robCompositions: An R-package for robust statistical analysis of compositional data

Matthias Templ[1,2,*]**, Karel Hron**[3]**, Peter Filzmoser**[1]

1. Department of Statistics and Probability Theory, Vienna University of Technology
2. Methods Unit, Statistics Austria
3. Palackỳ University Olomouc, Czech Republic
[*]Contact author: templ@tuwien.ac.at

Keywords: Multivariate Methods for Compositional Data, Robustness, R-package **robCompositions**

Compositional data are data that contain only relative information (see, e.g. Aitchison 1986)). Typical examples are data describing expenditures of persons on certain goods, or environmental data like the concentration of chemical elements in the soil. If all the compositional parts would be available, they would sum up to a total, like 100case of geochemical concentrations. Frequently, practical data sets include outliers, and thus a robust analysis is desirable. The R-package **robCompositions** (Templ et al., 2009) contains functions for robust statistical methods designed for compositional data, like principal component analysis (Filzmoser et al., 2009a) (including the robust compositional biplot), factor analysis (Filzmoser et al., 2009b), and discriminant analysis (Filzmoser et al., 2009c). Furthermore, methods to improve the quality of compositional data sets are implemented, like outlier detection (Filzmoser et al., 2008), and imputation of missing values (Hron et al, 2010). The latter one, based on a modified k-nearest neighbor algorithm and a model-based imputation, is also supported with measures of quality of imputation and diagnostic plots. The usage of the package will be illustrated on practical examples.

References

Aitchison, J. (1986). The Statistical Analysis of Compositional Data. *Chapman & Hall*, London.

K. Hron, M. Templ, P. Filzmoser (2010). Imputation of missing values for compositional data using classical and robust methods. *Computational Statistics and Data Analysis*, to appear.

P. Filzmoser and K. Hron (2008). Outlier detection for compositional data using robust methods. *Mathematical Geosciences*, 40(3), 233-248.

P. Filzmoser, K. Hron, and C. Reimann (2009a). Principal component analysis for compositional data with outliers. *Environmetrics*, 20, 621–632.

P. Filzmoser, K. Hron, C. Reimann, and R.G. Garrett (2009b). Robust factor analysis for compositional data. *Computers and Geosciences*, 35, 1854–1861.

P. Filzmoser, K. Hron, and M. Templ (2009c). Discriminant analysis for compositional data and robust parameter estimation. Research Report SM-2009-3, Department of Statistics and Probability Theory. Vienna University of Technology, 28 pages.

M. Templ, K. Horn, P. Filzmoser (2009). robCompositions: Robust Estimation for Compositional Data. Manual and Package, R package version 1.3.3.