

Interactive comment on “Mapping of soil properties at high resolution in Switzerland using boosted geoaddivitive models” by Madlene Nussbaum et al.

Anonymous Referee #2

Received and published: 8 August 2017

The manuscript is well written and deals with an important issue in digital soil mapping : How to take advantage of the multitude of co-variates and avoid problems of auto-correlation and model over fitting. The authors developed an automated routine using GAM models. They illustrate this routine with examples from different soil properties on a continuous scale (ECEC), binary scale (water logging) and ordinal scale (drainage class). I appreciate the complexity of the data and the statistics, but I would still insist on some quantitative measures to demonstrate the advantage of their approach: i) the selected variables in Table 2 should be more precise and the reader should be able to judge the relative importance and eventual interaction terms, ii) ii) Page 222 lines 4-5 How can the reader evaluate the benefits of the current approach in terms of reducing

C1

the risk of overfitting? iii) The case of the APEX data illustrates that some co-variates are not clearly described and one could even discuss their use in the model. Page 10 line 16: It is well-known that spectral information depends on the development stage of the vegetation for crops and that grasslands are less sensitive to development stage. When were the data acquired and what was the hypothesis on the inference on water logging/drainage class from these spectra? Minor remarks Page 2 line 28 ...boosted regression trees... Page 3 line 14 You refer to review paper of 2003 to discuss the recent trends in the application of GAM's. Please check for some more recent papers. I am sure that GAM's have been applied to predict soil properties. Table 2 What does 'UK-DMC' mean?

Interactive comment on SOIL Discuss., <https://doi.org/10.5194/soil-2017-13>, 2017.

C2