



Improvement of soil moisture regionalization based on random forest regression by applying score criteria for mobile cosmic-ray neutron sensing data

Daniel Altdorff^{1,3}, Ségolène Dega², Hendrik Paasche², and Martin Schrön²

¹Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany (daniel.altdorff@uni-potsdam.de)

²Helmholtz Centre for Environmental Research GmbH - UFZ, Department for Monitoring and Exploration Technologies, Leipzig, Germany

³Helmholtz Centre for Environmental Research – UFZ, Department of Computational Hydrosystems, Leipzig, Germany

Upscaling of soil water content (SWC) information towards the large-scale (>10 km) is highly desired to address the increasing demand on SWC products at various sectors. Random forest (RF) regression has been suggested as suitable method to generate large SWC maps from a limited amount of observations. RF deals with multiple prediction variables (predictors) to derive the missing values of a desired variable (e.g. SWC) based on their internal relationship. Cosmic ray neutron sensing (CRNS) is an alternative method for passive SWC mapping and monitoring, either by stationary CRNS sensors or by mobile CRNS roving. CRNS has a certain advantage over most classical hydrogeophysical approaches because of its footprint at the hectares-scale and beyond, particularly true for roving data, which qualifies CRNS data as suitable input for RF regressions. However, commonly CRNS roving data contain a high amount of noise and outlier values, related to the statistical distribution of neutron counting, which hinders the signal interpretation and could lower the quality of the RF regression performance. There are so far two ways to overcome the noise problem and to achieve a higher data stability; i) increasing of the aggregation time, which decreases the signal uncertainty but also reduces the spatial resolution and ii) applying smoothing algorithms, e.g. interpolation or moving averages, which results in more stable values, but it does not solve the outlier problem.

We used SWC data from CRNS roving along the Selketal catchment at the Harz mountain, Germany, to test the performance of a score criteria for an adaptive removal of potential outliers. The score criteria are internal test parameters, providing an indication about the probability of values that might be an outlier or not. Therefore, each observation was subject to a group of queries, asking its conformity to the surrounding values by selected statistical parameters. Based on the total score of the queries, the potentially unreliable observations were removed using various thresholds and used as input for the RF regression. RF regression was performed using static (e.g. topographical indices, soil properties) and dynamic (precipitation) predictors generating SWC maps from an area of ~2700 km². SWC input data were split into training (~2/3) and validation sets (~1/3).

Preliminary results showed that the application of the score criteria resulted in more stable spatial pattern and improved the R^2 from 0.099 to 0.196, 0.266 and 0.308 for score 6, 4 and 3, respectively. Achieved root mean squared error also decreased with stronger filtering, ranging from 0.14 for the original datasets to 0.078 for score 3. However, by using the score 3 threshold, 22.4% of the data were omitted. Hence, an optimization between the amount of excluded data and the resulting improvement of prediction needs to be developed and tested. The implementation of the spatial relationship in-between the observations and a weighting of the score values according to their importance should further increase the performance. Due to its easy application and its adjustable criteria selection, the proposed filtering approach has the potential to become more popular in CRNS roving studies.