Process identification and modelling for compositional data

J. Bacon-Shone[1] and E. Grunsky[2]
[1]Social Sciences Research Centre, University of Hong Kong; *johnbs@hku.hk*
[2] Dept. of Earth & Environmental Sciences, University of Waterloo

Abstract

This paper has a focus on process identification and modelling for compositional data given different possible processes:
1. static process (single distribution) (i.e. examples like those in (Aitchison 1982))
2. smooth process (e.g. time process showing smooth gradual change, e.g. (Thomas and Aitchison 2003))
3. disjoint process (A or B separately but not together) (i.e. where minerals do not occur together because the necessary conditions are so different)
4. stoichoiometric process (A replaces B in a fixed ratio) (examples in (Grunsky and Bacon-Shone 2011) showing linear dependence in simplex)
5. amalgamation/hierarchical process (examples in (Bacon-Shone and Grunsky 2015) that illustrate different behaviour within groups and across groups) (e.g. household expenditure, first choose whether to purchase alcohol, then choose what form of alcohol versus choose whether to purchase clothes and then what form of clothing)

The paper also discusses the effects of boundaries and detection - how they affect process identification and the connection to count data.

Other key issues include the consequences of these different process including zeroes and patterns of covariability in logratio and simplex space.

We explain how the different process lead to different patterns using graphical illustrations (e.g. biplot in simplex and normal space) and how to distinguish processes, especially if there are zeroes.

Lastly, we use an OSNACA database to illustrate processes 3 and 4 in a simple way.

References

Aitchison, J. (1982). "The statistical analysis of compositional data." Journal of the Royal Statistical Society. Series B (Methodological) **44**(2): 139-177.

Bacon-Shone, J. and E. Grunsky (2015). Compound compositional data processes. International Workshop on Compositional Data analysis, CoDaWork 2015.

Grunsky, E. and J. Bacon-Shone (2011). The stoichiometry of mineral compositions. CODAWORK 2011. Girona, Spain.

Thomas, C. and J. Aitchison (2003). "Exploration of geological variability and possible processes through the use of compositional data analysis: an example using scottish metamorphosed."