Predicting and forecasting root zone soil moisture with Random Forests

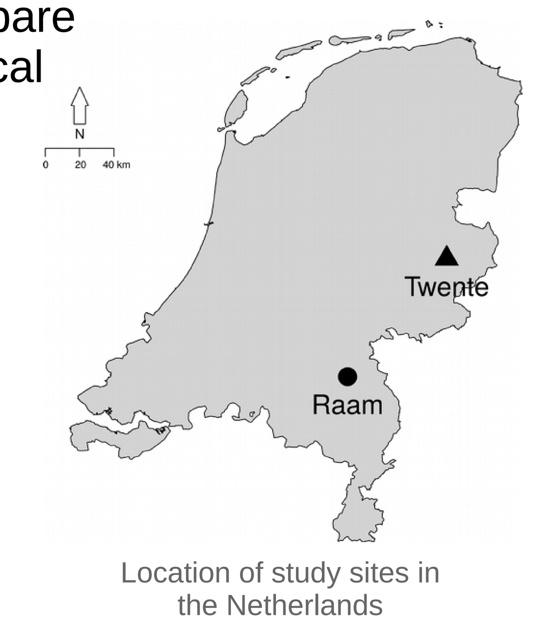
Coleen Carranza[1] and Martine van der Ploeg[2]

[1]Soil Physics and Land Management Group, Wageningen University, Netherlands
[2]Hydrology and Quantitative Water Management Group, Wageningen University, Netherlands

1. Introduction

- Root zone soil moisture (θ_{rz}) is important in the hydrologic cycle and has been used for drought monitoring, water storage estimation and carbon cycle monitoring;
- Using physical hydrological models are the most common approach for estimation of θ_{rz};
- Soil moisture datasets from different monitoring sites are continuously increasing and becoming more available (e.g. International Soil Moisture Network (ISMN));
- There is an opportunity to test data-driven methods such as Random Forests to estimate real-world θ_{rz} conditions which are currently far less common than physical models.

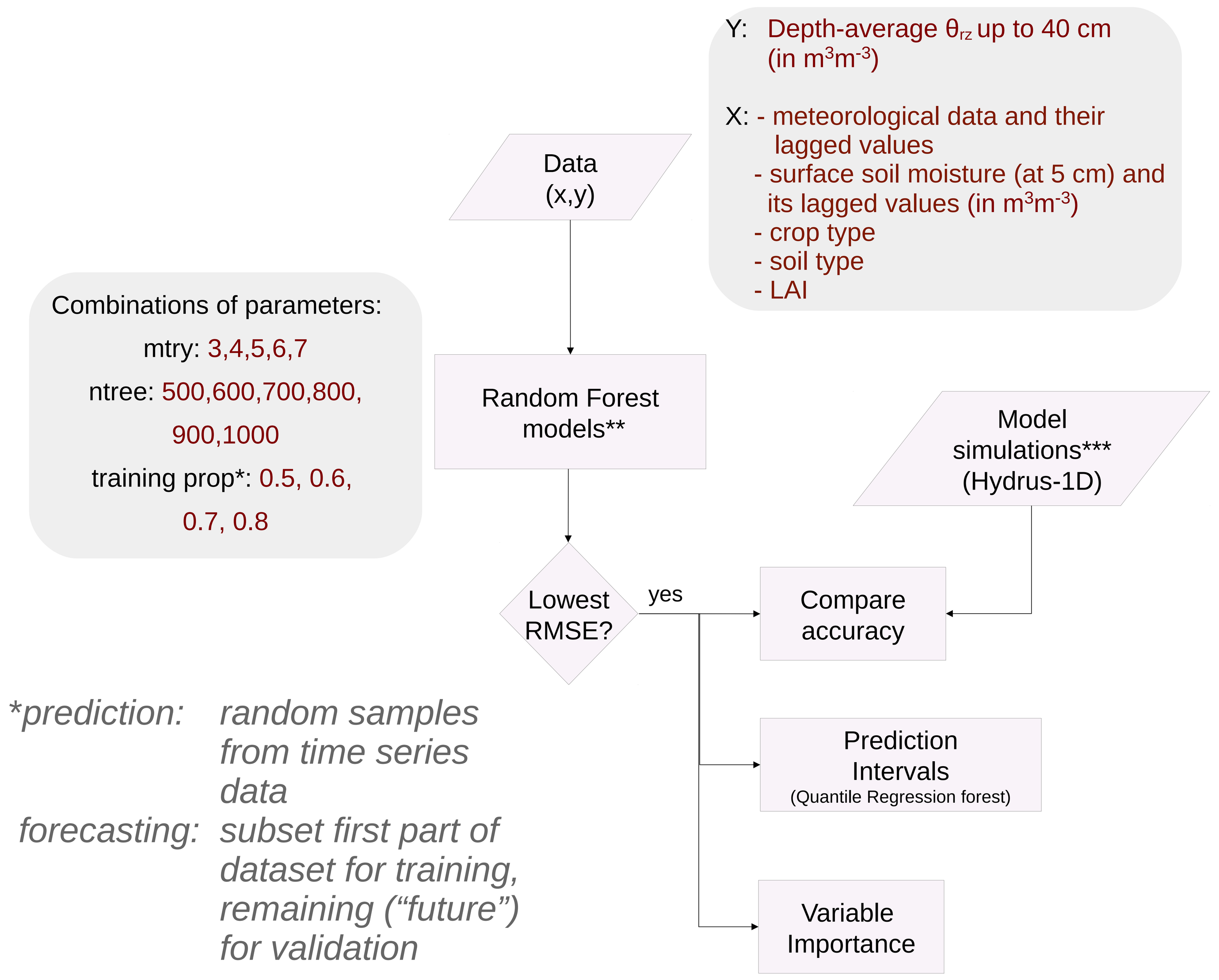
Objective: Perform Random Forest (RF) to predict and forecast RZSM and compare results with simulations from a physical hydrological model (Hydrus-1D).



