SOIL Discuss.,
https://doi.org/10.5194/soil-2017-13-AC1, 2017

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 you 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.

Cross-validation vs. validation with independent data (comments on P4 L 80–90, P7–8, Section 3.3.3)

More precisely, it would be good to explain why both a K-fold cross-validation as well as a validation based on an independent set of data were needed.

P.4 L 80–90: It's unclear to me when K-fold cross-validation will be used and when an independent set of data will be considered for validation. Moreover, why a "10-fold" has been considered (and not something different than 10)?

The main goal for geoGAM modelling was computing predictions of soil properties. We therefore tuned the parameters of our model building procedure (e. g. number of boosting iterations, decisions on removal of covariates and merging of factor levels) based on cross-validation (cv) statistics evaluated with the calibration data set (and not based on "goodness-of-fit" measures computed with the same data). However, a model, built by repeatedly using cv, fits the calibration data "too closely", even if in a single application of cv, predictions and observations for a given subset are (seemingly) independent from each other. If we use cv repeatedly to tune parameters in a sequence of model building steps then predictions and observations of cv subset k in step i are no longer independent because the predictions depend on the observations of the kth subset through the parameters tuned before in steps 1 to $i - 1$. cv prediction errors in steps $i > 2$ are likely too small, and the respective cv statistics are too optimistic compared to statistics evaluated for independent data. We can therefore consider cv statistics as conservative goodness-of-fit measures. Hence, to get an honest picture of the accuracy of predictions for new data, one should therefore use an independent validation data set.

For cross-validation Hastie et al. (2009, p. 254) recommend to use $k = 10$ subsets. Our data was spatially clustered. We also chose $k = 10$ to keep the k calibration sets large enough to possibly still cover the full study area after the random splitting. With $k = 5$

some areas would most likely not any longer have been present in all k calibration sets.

P. 7–8 Section 3.3.3 It's fairly hard for me to understand when a certain statistical measure (to test the model predictive performance ect...) have been used on (i) the calibration data set or (ii) in the context of the K-fold cross-validation or (iii) considering the independent set of validation data-points. (See related general comment). Anyway, I guess it would be good to clarify this further in the MS and probably improve as well the structure of this section of the text in the light of this comment.

We did not use the criteria introduced in section 3.3.3 as "goodness-of-fit" measures (i) for the calibration data set, but we used them as cross-validation statistics (ii) for model building, and as statistics to evaluate final model performance with independent validation data, see previous comment. We agree that the current text might be confusing and proposed clarifications (see supplement).

Discussion of results (comments on P9 L31–37, P15, Section 5.3)

Furthermore, I believe that the results needs to be discussed much more in detail, by comparing this study's output with other models' performances when predicting + mapping the considered variables in the literature.

P. 9 L 31–37: It would be good to compare these results with results from other models as presented in the literature (See related general comment). I'm aware this need to be done in "Section 5.3" (see below / ultimate comment)

p. 15, Section 5.3. I believe that the discussion of the predictive performance of the fitted models can be worked out much more in detail by comparing this study's output with other models' performances when predicting + mapping the considered variables in the literature.

As suggested we added more references on digital soil mapping of ECEC and discuss now the accuracy of the predictions achieved in the various studies (see section 5.3 in supplement). Moreover, we mentioned the results 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.

We refrained from expanding the discussion on the accuracy of predictions for drainage class and presence of waterlogging. We cited a review that refers to 8 studies and 4 additional publications. Drainage classes were not often mapped, and their definition depends on local soil classification. Moreover, the validation statistics used in the various studies differ what makes comparison difficult. In the two last paragraphs of section 5.3 we nevertheless report for our study the statistics that were used by the other authors, although we think that these criteria are suboptimal.

Answer to specific comments:

References in introduction (P2 L20–30, L45–46, L50–52)

You refer quite a lot to McBratney et al. 2003's overview paper on DSM. Although, I'm convinced that this is a very important paper and you certainly need to mention it, I believe that it would be much better to integrate as well much more specific examples when refereeing to specific modelling techniques as well as more recent publications. / P. 2 L 45-46: Add in specific references? / P. 2 L50-52: You say "Lately" but subsequently refers to McBratney et al. 2003, which is a fairly old reference by now, so please, add more recent and specific references.

We agree on these comments and suggest to add more specific citations (see supplement).

Accuracy of coordinates (comment on P6 L12–13)

You state that the accuracy of the coordinates in about 25 m. Is this something you interpret that way (because records have been made in the field on topographical maps) or has this info been documented somewhere?

