Geophysical Research Abstracts
Vol. 16, EGU2014-4195, 2014
EGU General Assembly 2014

Statistical aspects when analyzing geochemical compositions

Peter Filzmoser (1), Clemens Reimann (2), and Manfred Birke (3)

(1) Department of Statistics and Probability Theory, Vienna University of Technology, Vienna, Austria (P.Filzmoser@tuwien.ac.at), (2) Geological Survey of Norway (NGU), Trondheim, Norway (Clemens.Reimann@ngu.no), (3) Federal Institute for Geosciences and Natural Resources (BGR), Hannover, Germany (Manfred.Birke@bgr.de)

From a statistical point of view, the GEMAS (Geochemical Mapping of Agricultural and grazing land Soil) project deals with data mainly consisting of concentrations of chemical elements in the soil. It is logical that an increase of one element concentration in the sample will have an effect on the other concentrations, namely this usually causes a decrease, in particular if the increasing element is very dominant in the sample. This phenomenon can be observed even if the total sum of all concentrations is less than 1 (if the concentrations are expressed in proportions). In statistical terms, such data are called compositional data. In an appropriate approach, the dependency of the element concentrations among each other can be investigated by analyzing ratios of the different element concentrations, rather than directly the raw data.

The above problem can also be viewed from a geometrical perspective. The raw concentration data are not represented in the usual Euclidean geometry, but in the so-called Aitchison geometry on the simplex. The statistical methods, however, that are usually employed for analyzing geochemical data are only designed for the Euclidean geometry.

Analyzing ratios, or working in the appropriate geometry, comes down to first transform the compositional data. Nowadays, the ilr (isometric logratio) transformation is frequently used for this purpose. It is preferred because of its advantageous mathematical properties, but the results usually need to be back-transformed for an interpretability. Note that a log-transformation - which is often used in geochemistry - does neither transform to the Euclidean space, nor it allows to focus on ratios.

Using the GEMAS data set, we will outline how the ilr transformation can successfully be used for data analysis, and we will compare the results with those derived from a simple log-transformation. It turns out that the concept of analyzing ratios has strong implications. This already concerns univariate mapping of geochemical data, where the “relative” view may provide quite a different insight than the “absolute” view. Also multivariate statistical analyses may lead to different conclusions when ilr-transformations rather than log-transformations are employed.

The GEMAS atlas is an example where the ilr approach has been used consistently.