SOIL Discuss.,
https://doi.org/10.5194/soil-2017-13-AC2, 2017

SOIL Discussions Open Access EGU

Interactive comment on "Mapping of soil properties at high resolution in Switzerland using boosted geoadditive models" *by* Madlene Nussbaum et al.

Madlene Nussbaum et al.

madlene.nussbaum@env.ethz.ch

Received and published: 3 September 2017

Many thanks for your helpful feedback. We comment on your review in the subsequent text (P: page, L: line of the manuscript in two-column layout). Please further consider our suggestions for changes of the manuscript in the supplement to this document.

Details on covariate effects (Table 2)

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.

Complying with similar requests by referee 1, we added further details on covariates (radii of local windows used for computing terrain attributes, spatial resolution, etc.) to Table 2. Adding even more details — e. g. on non-linear effects of covariates — and keeping the table at the same time still well organized would be difficult. In the caption of Table 2 we refer to the partial residual plots in the Supplement of the manuscript where covariate effects are displayed. To keep the final models simple we did not include any interaction effects, which becomes apparent from the partial residual plots in Figures S1 and S4 to S7 in the Supplement.

Relative importance of covariates cannot be easily established for additive models (GAM) as opposed to tree-based methods. Estimated coefficients of one covariate depend on other covariates in the model. Moreover, because of collinearity, some covariates may be replaced by others without much loss of accuracy. To report covariate importance one could resort to a (disputed) partitioning of the coefficient of determination R^2 (Groemping 2006). We refrained from applying an ad-hoc method to evaluate covariate importance.

Evaluation of overfitting

Page 22 lines 4-5 (P15 L2 in two-column manuscript): How can the reader evaluate the benefits of the current approach in terms of reducing the risk of overfitting?

We used differences of statistics characterizing the accuracy of predictions for cross-

validation and independent validation as indicators for over-fitting. In view of a referee's comments on Nussbaum et al. (2017), we acknowledge that our use of the term "over-fitting" is not in accordance with its strict definition by Hastie et al. (2009). We therefore propose to clarify this at the begin of section 5.3 by replacing the first two sentences by the following text:

"For the final models, cross-validation statistics were similar to results obtained for the independent validation data. Through repeated cross-validation on the same subsets the cross-validation statistics are considered as conservative *goodness-of-fit* statistics (see our answer to a comment by referee 1). Hence, we conclude that the model did not over-fit the calibration data."

APEX spectral data as covariates

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 (P6 L96 in two-column manuscript): 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?

Firstly, we agree that the information on covariates is rather sparse in the manuscript. Nevertheless, we suggest not to expand the manuscript with details on APEX imagery. The manuscript focuses on the geoGAM framework, hence we tried to avoid adding too much information specific to the study regions. As you well noted APEX bands were not relevant covariates, therefore we only included citations, to the benefit of readers interested in details, but did not go into details. According to Schaepmann et al. (2015) the flight campaigns took place in autumn and spring (09/2013, 04/2014) and captured

the vegetation in different stages. Furthermore, there were bare soil areas and soil moisture likely differed on the two occasions. Table S2 in the Supplement of Nussbaum et al. (2017) — a follow-up study focusing on a comparison of DSM methods for the same study regions and in the meantime published as a citable discussion paper — explains that we used an indicator covariate to account for different sampling dates. However, we likely could not fully correct discrepancies between the two flight campaigns, and this may be one reason that the APEX covariates were not selected for the final models.

Secondly, we also agree that the usage of certain covariates can be questioned. We did not expect to find a relationship between APEX spectral bands and the drainage class or presence/absence of waterlogging. But we included the covariates for two reasons: (i) We presumed that a priori exclusion of covariates based on expert knowledge would in general hamper accuracy of predictions (see e.g. Brungard et al. 2015). (ii) An automatic model building procedure should be capable to exclude non-relevant covariates. Indeed, none of the APEX bands was included in the final geoGAM models for these responses.

Minor remarks

We added clarifications and references as proposed (see supplement).

References

Brungard, C. W., Boettinger, J. L., Duniway, M. C., Wills, S. A., and Edwards Jr., T. C.: Machine learning for predicting soil classes in three semi-arid landscapes, Geoderma, 239–240, 68–83, 10.1016/j.geoderma.2014.09.019, 2015.

Groemping, U.: Relative Importance for Linear Regression in R: The Package relaimpo, Journal of Statistical Software, 17, 1–27, 10.18637/jss.v017.i01, 2006.

Hastie, T., Tibshirani, R., and Friedman, J.: The Elements of Statistical Learning; Data Mining, Inference and Prediction, Springer, New York, 2 edn., 2009.

Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman, M., and Papritz: Evaluation of digital soil mapping approaches with large sets of environmental covariates, SOIL Discussions, 2017, 1–32, 10.5194/soil-2017-14, http://www.soil-discuss.net/soil-2017-14/, 2017.

Schaepman, M., Jehle, M., Hueni, A., D'Odorico, P., Damm, A., Weyermann, J., Schneider, F., Laurent, V., Popp, C., Seidel, F., Lenhard, K., Gege, P., Küchler, C., Brazile, J., Kohler, P., Vos, L., Meuleman, K., Meynart, R., Schläpfer, D., and Itten, K.: Advanced radiometry measurements and Earth science applications with the Airborne Prism Experiment (APEX), Remote Sensing of the Environment, 158, 207–219, 10.1016/j.rse.2014.11.014, 2015.

Please also note the supplement to this comment:
https://www.soil-discuss.net/soil-2017-13/soil-2017-13-AC2-supplement.pdf

---

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