

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 #2

Received and published: 20 October 2020

The presented manuscript proposes a comparison, in the context of the use of compositional data analysis (CoDA) techniques to perform digital soil mapping of particle-sized fractions, of different ILR transformation choices and different prediction algorithms.

The authors, after having provided a brief analysis of the current literature on the use of compositional data in geosciences, they perform three different ILR transformations of the data, and then proceed to assess and compare the prediction accuracy of several statistical learning methods, namely linear regression (glm with gaussian errors and identity link is classical least squares, gaussian regression), universal kriging and

C1

random forests, via the use of a real world dataset.

They then conclude by assessing what is the best algorithm in terms of prediction by inspecting several performance metrics.

While I do think that the general topic of investigation is of quite interest for an audience of geosciences practitioners, and so it is coherent with the aims of this Journal, I am quite concerned by the execution of the paper, and I think some very serious points need to be tackled before this paper is able to be considered suitable for publication:

1. The wording is very obscure at times, hindering the very comprehension of the matters at hand
2. Judging by how the performance metrics are chosen, the prediction problems solved by the authors are all scalar ones, and so the methods seem to have been applied separately to the different components. This is wrong, as it is fundamental in a compositional setting to inspect the cross-correlations between variables (and thus use multivariate prediction methods)
3. Given that linear methods (such as linear regression and regression kriging) are invariant to the choice of ILR basis, I am baffled by seeing results for these methods that are different across different ILR transformations.
4. The estimation of a bias metric via the use of RMSE on unbiased estimators (such as LM and RK) is simply incorrect.

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

C2