



Correlation between variables in compositional data

Peter Filzmoser (1), Petra Kynčlova (2), Karel Hron (3), and Clemens Reimann (4)

(1) Vienna University of Technology, Dept. of Statistics and Probability Theory, Vienna, Austria (P.Filzmoser@tuwien.ac.at), (2) UNIDO, Vienna, Austria, (3) Palacky University, Olomouc, Czech Republic, (4) Geological Survey of Norway (NGU), Trondheim, Norway

In compositional data analysis, the interest is in analyzing relative information rather than the measured concentration values directly. This can be done by considering the log-ratios between all pairs of variables, and by constructing an orthonormal basis describing this information. In the context of correlation analysis, Kynčlová and others (2017) introduced the concept of symmetric balances, which are specific coordinates that capture all relative information of two compositional parts of interest to the remaining parts. The resulting balances are orthonormal, and thus they can be used for analyses relying on the Euclidean geometry, starting from graphical inspection to correlation analysis, testing for uncorrelatedness, and robust estimation of association.

Based on real data, we will demonstrate that already graphical inspection of the balances provides interesting insights into the data: trends, groups, and associations can be revealed, and it can be instructive to link such findings to absolute information. In the context of geochemistry it turns out that correlations based on symmetric balances serve as a sensitive tool to reveal underlying geochemical processes.

References

Kynčlová, P., Hron, K., Filzmoser, P. (2017). Correlation between compositional parts based on symmetric balances. *Mathematical Geosciences*. To appear.