Statistical analysis of geochemical compositions: Problems, perspectives and solutions

CrossMark

This special issue is devoted to multivariate statistical analysis of geochemical compositions. Compositional data refer to data where the relevant information is contained in the ratios between the values of the variables. In his seminal book, "The Statistical Analysis of Compositional Data", Aitchison (1986) introduced the log-ratio approach to treat data in percentages, proportions, concentrations, etc., a field that has expanded with both theoretical and practical developments, a trend that is only recently being followed by a significant adoption in the field of Geochemistry.

This special issue focuses on reliable applications of compositional multivariate statistical methods, based on solid theoretical grounds. According to the announcement, the papers should suitcase the advantages of log-ratio approaches with respect to other approaches, discuss practical problems and solutions related to those methods, or introduce innovative methodologies tailored to geochemical data. The intended application fields include, but were not restricted to, Chemometrics, Regional Geochemistry and Environmental Sciences.

This issue includes nine articles which we have organized alphabetically according to the first author's name.

The first paper of Blake et al. is devoted to an investigation of the source aquifers of the thermal groundwaters in a region of Ireland. The authors make use of different multivariate statistical methods, by incorporating the compositional nature of the data. This helped to identify at least two different aquifer types, one deep and one shallow.

Buccianti and Zuo focus on exploring the relationship among some concepts, often considered to be unrelated, such as weathering reactions, compositional data and fractals by means of distribution analysis. The isometric logratio coordinates, constructed using the sequential binary partitioning, are identified to maintain, as far as possible, the similarity with a correspondent weathering reaction affecting the Arno river catchment (Tuscany, central Italy) as described by the Law of Mass Action.

The paper by Ellefsen and Smith proposes a specific clustering procedure to partition the field samples successively into two clusters. In this way, a hierarchy of clusters is manually created, and the different levels of the hierarchy show processes that occur at different scales. For clustering, a Bayesian finite mixture model with two probability distributions is used for isometric logratio coordinates.

Fačevicová et al. present a statistical characterization of the Devonian-Carboniferous boundary, comparing results using a logratio perspective with those obtained from a classical point of view, without any prior transformation. The paper suggests that a combination of both approaches brings more insights than the use of any of the two methods separately.

The paper by Flood et al. is devoted to modeling grain-size data from a set of Holocene sedimentary cores from a region in India. In this context, it is shown how calibration models are used by considering the compositional nature of the data.

Jackson et al. present a case study relating chronic kidney disease prevalence with the regional geochemistry of Northern Ireland, by using a combination of logratio regression methods. The strength of the paper lies on the combination of data-driven and knowledge-driven methods to highlight the relationship between the disease prevalence and the relation between As, Pb, Cd, Ca and most specially Fe.

Mueller and Grunsky present a multivariate spatial analysis of a regionalized composition, with the goal to discriminate between lithological units. Their method combines linear discriminant analysis and minimum-maximum autocorrelation factors, a technique generalizing the widely known principal component analysis in order to account for spatial dependence between nearby samples.

Tolosana-Delgado and McKinley apply the logratio methodology for exploratory analysis, based on PCA and its graphical representation as a biplot, of a complex geological data set, the Tellus geochemical soil survey of Northern Ireland (NI). Results show that a log-ratio PCA (robust or classical) of all available elements is the most powerful exploratory setting among possible preprocessing alternatives (dealing with the original or standardized data).

Finally, in the paper by Xiong and Zuo, two models for mapping felsic intrusions are compared. The authors consider a factor ratio model based on the chemical (compositional) data characteristics, and a hybrid model combining PCA and local singularity analysis.

Peter Filzmoser*
Vienna University of Technology, Institute of Statistics and Mathematical Methods in Economics, Wiedner Hauptstr. 8-10, 1040 Vienna, Austria

Karel Hron
Faculty of Science, Palacký University, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 12, 771 46 Olomouc, Czechia
E-mail address: hronk@seznam.cz.

Raimon Tolosana-Delgado
Helmholtz-Zentrum Dresden-Rossendorf Chemnitzer, Str. 40, 09599
Freiberg, Germany
E-mail address: r.tolosana@hzdr.de.

[*] Corresponding author.
E-mail address: P.Filzmoser@tuwien.ac.at (P. Filzmoser).