Location of study sites in the Netherlands

2. Materials and Methods

Study area: 4 stations within existing soil moisture networks (1 in Raam and 3 in Twente) which are closest to meteorological stations of KNMI (Royal Dutch Meterological Institute).



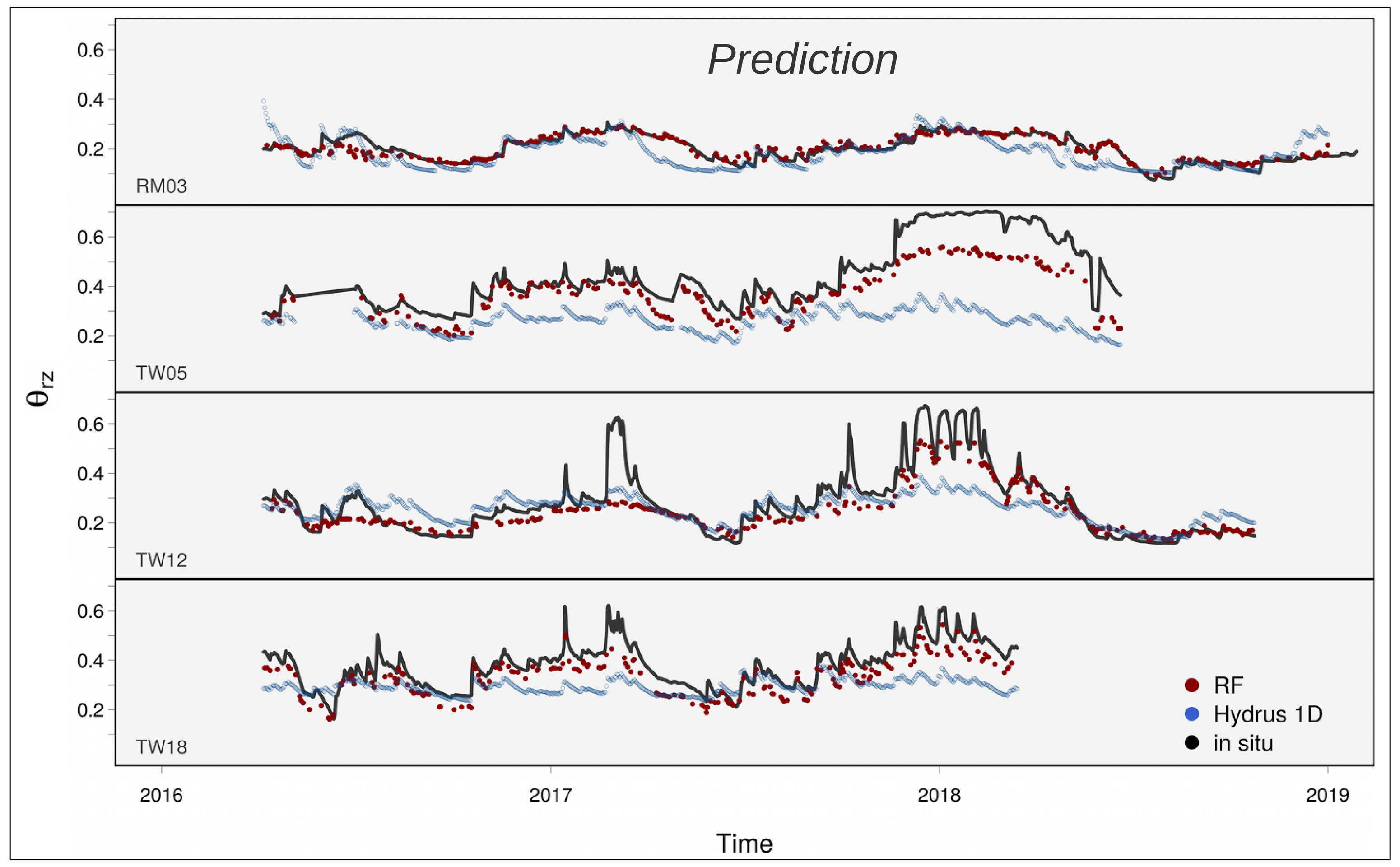
*prediction: *random samples from time series data*

