for this Special Issue!

We kindly acknowledge the chief editors of the journal, Benedetto de Vivo and Robert Ayuso, for the opportunity to compile this special issue.

The Guest Editors.

Acknowledgments

We kindly acknowledge the chief editors of the journal, Benedetto de Vivo and Robert Ayuso, for the opportunity to compile this special issue. Javier Palarea-Albaladejo and Josep Antoni Martín-Fernández are thanked for the revision of this editorial and for their valuable suggestions and comments.

References

- Aitchison, J., 1982. The statistical analysis of compositional data (with discussion). *J. R. Stat. Soc. Ser. B (Statistical Methodology)* 44 (2), 139–177.
- Aitchison, J., 1986. The statistical analysis of compositional data. *Monographs on Statistics and Applied Probability* Chapman & Hall Ltd., London. Reprinted (2003) with additional material by The Blackburn press, Caldwell, NJ.
- Aitchison, J., 1997. The one-hour course in compositional data analysis or compositional data analysis is simple. In: Pawłowsky-Glahn, V. (Ed.), *Proceedings of IAMG'97 – The III Annual Conference of the International Association for Mathematical Geology. International centre for Numerical Methods in Engineering (CIMNE), Barcelona (E)*, pp. 3–35 (Vol. I, II and addendum).
- Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J., Pawłowsky-Glahn, V., 2000. Logratio analysis and compositional distance. *Math. Geol.* 32 (3), 271–275.
- Billheimer, D., Guttorm, P., Fagan, W., 2001. Statistical interpretation of species composition. *J. Am. Stat. Assoc.* 96 (456), 1205–1214.
- Buccianti, A., 2011. Natural laws governing the distribution of the elements in geochemistry: the role of the log-ratio approach. In: Pawłowsky-Glahn, Buccianti (Eds.), *Compositional Data Analysis. Theory and Applications*. John Wiley & Sons, pp. 255–266.
- Buccianti, A., Nisi, B., Martín-Fernández, J.A., Palarea-Albaladejo, J., 2014. Methods to investigate the geochemistry of groundwaters with values for nitrogen compounds below the detection limit. *J. Geochem. Explor.* <http://dx.doi.org/10.1016/j.gexplo.2014.01.014>.
- Buccianti, A., Pawłowsky-Glahn, V., 2005. New perspectives on water chemistry and compositional data analysis. *Mathematical Geology* 37 (7), 703–727.
- Compositional data analysis in the geosciences 2006. In: Buccianti, A., Mateu-Figueras, G., Pawłowsky-Glahn, V. (Eds.), *From theory to Practice*. Geological Society, London, Special Publication 264.
- Connes, A., 1982. A view of mathematics. *Encyclopedia of Life Support Systems (EOLSS)*, UNESCO chapter. http://www.omath.org.il/image/users/112431/ftp/my_files/Alain%20connes.pdf?id=8883062.
- Daunis-I-Estadella, J., Barceló-Vidal, C., Buccianti, A., 2006. Exploratory compositional data analysis. In: Buccianti, A., Mateu-Figueras, G., Pawłowsky-Glahn, V. (Eds.), *Compositional data Analysis in the Geosciences: from theory to practice*. Geological Society, London, Special Publications, 264, pp. 161–174.
- Egozcue, J.J., Pawłowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C., 2003. Isometric logratio transformation for compositional data analysis. *Math. Geol.* 35 (3), 279–300.
- Egozcue, J.J., Pawłowsky-Glahn, V., 2005. Groups of parts and their balances in compositional data analysis. *Math. Geol.* 37 (7), 795–828.

- Egozcue, J.J., Pawłowsky-Glahn, V., 2006. Simplicial geometry for compositional data. In: Buccianti, A., Mateu-Figueras, G., Pawłowsky-Glahn, V. (Eds.), *Compositional data Analysis in the Geosciences: from theory to practice*. Geological Society, London, Special Publications, 264, pp. 145–159.
- Engle, M.A., Blondes, M.S., 2014. Linking compositional data analysis with thermodynamic geochemical modeling: oilfield brines from the Permian Basin, USA. *J. Geochem. Explor.* <http://dx.doi.org/10.1016/j.gexplo.2014.02.025>.
- Grunsky, E.C., Mueller, U.A., Corrigan, D., 2014. A study of the lake sediment geochemistry of the Melville Peninsula using multivariate methods: applications for predictive geological mapping. *J. Geochem. Explor.* <http://dx.doi.org/10.1016/j.gexplo.2013.07.013>.
- Palarea-Albaladejo, J., Martín-Fernández, J.A., Buccianti, A., 2014. Compositional methods for estimating elemental concentrations below the limit of detection in practice using R. *J. Geochem. Explor.* <http://dx.doi.org/10.1016/j.gexplo.2013.09.003>.
- Parent, L.E., Parent, S.E., Ziadi, N., 2014. Biogeochemistry of soil inorganic and organic phosphorus: a compositional analysis with balances. *J. Geochem. Explor.* <http://dx.doi.org/10.1016/j.gexplo.2014.01.030>.
- Pawłowsky-Glahn, V., Egozcue, J.J., 2001. Geometric approach to statistical analysis on the simplex. *Stoch. Env. Res. Risk A. (SERRA)* 15 (5), 384–398.
- Pawłowsky-Glahn, V., Buccianti, A., 2011. *Compositional data analysis. Theory and Application* Wiley & Sons Ltd Publication, UK.
- 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 measurement of organs. *Proceedings of the Royal Society of London, LX* pp. 489–502.
- Templ, M., Hron, K., Filzmoser, P., 2011. robCompositions: an R-package for robust statistical analysis of compositional data. In: Pawłowsky-Glahn, Buccianti (Ed.), *Compositional Data Analysis. Theory and Applications*. John Wiley & Sons, pp. 341–355.
- Thió-Henestrosa, S., Daunis-i-Estadella, J., 2011. Exploratory analysis using CoDaPack 3D. In: Pawłowsky-Glahn, Buccianti (Eds.), *Compositional Data Analysis. Theory and Applications*. John Wiley & Sons, pp. 329–340.
- Tolosana-Delgado, R., van den Boogaart, Gerald K., 2014. Towards compositional geochemical potential mapping. *J. Geochem. Explor.* <http://dx.doi.org/10.1016/j.gexplo.2014.02.026>.
- Wang, W., Zhao, J., Cheng, Q., 2014. Mapping of Fe mineralization-associated geochemical signatures using logratio transformed stream sediment geochemical data in eastern Tianshan, China. *J. Geochem. Explor.* <http://dx.doi.org/10.1016/j.gexplo.2013.11.008>.
- Zeng, L., Cheng, L., Cheng, Q., Zhang, S., 2014. A refinement of Lange's plagioclase–liquid hygrometer/thermometer based on quadratic log-contrast models for experiments with mixtures. *J. Geochem. Explor.* <http://dx.doi.org/10.1016/j.gexplo.2013.12.008>.