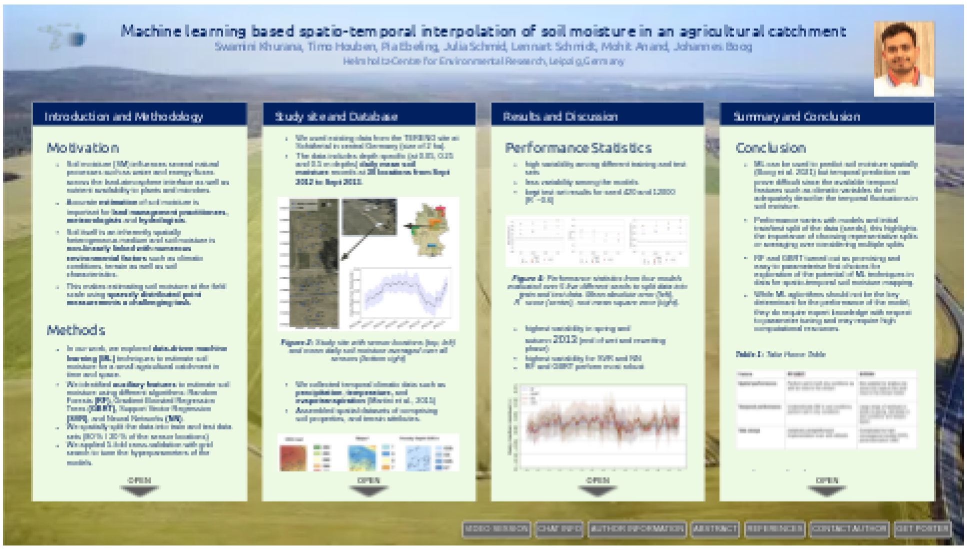
Machine learning based spatio-temporal interpolation of soil moisture in an agricultural catchment



Swamini Khurana, Timo Houben, Pia Ebeling, Julia Schmid, Lennart Schmidt, Mohit Anand, Johannes Boog

Helmholtz-Centre for Environmental Research, Leipzig, Germany

PRESENTED AT:

INTRODUCTION AND METHODOLOGY

Motivation

- Soil moisture (SM) influences several natural processes such as water and energy fluxes across the land-atmosphere interface as well as nutrient availability to plants and microbes.

- Accurate **estimation** of soil moisture is important for **land management practitioners, meteorologists** and **hydrologists**.

- Soil itself is an inherently spatially heterogeneous medium and soil moisture is **non-linearly linked with numerous environmental factors** such as climatic conditions, terrain as well as soil characteristics.

- This makes estimating soil moisture at the field scale using **sparsely distributed point measurements a challenging task.**

Methods

- In our work, we explored **data-driven machine learning (ML)** techniques to estimate soil moisture for a small agricultural catchment in time and space.
- We identified **auxiliary features** to estimate soil moisture using different algorithms: Random Forests **(RF),** Gradient Boosted Regression Trees **(GBRT),** Support Vector Regression **(SVR)**, and Neural Networks **(NN)**.
- We spatially split the data into train and test data sets (80 % / 20 % of the sensor locations)
- We applied 5-fold cross-validation with grid search to tune the hyperparameters of the models.

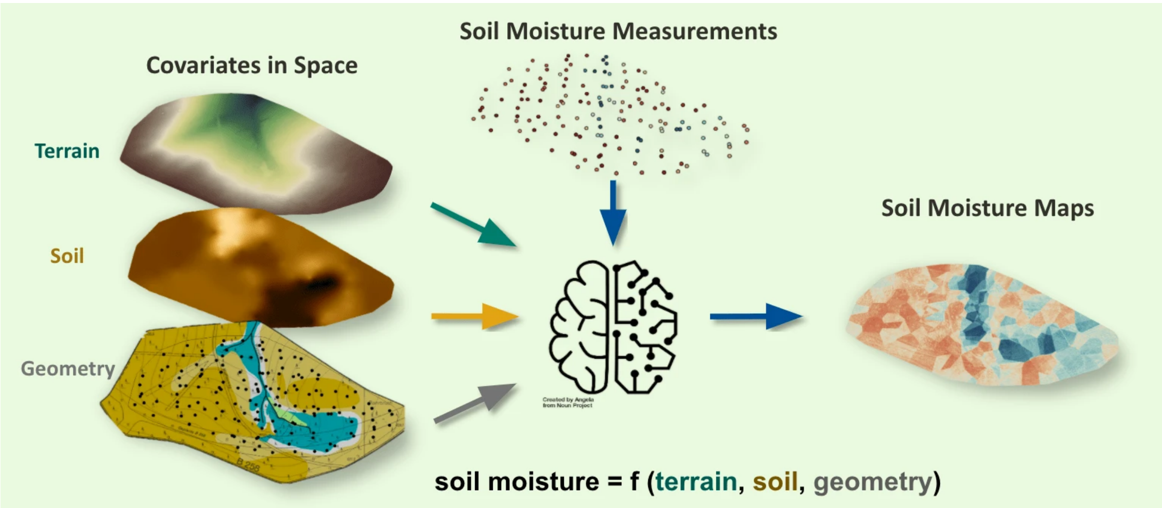


Figure 1: *Conceptual overview*

Research Questions

1. How accurately do machine-learning algorithms estimate soil moisture from auxiliary (environmental) data in time and space?

