Geophysical Research Abstracts
Vol. 18, EGU2016-3143, 2016
EGU General Assembly 2016

Scaling laws and properties of compositional data

Antonella Buccianti (1), Stefano Albanese (2), AnnaMaria Lima (2), Giulia Minolfi (2), and Benedetto De Vivo (2)

(1) Università di Firenze, Dipartimento di Scienze della Terra (I), (2) Università di Napoli Federico II, Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse (I)

Many random processes occur in geochemistry. Accurate predictions of the manner in which elements or chemical species interact each other are needed to construct models able to treat presence of random components.
Geochemical variables actually observed are the consequence of several events, some of which may be poorly defined or imperfectly understood. Variables tend to change with time/space but, despite their complexity, may share specific common traits and it is possible to model them stochastically.
Description of the frequency distribution of the geochemical abundances has been an important target of research, attracting attention for at least 100 years, starting with CLARKE (1889) and continued by GOLDSCHMIDT (1933) and WEDEPOHL (1955). However, it was AHRENS (1954a,b) who focussed on the effect of skewness distributions, for example the log-normal distribution, regarded by him as a fundamental law of geochemistry.
Although modeling of frequency distributions with some probabilistic models (for example Gaussian, log-normal, Pareto) has been well discussed in several fields of application, little attention has been devoted to the features of compositional data. When compositional nature of data is taken into account, the most typical distribution models for compositions are the Dirichlet and the additive logistic normal (or normal on the simplex) (AITCHISON et al. 2003; MATEU-FIGUERAS et al. 2005; MATEU-FIGUERAS and PAWLOWSKY-GLAHN 2008; MATEU-FIGUERAS et al. 2013). As an alternative, because compositional data have to be transformed from simplex space to real space, coordinates obtained by the ilr transformation or by application of the concept of balance can be analyzed by classical methods (EGOZCUE et al. 2003).
In this contribution an approach coherent with the properties of compositional information is proposed and used to investigate the shape of the frequency distribution of compositional data. The purpose is to understand data-generation processes from the perspective of compositional theory.
The approach is based on the use of the isometric log-ratio transformation, characterized by theoretical and practical advantages, but requiring a more complex geochemical interpretation compared with the investigation of single variables. The proposed methodology directs attention to model the frequency distributions of more complex indices, linking all the terms of the composition to better represent the dynamics of geochemical processes. An example of its application is presented and discussed by considering topsoil geochemistry of Campania Region (southern Italy). The investigated multi-element data archive contains, among others, Al, As, B, Ba, Ca, Co, Cr, Cu, Fe, K, La, Mg, Mn, Mo, Na, Ni, P, Pb, Sr, Th, Ti, V and Zn (mg/kg) contents determined in 3535 new topsoils as well as information on coordinates, geology, land cover. (BUCCIANTI et al., 2015).

AHRENS, L. ,1954a. Geochim. Cosm. Acta 6, 121–131.
AHRENS, L., 1954b. Geochim. Cosm. Acta 5, 49–73.
AITCHISON, J., et al., 2003. Math Geol 35(6), 667–680.
BUCCIANTI et al., 2015. Jour. Geoch. Explor., 159, 302-316.
CLARKE, F., 1889. Phil. Society of Washington Bull. 11, 131–142.
EGOZCUE, J.J. et al., 2003. Math Geol 35(3), 279–300.
MATEU-FIGUERAS, G. et al, (2005), Stoch. Environ. Res. Risk Ass. 19(3), 205–214.