Compositional data-driven alternatives to single geochemical component maps

K. G. VAN DEN BOOGAART[1,8*], J. MCKINLEY[2], P. DE CARITAT[3], P. FILZMOSER[4], E. GRUNSKY[5], K. HRON[6], C. REIMANN[7], R. TOLOSANA DELGADO[1]

[1]Helmholtz Institute Freiberg for Resource Technoloy - Helmholtz Center Dresden Rossendorf, Germany
boogaart@hzdr.de
[2]School of Geography, Archaeology and Palaeoecology - Queens Univ. Belfast, UK
[3]Geoscience Australia, Australia
[4]Dept. of Statistics and Mathematical Methods in Economics - Vienna University of Technology, Austria
[5]Geological Survey of Canada - Natural Resources Canada, Canada
[6]Dept. of Mathematical Analysis and Applications of Mathematics - Palacky University Olomouc, Czech Republic
[7]Geological Survey of Norway, Norway
[8]Institute for Stochastics - TU Bergakademie Freiberg, Germany
[*] presenting author

Conventional practice in Regional Geochemistry includes as a final step of any geochemical campaign the generation of a series of maps, to show the spatial distribution of each of the components considered. Such maps, though necessary, do not comply with the compositional, relative nature of the data, which unfortunately make any conclusion based on them sensitive to spurious correlation problems. This is one of the reasons why these maps are never interpreted isolated. This contribution aims at gathering a series of statistical methods to produce individual maps of multiplicative combinations of components (logcontrasts), much in the flavor of equilibrium constants, which are designed on purpose to capture certain aspects of the data.

We distinguish between supervised and unsupervised methods, where the first require an external, non-compositional variable (besides the compositional geochemical information) available in an analogous training set. This external variable can be a quantity (soil density, collocated magnetics, collocated ratio of Th/U spectral gamma counts, proportion of clay particle fraction, etc) or a category (rock type, land use type, etc). In the supervised methods, a regression-like model between the external variable and the geochemical composition is derived in the training set, and then this model is mapped on the whole region. This case is illustrated with the Tellus dataset, covering Northern Ireland at a density of 1 soil sample per 2 square km, where we map the presence of blanket peat and the underlying geology.

The unsupervised methods considered include principal components and principal balances (Pawlowsky-Glahn et al., CoDaWork2013), i.e. logcontrasts of the data that are devised to capture very large variability or else be quasi-constant. Using the Tellus dataset again, it is found that geological features are highlighted by the quasi-constant ratios Hf/Nb and their ratio against SiO2; Rb/K2O and Zr/Na2O and the balance between these two groups of two variables; the balance of Al2O3 and TiO2 vs. MgO; or the balance of Cr, Ni and Co vs. V and Fe2O3. The largest variability appears to be related to the presence/absence of peat.