2. Do machine learning algorithms differ in their performance in estimating soil moisture in space? Do the patterns in prediction vary and, if so, where?

3. Which environmental variables are important to estimate soil moisture in space and time?

STUDY SITE AND DATABASE

- We used existing data from the TERENO site at Schäfertal in central Germany (size of 2 ha).
- The data includes depth specific (at 0.05, 0.25 and 0.5 m depths) **daily mean soil moisture** records at **30 locations from Sept 2012 to Sept 2013.**

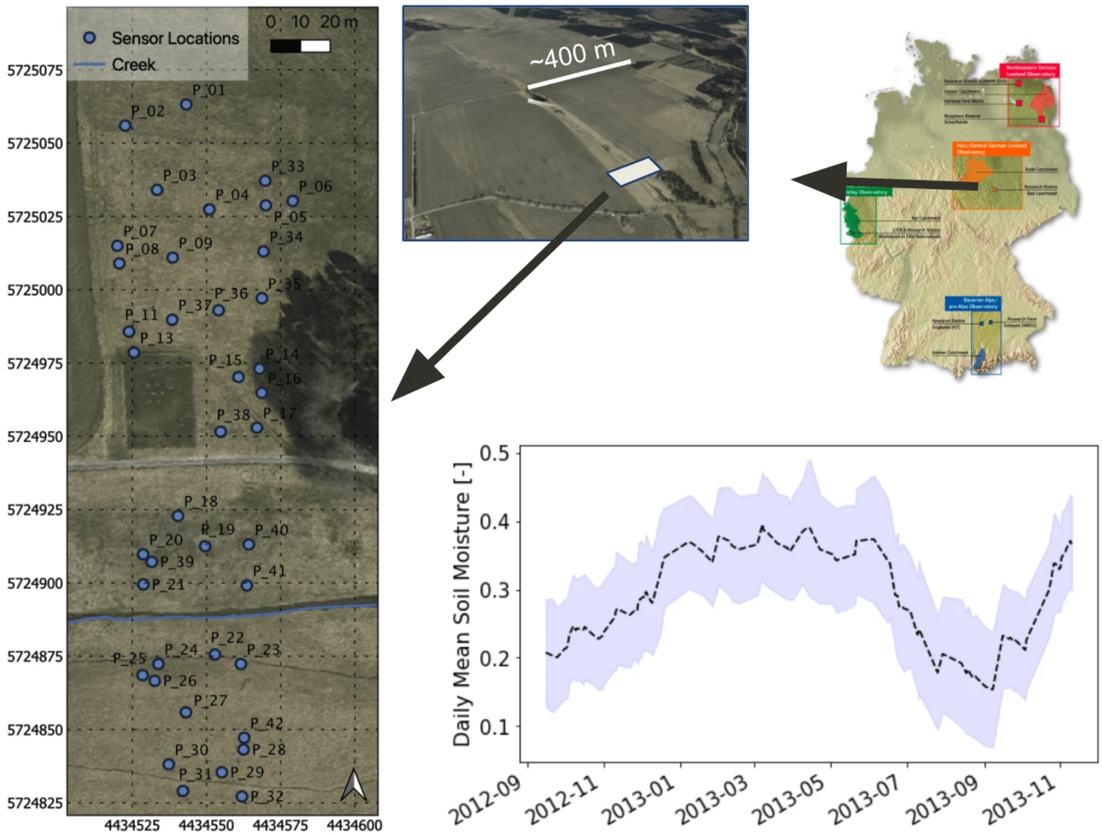


Figure 2: *Study site with sensor locations (top, left) and mean daily soil moisture averaged over all sensors (bottom right)*

- We collected temporal climatic data such as **precipitation**, **temperature,** and **evapotranspiration** (Martini et al., 2015)
- Assembled spatial datasets of comprising soil properties, and terrain attributes.

