

Interactive comment on “Compositional balance should be considered in soil particle-size fractions mapping using hybrid interpolators” by Mo Zhang and Wenjiao Shi

Anonymous Referee #1

Received and published: 30 September 2020

This paper presents an investigation on the use of compositional balances for the analysis and spatial mapping of particle-size fractions. In particular, it aims to compare the use of ILR balances for the purpose of performing: (1) linear regression (named GLM, although it is just LM in the study), (2) regression kriging, (3) random forest prediction.

I believe that the general topic of the investigation could be interesting for the applied field of study. However, in my view, the study has severe limitations, and should not be considered further for publication in this journal. The main weak points of the study are summarized in the following points.

1. Although only partially explained, the methods being compared appears to be all

C1

applied separately to the three univariate ILR coordinates. This is not correct, as the ILR coordinates are typically correlated, so the analysis should also consider the cross-correlations among coordinates. All the presented results are thus suboptimal, and does not provide any effective guidance for other studies in the field. This also leads inconsistencies in the results (see my following point 2.).

2. The results of linear models and kriging, if correctly applied, should not depend on the ILR basis being chosen. This has been proved by previous studies on the use of linear models and kriging for compositional data (see, e.g., Pawlowsky-Glahn et al, (2015)). The authors cited in the text (Fiserova & Hron, 2011) indeed suggest to choose the ILR basis driven by interpretation purposes (which may be eased by a particular basis), but they do not refer to the influence of the basis on the results themselves, as these are independent on the basis being chosen if the method can be restated in term of a projection in the simplex (as LM and RK). As such, studying the effect of the choice of the ILR basis in these cases does not provide any meaningful information, beside the evidence that the methods discussed in the manuscript were not applied correctly (see point 1.).

3. Studying the bias of linear models and regression kriging is not meaningful, because both are unbiased methods. If bias is found, it derives from an incorrect definition of the notion of bias, which should be considered in the geometry of the simplex. Moreover, all the statistics and summaries should be considered in a multivariate setting, and the consideration of univariate components of psfs should be completely avoided (particularly if the aim is to approach them in a compositional setting). Overall, the paper does not discuss clearly the background on compositional data analysis, and the comments to analyses and results are often formally inappropriate, showing inconsistencies and general confusion on the concepts related with the theory of compositional data analysis.

4. The discussion is very confused, and the overall message strongly hindered by incorrect English wording.

C2

References: Fišerová E, Hron K (2011) On the interpretation of orthonormal coordinates for compositional data. *Mathematical Geosciences* 43(4):455–468
Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R (2015) *Modeling and analysis of compositional data*. John Wiley & Sons

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2020-384>, 2020.