

Interactive comment on “Machine learning and soil sciences: A review aided by machine learning tools” by José Padarian et al.

Wei Shangguan

shgwei@mail.sysu.edu.cn

Received and published: 25 October 2019

I am not an invited reviewer. But I think this review is very interesting and may have fundamental influence on the application of machine learning on soil science. Since it is an open discussion paper, I would like to offer my comments here. I hope this may help to improve this review a little. 1, The term “advanced ML” is mentioned several times. What do you mean by this? Any criteria? And please offer a list of the included ML algorithms by advanced ML. 2, it will be useful if the author can offer some details about the performance of ML and simpler approaches (at least the best and/or popular ones) with validation statistics such as R², RMSE and ME, which are reported by most studies. A boxplot may do the job. 3, Authors reviewed both the best performance (though no details offered) ML and the most used ML. Could you compare these two?

C1

There may be a gap between them and need attention for researchers to choose their ML wisely. As discussed by the author, performance and interpretability may both affect the choice of ML or other methods. Suggestion and insights may be offered by reviewing the most used, performance and interpretability (even not well defined, you may still classify them such as low, medium, high, potential low ...). 4, consider using the cited number of a paper in addition to the number of paper only. When defining the most used paper, you may use the cited number as a weight to each published paper. In this way, we may see a different pattern from Figure 7. 5, some short names have no full names, e.g. PLSR, PCR, kNN. And there are so many short names in the paper. Consider make a list of them as appendix. If it only appears one or two times, do not use a short name.

Interactive comment on SOIL Discuss., <https://doi.org/10.5194/soil-2019-57>, 2019.

C2