

## ***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***

**Anonymous Referee #2**

Received and published: 29 January 2020

I went through the manuscript submitted by the authors. Compared to previous versions of the manuscript appeared online, I appreciate the final discussion on the methods being used, particularly on the strengths and weaknesses of the compositional transformations as well as on the limitations of the not-transformed approach being used in the analysis.

Although the discussion is overall well-done, I believe that the manuscript would benefit from a clarification on the approach of analysis included earlier in the manuscript (with particular focus on the way methods are used). In Section 2.4.6 the authors should

C1

more clearly state that all the methods will be used separately on the components. Setting apart the issue on the appropriateness of the methods, I believe that a clear statement on the approach used and on the limitations of the presented study would provide a fair setting for the presentation of the interesting results presented by the authors, making the reader aware from the very beginning on the boundaries of usability of the presented analysis.

I thus suggest a revision that (1) clearly states that the methods will be applied independently on the components, before the presentation of the results (e.g., in the section on methods); this should be also better clarified in the supplement

(2) discuss from the very beginning that the range of applicability of the study is limited to this type of modeling (component-wise), whereas a joint modeling could lead to different results

(3) mention which of the methods being compared could be also used to build a joint regression model for the psf; this would better contextualize the modeling choice of the authors, as well as the limitations of the study.

On the latter regard, I believe that the following methods could be also easily applied to multivariate vectors: KNN, Multilayer perceptron neural network, Random Forest, Support vector machine.

Finally, the authors should mention how to deal with the possible presence of negative predictions when the analysis is carried out on not-transformed data, as this is not discussed in the paper.

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2019-648>, 2020.

C2