

Paper Number: 3074

Correlation analysis for compositional data using logratio coordinates



Hron, K.¹, Filzmoser, P.², Kynčlová, P.²

¹Department of Mathematical Analysis and Applications of Mathematics, Palacky University, 17. listopadu 12, 771 46 Olomouc, Czech Republic, email: hronk@seznam.cz

²Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Wiedner Hauptstrasse 8-10, A-1040 Vienna, Austria

Compositional data that commonly occur in geosciences in terms of proportions, percentages, or mg/kg, have specific properties that are not compatible with geometrical assumptions of most standard statistical methods [1,5]. Particularly, their constant sum representations (like the above listed cases) lead to a distortion of the correlation structure, commonly known as *spurious correlation* in literature. Since the seminal paper of Pearson [6] where the failure of the correlation coefficient in case of compositional data was pointed out, many attempts were made to define a reasonable measure of association between compositional parts. One possibility was proposed by Aitchison [1] with the elements of the variation matrix, i.e. to take the variance of the logratio of two compositional parts for this purpose. Accordingly, small variance indicates nearly a constant ratio between both parts and, consequently, their possible interchangeability. Although even some updates of this measure were presented to enhance its applicability [3], its interpretation in the sense of a correlation coefficient (particularly, the possibility of having either *positive* or *negative* relation between relative information conveyed by both parts) is not possible. Aim of this contribution is to present possible approaches to a correlation measure between compositional parts, based on the use of isometric logratio coordinates [2], concretely their special choice according to [4]. Consequently, new orthonormal coordinates are obtained, capturing relative information (in form of aggregated logratios) on both compositional parts of interest simultaneously, where the standard correlation analysis can be applied. Theoretical results will be supported by a case study from geochemistry.

References:

- [1] Aitchison J (1986) *The Statistical Analysis of Compositional Data*: Chapman and Hall
- [2] Egozcue JJ et al. (2003) *Mathematical Geology* 35(3): 279-300
- [3] Filzmoser P et al. (2010) *Science of the Total Environment* 408(19), 4230-4238
- [4] Fišerová E and Hron K (2011) *Mathematical Geosciences* 43(4): 455-468
- [5] Pawlowsky-Glahn V et al. (2015) *Modeling and Analysis of Compositional Data*: Wiley
- [6] Pearson (1897) *Proceedings of the Royal Society of London* LX, 489-502

