Variation diagrams to statistically model the behavior of geochemical variables: Theory and applications

A. Buccianti [a,*], J.J. Egozcue [b], V. Pawlowsky-Glahn [c]

[a] Dip. di Scienze della Terra, Università degli Studi di Firenze, Firenze, Italy
[b] Dep. Matemàtica Aplicada III, Univ. Politècnica de Catalunya, Barcelona, Spain
[c] Dept. d'Informàtica, Matemàtica Aplicada, i Estadística, Universitat de Girona, Girona, Spain

A R T I C L E   I N F O

S U M M A R Y

Variation diagrams are revisited from a compositional perspective. An alternative, consistent with the Aitchison geometry of the simplex, is proposed. The new multiple binary diagrams compare a reference balance, b_1, fixed on the x-axis, and a set of secondary balances, b_2, on the y-axis. The diagrams fulfill compositional principles of invariance under change of scale, guaranteeing consistency of interpretations and working better than diagrams based on raw concentrations of chemical species. The new diagrams allow us to investigate the nature of compositional changes, and inform about the elements or chemical species responsible for the variation when data are a priori ranked following some compositional (or non-compositional) criteria. An application example using the chemistry of large rivers illustrates the procedure. Data are extracted from the Global Register of river inputs, which contains the water quality of about 550 rivers having exorheic basins exceeding $10{,}000$ km^2 and water discharge exceeding 10 km^3/yr. The chemistry of only 55 large rivers of the world (in μmol/L, content of HCO_3^-, Cl^-, SO_4^{2-}, K^+, Na^+, Mg^{2+}, Ca^{2+} and $SiO_{2(aq)}$) as well as their discharge values are considered. The choice is based on the completeness of analytical data and representativeness with respect to water/rocks interaction processes. Results indicate that only the balances $b_1(Ca^{2+})$ and $b_2(HCO_3^-)$ have a high, significant correlation. They are the only two chemical components able to trace weathering processes on this large scale. On the contrary, large data scattering is found in all other cases, indicating a higher sensibility to different environmental parameters, like local lithology, climate or pollution. When all the balances are analyzed as a function of discharge values very low correlation is found, indicating that the chemostatic behavior appears to be a common condition. When samples are ranked for increasing values of $b_1(Ca^{2+})$, or alternatively by discharge values, the statistical investigation of chemical changes reveals interesting features of the analyzed compositions.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

A table of geochemical data—related to the chemistry of some solid, liquid or gaseous material—may show an almost incomprehensible variation in the concentration of individual components, elements or chemical species. However, since the samples are likely to be geochemically related, a major goal for investigators is to find a way in which the variation between individual samples can be simplified and condensed. The tool which is most commonly used, often considered an invaluable resource, is the variation diagram. Variation diagrams often provide evidence about the behavior of

chemical elements or species involved in geochemical processes, but generally the illustration of the phenomena remains at a descriptive step. However, when statistics is applied, misleading results can be obtained (Aitchison and Egozcue, 2005; Egozcue, 2009). These diagrams are constructed for a sample space whose geometry is not compatible with the properties of compositional data. Compositional data are by definition parts of some given numerical total which only carry relative information between them (Aitchison, 1986). They are usually represented as data that sum to a constant, like 100 for percentages or 10^6 for ppm, making the so called closure effect apparent (Pearson, 1897; Chayes, 1960; Aitchison, 1982). As a consequence, the proportion of one component will change, possibly without any relationship to any geochemical process. When variables are linked by this internal rule, changes cannot be visualized on a Cartesian plane defined by two

orthogonal axes. If we want to capture all the information contained in compositional data, with the aim to better understand how geochemical processes are working, new tools have to be exploited (Pawlowsky-Glahn and Egozcue, 2001; Pawlowsky-Glahn and Buccianti, 2011; Buccianti, 2013; Engle and Rowan, 2013).

This contribution concerns the development of variation diagrams coherent with the algebraic–geometric properties of compositional data summarized in Egozcue and Pawlowsky-Glahn (2006a). Diagrams are constructed representing on the horizontal axis a variable whose positive changes (increase) can be associated with the persistent action of some geochemical process. Thus, the *x*-axis represents a reference term against which the behavior of all the other terms of the composition can be investigated. This variable, acting as a very simple index, can be easily substituted by more complex combinations of parts and external variables preserving coherence with the properties of compositional data.

The paper is structured as follows. In Section 2 the use of variation diagrams in different fields of the Geosciences is revisited. Section 3 is divided in two: Section 3.1 introduces the database of the Global Register of River Inputs, from which the data used for illustration have been extracted, and SubSection 3.2 presents a summary of the Aitchison geometry of the simplex, the sample space of compositional data. Section 4 presents in detail the new variation diagrams for compositional data. The chemistry of 55 large rivers of the world is used in Section 5 for illustration of the proposed alternative to the variation diagram.