forecasting: subset first part of dataset for training, remaining ("future") for validation

*** 120 sets of parameter combinations (based on list above) were used for training RF models. Samples from each site were combined train a single RF model*

**** pore flow simulations with Hydrus-1D utilized the same independent variables (X) as RF. Additional inputs to perform the simulations were: soil hydraulic parameters, root water uptake function, site characteristics (e.g. Lat, Long) of meteorological stations*

3. Results



A. Best RF parameter combination

mtry	5
ntree	900
training prop.	0.7

***Randomly selected RF training points not shown to visualize RF validation points (red) more clearly.

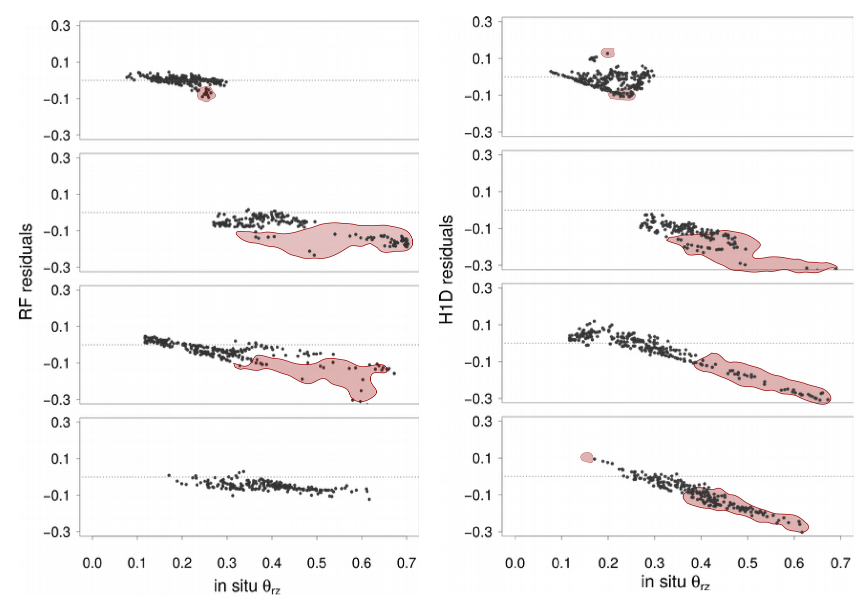
B. Accuracy Metrics: how does RF compare with H1D results?

	RMSE		Corr.		Bias		Unbiased RMSE	
	RF	H1D	**RF**	H1D	**RF**	H1D	**RF**	H1D
RM03	**0.021**	0.046	**0.912**	0.731	**-0.001**	-0.019	**0.021**	0.042
TW05	**0.093**	0.223	**0.94**	0.542	**-0.074**	-0.186	**0.056**	0.122
TW12	**0.069**	0.095	**0.919**	0.831	**-0.034**	-0.026	**0.06**	0.091
TW18	**0.055**	0.115	**0.971**	0.618	**-0.05**	-0.088	**0.023**	0.074

