Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-584-RC2, 2019

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Systematic comparison of five machine-learning methods in classification and interpolation of soil particle size fractions using different transformed data" *by* Mo Zhang and Wenjiao Shi

Tomislav Hengl (Referee)

tom.hengl@envirometrix.net

Received and published: 23 April 2019

Paper by Zhang and Shi on comparing ML methods for predicting PSFs is novel and comprehensive. It is also to my knowledge first time somebody has compared using compositional transformations for soil texture mapping in such a detail (I remember following a short course by Matthias Templ which uses also geochemical data for interpolation purposes, but this paper extends to wider range of ML techniques). The authors used most of relevant packages (use of caret package is especially welcomed), I only regret that the data and the R code is not shared in a way that is more accessible

(github and zenodo.org). I have tried to download the data but I finished on a website for Chinese users only and which requires registration etc. This is my biggest criticism of this paper in fact.

I would recommend publication of this paper under condition that the following 3 things are clarified and improved:

1. Section 2.2 needs to be extended. How was the sampling plan generated? Did you sample per soil horizon or at fixed depths? Did you take bulk samples or are the samples close-to-point support? Which laboratory methods you used and what is the average measurement error per texture fraction? These issues need to be clarified.

2. Figures 2 (irregular scales of x-axis) and 8 (unreadable text and missing description of point colors) need to be improved following the comments I made in the text.

3. The authors need to make it clear in the abstract (and discussion) what is their final conclusion considering: is transforming fractions necessary or not? which transformation is 'the best' (mapping accuracy wise)? what would you recommend as the best strategy to map PSFs / texture classes? and what are the most important implications of this work?

An additional recommendation, which is optional, is not to use Lat and Lon as predictors in a ML framework (it leads to obvious artifacts). Instead I would advise the authors to combine geographical distances to points, which is explained in detail in https://peerj.com/articles/5518/ and which also helps solve problems of spatial autocorrelation etc. Again, this is optional recommendation and the paper would be equally useful without this adaptation. I also recommend looking at the work by Tolosana-Delgado et al. on interpolating compositional data (https://link.springer.com/article/10.1007%2Fs11004-018-9769-3), which is an important reference maybe missed by the authors.

Please also note the supplement to this comment:

https://www.hydrol-earth-syst-sci-discuss.net/hess-2018-584/hess-2018-584-RC2-supplement.pdf

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-584, 2019.