2. Variation diagrams in the Geosciences. A review

2.1. General comments

A variation diagram is a bivariate graph, or scattergram, on which two selected components are plotted (Rollinson, 1993). The peculiarity of these diagrams is that on the horizontal axis a variable is often reported whose increase, compared to all other terms of the composition, is related to the evolution of natural processes. It thus represents a reference term. Diagrams like these have been popular for the investigations of rocks since 1909, when Alfred Harker proposed their use (Harker, 1909; Rollinson, 1993). Variation diagrams are able to condense and rationalize a large volume of numerical information and, qualitatively, are assumed to show the presence of correlation (positive or negative) between chemical components in a given geological material. Traditionally, the presence of strong positive correlation is associated with geochemical coherence, and suggests that there is an underlying process able to explain the observed relationship. In this context, the building up of variation diagrams with mathematical ratios (binary diagrams of molar ratios; stoichiometrical ratios with concentration expressed in meq/L for water; diagrams of one or more combined constituents versus the total concentration or the mean concentration of major and trace elements of some reservoir as, for example, the continental crust) are often considered helpful to make comparisons possible (Chetelat et al., 2013). Diagrams of the concentration of a soluble element, normalized versus Al, are often used to build up weathering indices (for example X/Al, with $X = Na, K, Mg$ or Ca) in the investigation of the partition of elements between the dissolved phase and the solid residues during weathering processes (Bouchez et al., 2012). In this respect, a discussion on the use of high top-/bottom-soil ratios, or high values of *enrichment factors* (EFs) is used as a proof for major anthropogenic impact on the geochemistry of the Earth surface as reported in Sucharovà et al. (2012). The idea behind calculating such ratios is that soils taken at depth can represent the *average crust*, thus providing the geochemical background for the soils collected at the Earth surface. Diagrams of ratios with common denominator are

frequently used in the investigation of mixing problems, particularly with isotopic ratios (Rollinson, 1993).

2.2. Variation diagrams and water chemistry

From a general point of view, the study of water chemistry using ratios has been developed in different fields of research and has evolved in time. The ratio of chloride to other ions has been used in studies of water contaminated with common salt (sodium chloride) or associated with petroleum (Schoeller, 1955). White (1960) published a set of median ratios of ion concentrations in ppm, which he believed to be representative of water of different origins. Garrels and MacKenzie (1967) used the ratio of silica to dissolved solids in identifying water influenced by the solution of silicate minerals. In some cases the type of mineral itself may be deduced from the ratios among major cations. Garrels (1967), investigating the chemistry of groundwater from igneous rocks, developed a model to explain it, based on the molar ratios Na^+/Ca^{2+} versus the molar ratio of $HCO_3^-/SiO_{2(aq)}$. Meisler and Becher (1967) used the ratio of calcium to magnesium to study water interacting with limestone and dolomite to trace seawater contamination. On the other hand, the ratio of sodium to total cations is considered to be useful in areas of natural cation exchange. Gibbs (1970) was able to give a framework to the idea of working with ratios investigating the chemical composition of rivers. In terms of cations, he considered Na^+ (high saline) and Ca^{2+} (freshwater) to be the best representatives of surface water end-members. Similarly, the anions Cl^- (high saline) and HCO_3^- (freshwater) were selected to represent surface water end members. Plots of the relative abundances of the cations and anions ($Na^+/(Na^+ + Ca^{2+})$ or $Cl^-/(Cl^- + HCO_3^-)$) versus total dissolved solids gave two diagonal fields (and a boomerang-shaped diagram) anchored by the two end-members. Based on these graphical representations, Gibbs suggested that there are three mechanisms that control the chemistry of surface water. The first is atmospheric precipitation, represented by surface water, with low total dissolved solids and relatively high Na^+ and Cl^-, that plot in the lower right portion of his diagram. The second mechanism is essentially weathering (referred to as rock dominance), in which the chemical breakdown of rocks and minerals provides the dissolved components, depending on climate, relief, and exposed rock types. The third mechanism is evaporation–crystallization, when the concentration of the total dissolved solids increases due to evaporation. An ion-ratio calculation technique using molar concentrations of Na^+, K^+ and Ca^{2+} proposed by Fournier and Truesdell (1973) has been widely used for estimating temperatures of geothermal reservoirs. Stallard and Edmond (1983), criticizing Gibbs's results, proposed a classification of river water chemistry that emphasizes the importance of rock weathering, based on the values of the molar ratios $SiO_{2(aq)}/(Na^+ + K^+)$, $Na^+/(Na^+ + Ca^{2+})$, Na^+/Cl^-, $(Ca^{2+} + Mg^{2+})/(0.5HCO_3^- + SO_4^{2-})$, the total cation charge ($\mu eq/L$) and Total Dissolved Solids (TDS, mg/L). Gaillardet et al. (1999), in their investigation of the chemistry of large rivers to estimate global silicate weathering and CO_2 consumption rates, use a number of elemental ratios focusing the attention on Na^+ normalized (common denominator) molar ratios as, e.g. Ca^{2+}/Na^+, K^+/Na^+, Mg^{2+}/Na^+, Cl^-/Na^+, SO_4^{2-}/Na^+, HCO_3^-/Na^+. Instead of absolute concentrations (being dependent on dilution and evaporation processes), elemental ratios and isotopic compositions are intensive parameters that permit the comparison between rivers draining areas of high runoff, and rivers draining arid areas. A recent application example can be found in Voss et al. (2014) for the Fraser River (Canada), where (according to the authors) the relationship between Ca^{2+}/Na^+ and Mg^{2+}/Na^+ of the dissolved load—compared with the end members (carbonate, silicate and evaporite)—indicates a dilution of carbonate-dominated weathering products