Unfortunately, we lack a thorough evaluation of the accuracy of geographical coordinates. During the surveys, site coordinates were recorded on topographical maps at a scale 1:25 000 without using positioning devices such as GPS or conventional land survey instruments. From the topographical map scale (the thickness of a pencil line (0.5–1 mm) corresponds to 12–25 m in the field) and by consulting soil surveyors we estimated possible deviations to about 25 m. By reporting this figure, we wanted to point out that the spatial resolution of some covariates was better than the accuracy of the spatial coordinates of sampled locations. We propose to change the text slightly to "we estimated accuracy of coordinates to about 25 m".

Pedotransferfunction for ECEC (comment on P6 L35–39)

Why did you consider this additional 5 % of data for which you needed a PTF to estimate ECEC? Furthermore, did you test the effect of having included this data on the overall outcome / model performance / uncertainty ect...?

We used these additional 73 sites (of which 13 were randomly assigned to the validation set) because soil data was rather scarce. Furthermore, the bulk 95 % of the data were not homogeneous either: Several institutions had gathered soil samples with varying support (soil pits, fixed-depth aliquots collected on 20 x 20 m^2 plots) over about 30 years and had sent the samples for analysis to different labs. We could only partly account for this variation (e. g. between-laboratory variation, Walthert et al. 2016) in

data harmonization. We did not evaluate how the use of data derived from PTF affected the uncertainty of the predictions. For such an analysis one would need to consider also other characteristics of legacy data (non-constant support, inter-laboratory variation, change of analytical methods, data coding schemes, . . .) that all contribute to uncertainty.

Amount of data used for independent validation (comments on P6 L40, L67)

P. 6 L 40 So that's 21.7 % of the data used for Validation. Why 21.7 %? P. 6 L 67 So that's 20.6 % of the data used for Validation. Why 20.6 %?

We aimed to use about 20 % of the sites for model validation. After the splitting the data into calibration and validation sites we had to remove some sites because values were missing for some covariates (e. g. small clouds in SPOT5 satellite image). Missing values were more frequent in the calibration sets, hence the share of validation sites was eventually slightly larger than 20 %.

Explained variance (P9 L50–55)

By saying that "the model explained about 40 % of the variance of the log-transformed and 37 % of the variance of the original data", I'm wondering if you did make this statement by comparing your RMSE-value with the STDEV value in the validation dataset? If yes, was it done in a (K-fold) cross-validation context and/or based on the independent validation dataset? Can you please clarify what you did to come up with this conclusion?

With "explained variance" we referred to the mean squared error skill score (Eq. 14 in the manuscript) which is the ratio of the mean squared error and the variance of the data. The reported “explained variances” all refer to the independent validation sets. We suggest to change the text to make this clear (see supplement).

Spatial standard deviation and raster resolution for topography (comments on P10, Table 2, P13 L 10–25, L 48–63)

You refer to “SD slope” and “SD elevation” ect... as being the standard deviation in local neighbourhood. Can you please clarify this a little bit more? Does I got it right that you did calculate the standard deviation based on an X.X raster window in order to obtain alternative measures for terrain/topographical complexity?

This is correct. We calculated standard deviations in a circular moving window to obtain terrain roughness measures. Different radii represent terrain changes at multiple scales. We expanded the explanation in the table caption (see supplement).

P. 10 Table 2: Can you specify the 'r' (raster-resolution, i.e. 2m or 25 m) value of the maps of the considered topographical variables?

We proposed to additionally add the radii of local neighbourhoods used to calculate the terrain attributes (see supplement).

P.13 L 10–25 + L 48–63. In the interpretation of the results as regards the influence of topography on the modelled soil characteristics it will be crucial to mention the raster-resolution, i.e. 2m or 25 m, because different levels of topographical detail will represents different process, i.e. small scales irregularity within a field (e.g. reflecting local surface irregularities related to agricultural practices) visible at 2 m resolution versus larger scale topographical general slope signature (e.g. reflecting variability induced by soil erosion processes) visible at 25m resolution.

We agree that not just algorithms and radii of moving neighbourhoods used to compute the various terrain attributes are relevant, but also the raster resolution of the original elevation data. We proposed to change the manuscript accordingly (see supplement).

References

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.

Walthert, L., Bridler, L., Keller, A., Lussi, M., and Grob, U.: Harmonisierung von Bodendaten im Projekt “Predictive mapping of soil properties for the evaluation of soil functions at regional scale (PMSoil)” des Nationalen Forschungsprogramms Boden (NFP 68), Bericht, Eidgenössische Forschungsanstalt WSL und Agroscope Reckenholz, Birmensdorf und Zürich, 10.3929/ethz-a-010801994, 2016.

Please also note the supplement to this comment:
https://www.soil-discuss.net/soil-2017-13/soil-2017-13-AC1-supplement.pdf

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