Geophysical Research Abstracts
Vol. 21, EGU2019-4981, 2019
EGU General Assembly 2019

Soil property maps for Swiss forests by machine learning based model averaging

Madlene Nussbaum (1), Andri Baltensweiler (2), and Lorenz Walthert (2)

(1) Berne University of Applied Sciences (BFH), School of Agricultural, Forest and Food Sciences (HAFL), Zollikofen, Switzerland, (2) Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland

Increasing weather extremes lead to challenges in forest management. Especially, the water storage capacity of soils plays a key role for tree species selection that can cope with reoccurring drought periods. Spatial assessments of such soil functions require maps of various basic soil properties down to deeper soil layers. In Switzerland, only a legacy soil map at scale of 1:200'000 exists which was created with focus on agricultural suitability and which contains limited information on forest soils.

During the last 30 years, the Swiss Federal Institute for Forest, Snow and Landscape Research (WSL) assembled a homogeneous nationwide forest soil dataset of 2'070 locations. From this database we derived six soil properties (clay, sand, gravel, bulk density, soil organic carbon, pH) in six depth intervals (0-5, 5-15, 15-30, 30-60, 60-100 and 100-200 cm) resulting in a total of 36 responses. In a previous study Random Forest (RF) was fitted to the same data, but the predictions were prone to artefacts showing unnatural patterns. To balance these artefacts we employed a model averaging (MA) approach combining seven automatically built models. Models were fitted with 175 environmental covariates by 1) model selection for linear models through grouped lasso, 2) robust external drift kriging (georob), 3) geoadditive models selecting penalized smoothing spline terms by componentwise gradient boosting (geoGAM), and three different tree-based methods 4) boosted regression trees (BRT), 5) RF, 6) rule-based linear regression (Cubist) and 7) support vector machines with non-linear basis functions.

Model validation was computed with 382 evenly spread soil profiles (~20 % of the total data set) not used for model calibration. Comparing the approaches RF had most often the highest R^2 (22 responses) although differences were small. Model averaging slightly outperformed the best of the seven single fits for 7 of the 36 responses. After visual inspection of the predicted maps we used MA to create the final data product to avoid prediction artefacts seen in RF. Model performance ranged from R^2 of 0.18 to 0.64 with a mean R^2 of 0.40 for the 36 responses. Performance mainly depended on the soil property. Bulk density worked best (mean R^2 0.60 over all soil depth intervals) while soil organic carbon was most difficult to predict (mean R^2 0.22). In contrast to other studies model performance did not decrease with soil depth. Only for the lowest interval of 100-200 cm the validation statistics dropped slightly.