

## ***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***

**Mo Zhang and Wenjiao Shi**

shiwj@lreis.ac.cn

Received and published: 14 March 2019

Dear Yen-Sen Lu: Thank you for your short comment about the parameters used in the ML methods. The parameter table shows all parameters of machine-learning methods for soil particle size fractions interpolation and (direct) soil texture classification. In our manuscript, the range of most dependent variables were demonstrated, and this table can give more detailed information with respect to these parameters. For KNN, three parameters were tuned and their values were analogous; for MLP, the maximum of iteration was 1000, the learning rate of the resilient backpropagation algorithm (the

C1

learning function we chose, i.e. Rprop) is adaptive, therefore only the number of units in the hidden layers (i.e. size) were adjusted; for RF, we provided the number of trees to grow (ntree) and the number of variables randomly sampled as candidates at each split (mtry); for SVM, the radial basis function was selected as the kernel function, and cost and gamma were revealed in this table; for XGB, seven parameters were adjusted. All parameters we adjusted were included in this table. If there are other parameters, then their values are the default. Additionally, we noticed that when these parameters were fixed within certain ranges, the values of assessment indicators (e.g. MEs, RMSEs, etc.) remain stable and the same rule were obtained. Moreover, although log ratio approaches were widely used in spatial prediction of soil properties in soil science, there has been concern that the optimal estimate of log ratio transformed data does not deliver the optimal estimate of the compositions back-transformed to the real space. In the process of parameters optimization, the optimal parameters of different machine-learning methods using log ratio transformed data were obtained; however, it cannot guarantee that the values of assessment indicators were optimal using these parameters when independent data set were used to validate the models. Thus, the “optimal” does not seem very significant and this is the reason why we just gave the relative ranges of these adjusted parameters and omitted this parameter table in our manuscript. But there is no doubt that it is necessary to acquire these parameters in detail if you want to reproduce the work and generate the same results.

Yours sincerely, Wenjiao Shi

Please also note the supplement to this comment:

<https://www.hydrol-earth-syst-sci-discuss.net/hess-2018-584/hess-2018-584-AC2-supplement.pdf>

---

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

C2