**RF - Random Forest H1D - Hydrus 1D*

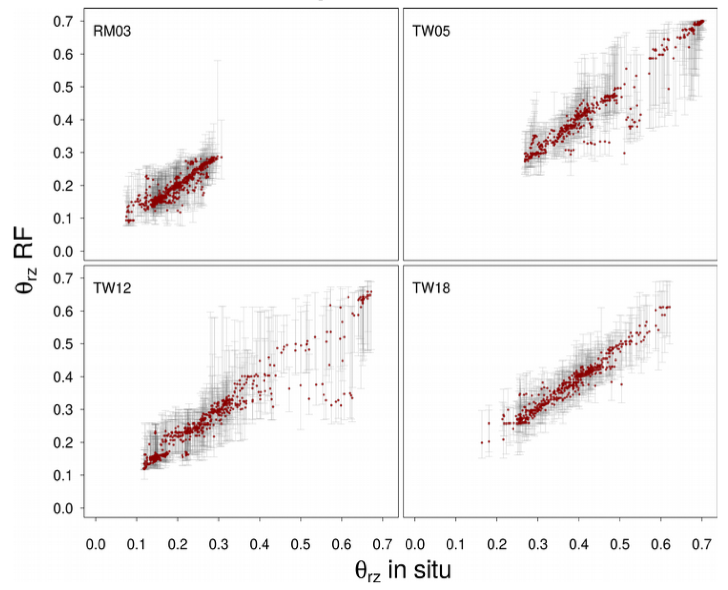
Values in green show better results.

C. Residuals: Which soil moisture values are poorly estimated?



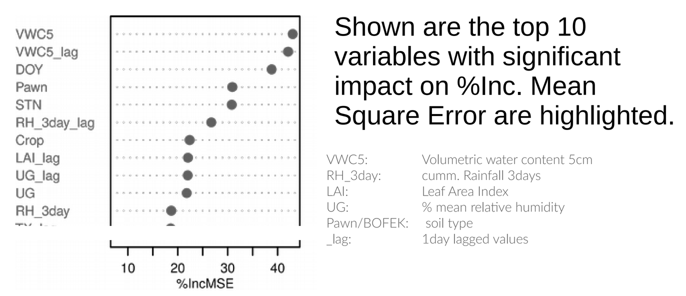
Points inside red regions have residuals >± 0.1 to show which soil moisture values are poorly predicted.

