Root zone soil moisture estimation with Random Forest

Authors: Coleen Carranza[a], Corjan Nolet[b], Michiel Pezij[c,d], Martine van der Ploeg[e]

[a] Soil Physics and Land Management Group, Wageningen University, Wageningen, The Netherlands
[b] FutureWater, Wageningen, The Netherlands
[c] Department of Water Engineering and Management, University of Twente, Enschede, The Netherlands
[d] HKV, Lelystad, The Netherlands
[e] Hydrology and Quantitative Water Management Group, Wageningen University, Wageningen, The Netherlands

Abstract

Accurate estimates of root zone soil moisture (RZSM) at relevant spatio-temporal scales are essential for many agricultural and hydrological applications. Applications of machine learning (ML) techniques to estimate root zone soil moisture are limited compared to commonly used process-based models based on flow and transport equations in the vadose zone. However, data-driven ML techniques present unique opportunities to develop quantitative models without having assumptions on the processes operating within the system being investigated. In this study, the Random Forest (RF) ensemble learning algorithm, is tested to demonstrate the capabilities and advantages of ML for RZSM estimation. Interpolation and extrapolation of RZSM on a daily timescale was carried out using RF over a small agricultural catchment from 2016 to 2018 using in situ measurements. Results show that RF predictions have slightly higher accuracy for interpolation and similar accuracy for extrapolation in comparison with RZSM simulated from a process-based model combined with data assimilation. RF predictions for extreme wet and dry conditions were, however, less accurate. This was inferred to be due to infrequent sampling of such conditions that led to poor learning in the trained RF model and to incomplete representation of relevant subsurface processes at the study sites in the RF covariates. Since RF does not depend on parameters required to estimate subsurface water flow, it is more advantageous than a process-based model in data-poor regions where soil hydraulic parameters are incomplete or missing, especially when the primary goal is only the estimation of soil moisture states.

Keywords:

Root zone soil moisture, Random Forest, Prediction, Process-based modeling.