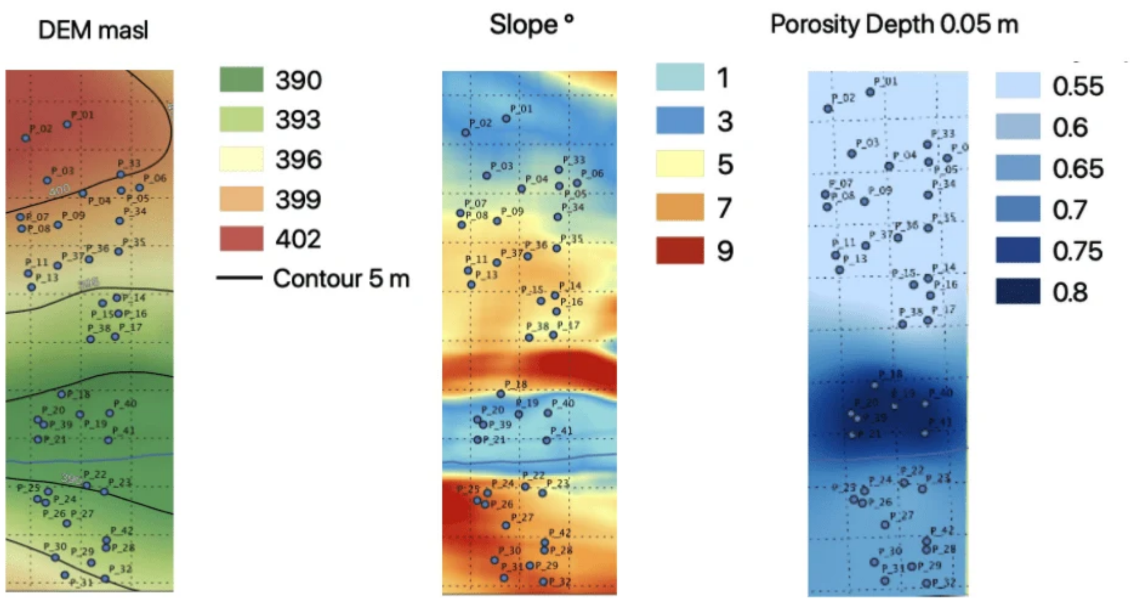


Figure 3: *The terrain covariates, digital elevation model (left) and hill slope (center), as well as one of the soil covariates, the interpolated porosity (right).*

RESULTS AND DISCUSSION

Performance Statistics

- high variability among different training and test sets
- less variability among the models
- best test set results for seed 420 and 12000 (R^2~0.6)

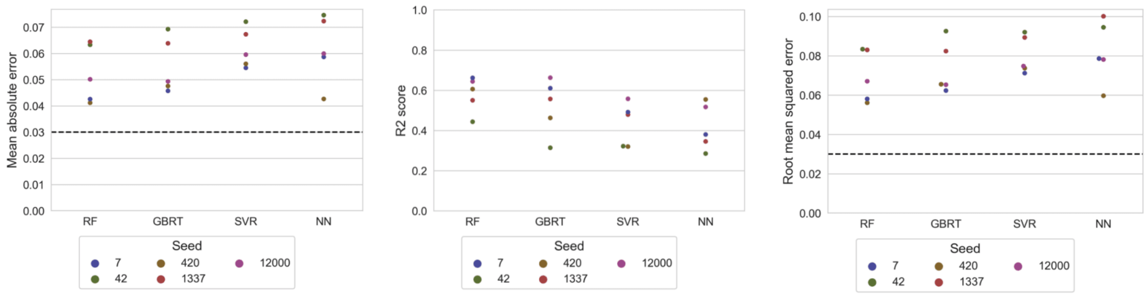


Figure 4: *Performance statistics from four models evaluated over 5 five different seeds to split data into train and test data. Mean absolute error (left), R^2 score (center), root mean square error (right).*

- highest variability in spring and autumn 2013 (end of wet and rewetting phase)
- highest variability for SVR and NN
- RF and GBRT perform most robust