D. RF Prediction interval: How variable are each RF predictions?

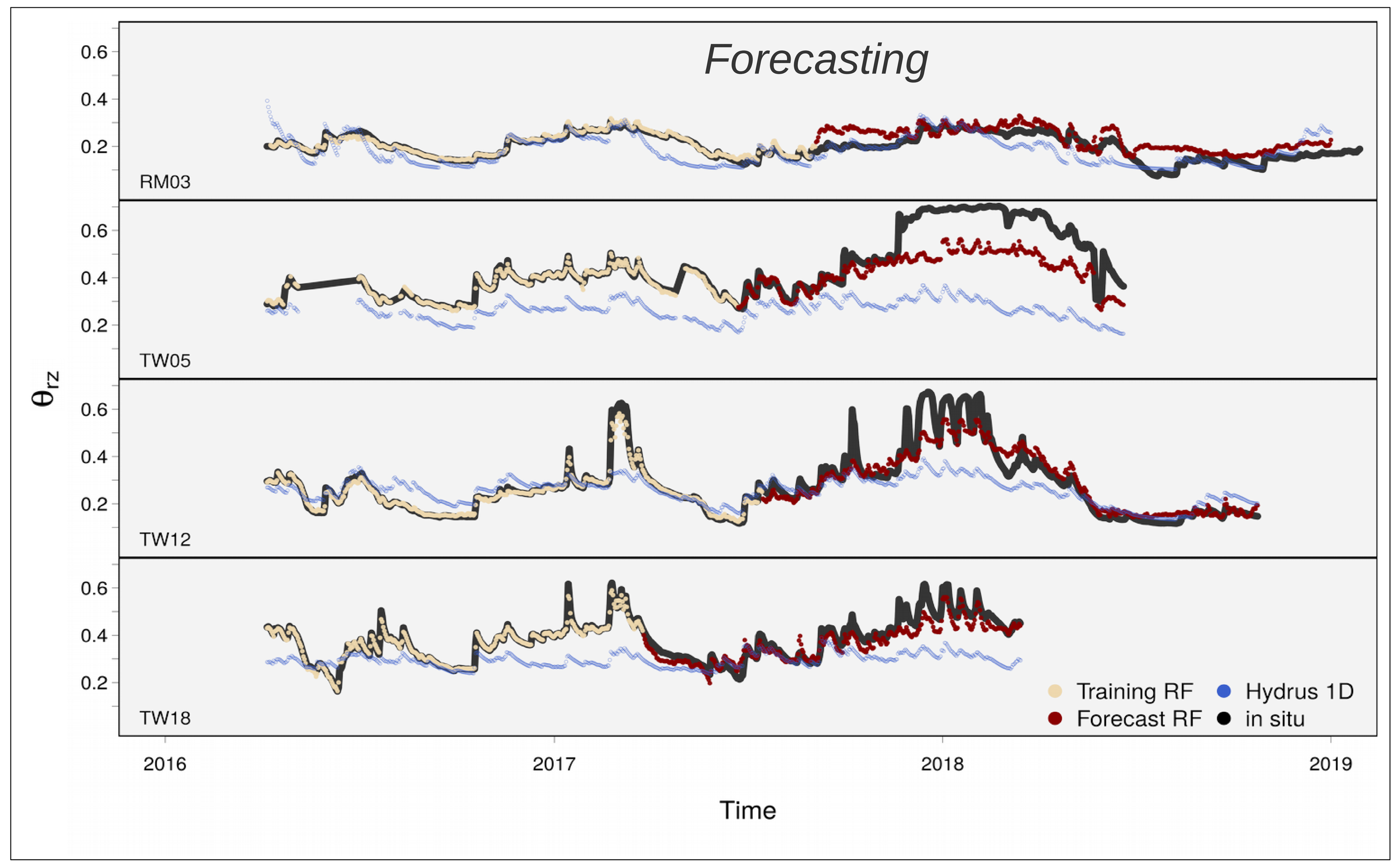


Red dots show the mean soil moisture prediction using RF. The gray bar per dot represents the prediction interval between 5% and 95% percentiles.

E. Variable Importance: Which variables significantly contributed to RF model?



Shown are the top 10 variables with significant impact on %Inc. Mean Square Error are highlighted.



A. Best RF parameter combination

mtry	7
ntree	1000
training prop.	0.5

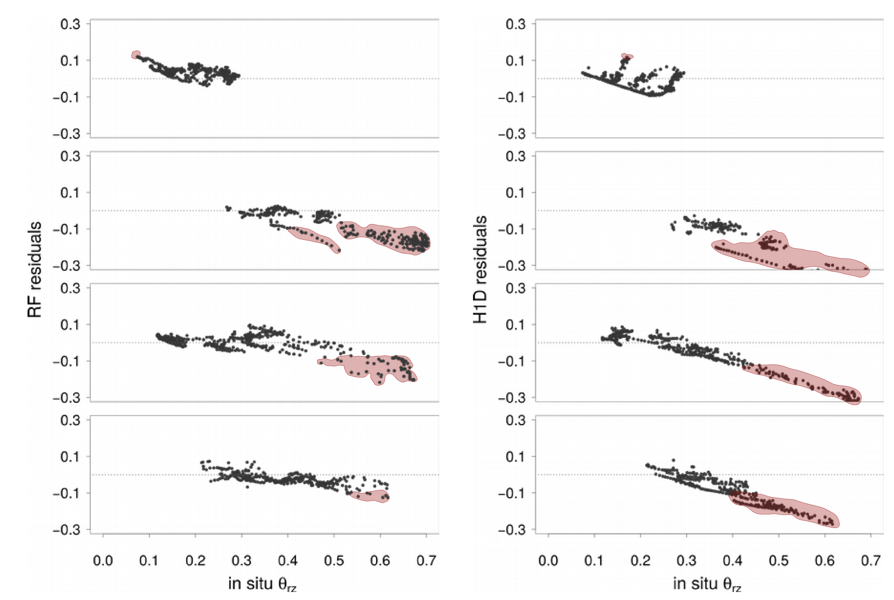
B. Accuracy Metrics: how does RF compare with H1D results?

