Geophysical Research Abstracts
Vol. 17, EGU2015-2106, 2015
EGU General Assembly 2015

Predictive mapping of soil properties at high resolution by component wise gradient boosting

Madlene Nussbaum (1), Andreas Papritz (1), Marielle Fraefel (2), Andri Baltensweiler (2), and Armin Keller (3)

(1) Swiss Federal Institute of Technology ETH Zurich, Departement of Environmental System Science, Institute of Terrestrial Ecosystems, Zurich, Switzerland (madlene.nussbaum@env.ethz.ch), (2) Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland, (3) Agroscope Reckenholz, Zürich, Switzerland

Accurate spatial information on soils is crucial for sustainable usage of the resource soil. Spatial planning, agriculture, forestry or natural hazards management need high resolution maps of potentials of soils for particular functions (e. g. water storage, nutrient supply). Soil functions are derived from basic soil properties like soil organic carbon or soil texture. For many regions precise maps of basic soil properties are missing. Hence, as a prerequisite for digital soil function mapping, maps of soil properties must be created with the desired resolution.

A wide range of statistical approaches (linear and additive models, external drift kriging, Random Forest) were used for this in the past. When numerous environmental covariates (e. g. hyper-spectral remote sensing data) are available the selection of the model with best predictive power is challenging. Besides the issue of covariate selection, one should allow for non-linear effects of covariates on soil properties.

To handle these difficulties we used a gradient boosting approach that included besides categorical covariates linear and smooth non-linear terms of continuous covariates as base learners. Residual auto-correlation and non-stationary relationships were modeled by smooth spatial surfaces. Gradient boosting of this flavor selects relevant covariates in a slow learning procedure and inherently models non-linear dependencies on covariates during the fitting process. The restriction to linear and smoothing spline base learners retains the interpretability of the fitted predictive models. The number of boosting iterations is the main tuning parameter and was determined by tenfold cross validation.

To explore the feasibility of the gradient boosting approach we mapped pH of forest topsoils in Canton of Zurich, Switzerland, at high (50 m) spatial resolution. Legacy pH measurements were available from 1200 sites in the in the forests of Canton of Zurich. Gradient boosting selected a sparse model with 11 out of 270 possible covariates. External validation with data of 166 sites not used for calibration was found to be satisfactory with a RMSE of 0.96 and R^2 of 0.54. We computed also prediction intervals by a model based bootstrap approach and validated externally the coverage of these intervals.