

SOIL Discuss., referee comment RC2
<https://doi.org/10.5194/soil-2020-102-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on soil-2020-102

Anonymous Referee #2

Referee comment on "On the benefits of clustering approaches in digital soil mapping: an application example concerning soil texture regionalization" by István Dunkl and Mareike Ließ, SOIL Discuss., <https://doi.org/10.5194/soil-2020-102-RC2>, 2021

Abstract: please add the brief obtained results, e.g., accuracy

Line 16: add a relevant cit.

Line 96: The soil texture of the soil legacy dataset used as model input or model output?

Line 107: please add a relevant cit. and software

Line 156: did you assign the number of clusters into 3?

Lines 160 to 184: are not clear. Adding a flowchart might enhance the clarification

Feature selection is also unclear. Did you specify the significant relationships between soil fractions and covariates? Please justify the paragraph

Did you transform soil fractions? If you use the raw data, the predicted maps do not guarantee the sum of 100%.

Random oversampling is just "copy and paste" of the original data. Do not think if the approach resulted in overfitting?

Line 245: please add a relevant explanation about the filter method in feature selection in the method section of the paper

Please calculate line concordance instead of R squared.

I also recommend reducing the size of the paper. There are many aspects of modeling are included in the current version. One possibility is to move them to the appendix.