	RMSE		Corr.		Bias		Unbiased RMSE	
	RF	H1D	**RF**	H1D	**RF**	H1D	**RF**	H1D
RM03	**0.05**	0.045	**0.833**	0.727	**0.039**	-0.009	**0.032**	0.044
TW05	**0.124**	0.285	**0.898**	0.386	**-0.096**	-0.254	**0.078**	0.13
TW12	**0.056**	0.115	**0.945**	0.894	**-0.006**	-0.051	**0.056**	0.103
TW18	**0.04**	0.118	**0.954**	0.601	**-0.024**	-0.088	**0.032**	0.078

**RF - Random Forest H1D - Hydrus 1D*

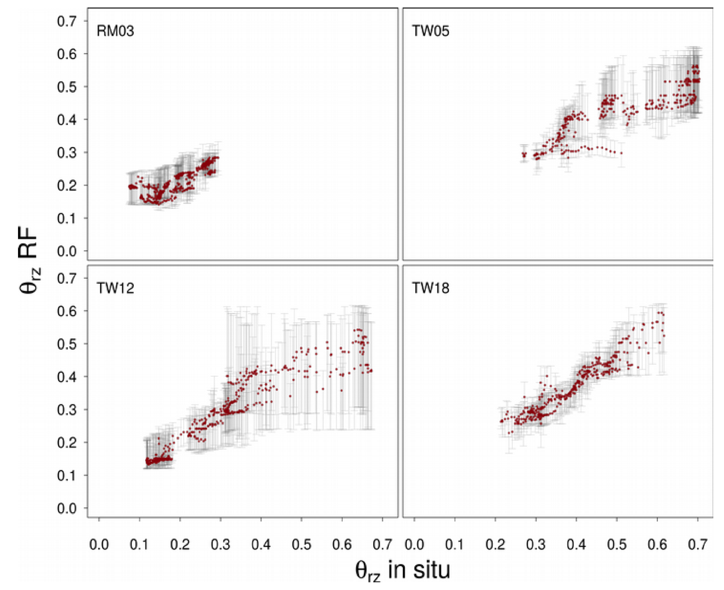
Values in green show better results.

C. Residuals: Which soil moisture values are poorly estimated?



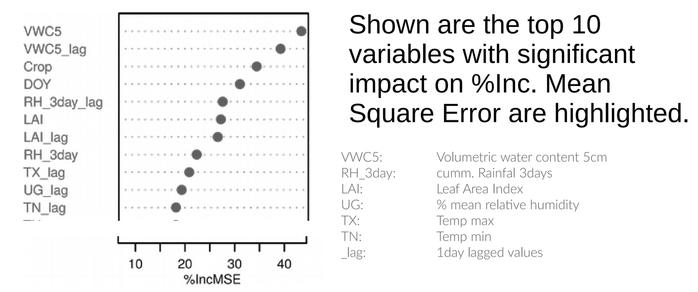
Points inside red regions have residuals >± 0.1 to show which soil moisture values are poorly predicted.

D. RF Prediction interval: How variable are each RF predictions?



Red dots show the mean soil moisture prediction using RF. The gray bar per dot represents the prediction interval between 5% and 95% percentiles.

E. Variable Importance: Which variables significantly contributed to RF model?



Shown are the top 10 variables with significant impact on %Inc. Mean Square Error are highlighted.

4. Discussion

- In general, RF predictions showed better results than H1D simulations (B), for both prediction and forecasting.
- Both H1D and RF underestimate very high soil moisture values (C). Overestimation of very low values is also observed, but is not as common
- For RF, variables which represent processes that promote occurrence of very high soil moisture should be included (e.g. likelihood of macropore occurrence, bulk density).
- Range of prediction intervals appear to be site-specific (D). However, for a couple of sites (TW05 and TW12),
- larger intervals are observed for high soil moisture content.
- Important variables (E) show combination of meteorological conditions, vegetation and, soil properties are necessary for accurate prediction. Lagged values appear to be as important as current values.

Why choose RF (or another data-driven method) over a physical model?

- Results from the study sites show capability of RF to accurately estimate θ_{rz}. It even surpassed the accuracy of Hydrus-1D estimates using a pore flow model.
- Soil hydraulic parameters not required to run RF compared to physical models, which means they can be especially applicable in areas when these are not available. Prediction method also can be used to fill data gaps in soil moisture time series
- When the main objective is to estimate soil moisture states, RF can do the task. However, if processes that control soil moisture states are also sought, physical models should be applied. Although for RF, a glimpse into these processes are given by the list of important variables.

Acknowledgements

This study is part of the project entitled Operational Water Management using Sentinel-1 Satellites (OWAS1S). The project is funded by Toegepaste and Technishe Wetenschappen (TTW) which is part of the Netherlands Organization for Scientific Research (NWO).