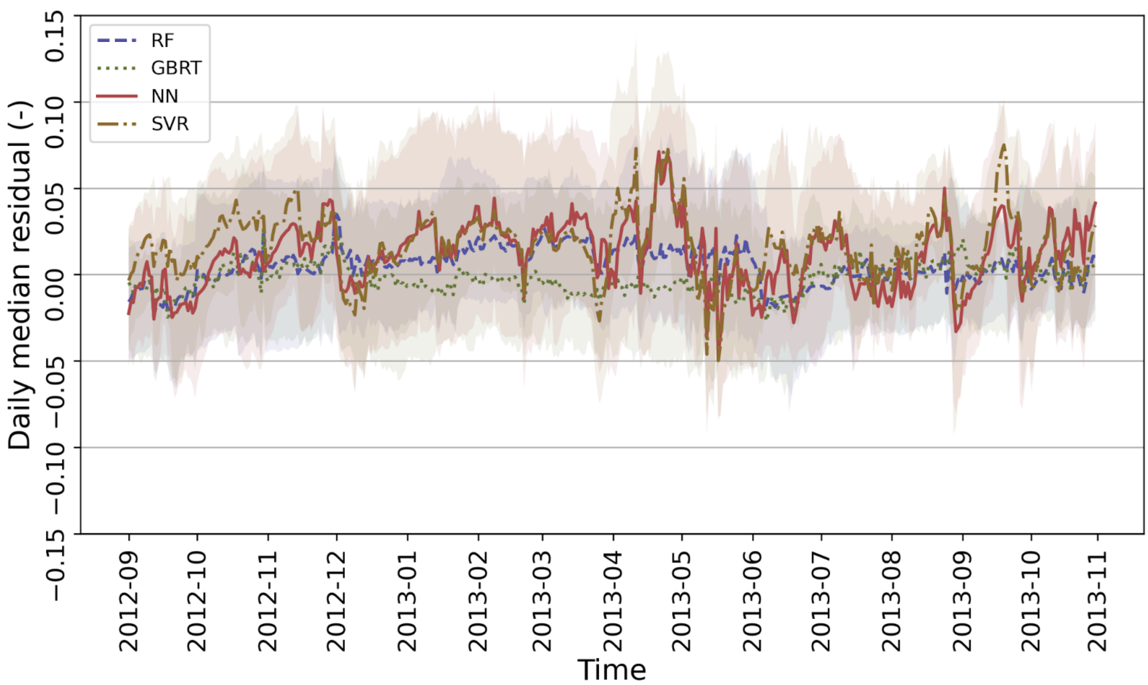


Figure 5: *Daily mean residuals and the interquartile range averaged over test data from all sensor locations, seperately evaluated for each model.*

Scatter Plots

- Residuals seem to be mostly attributable to extreme dry conditions (>0.3!) for all models.

- Performance in wet conditions is variable but RF/GBRT systematically underestimate while SVR/NN is more cloudy (no systematic error).

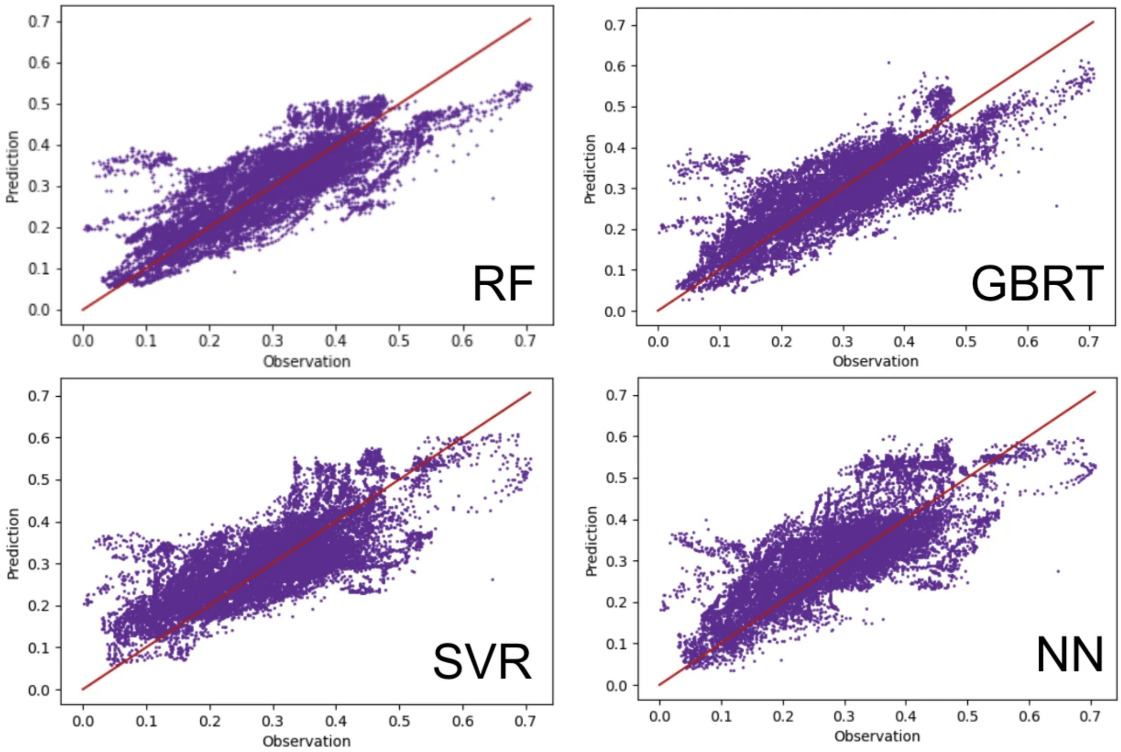


Figure 6: *Scatter plots of every model prediction vs observation and seed 12,000*

Feature Importances

- most important features:
  - day of the year (dayofyear_sin/cos, resolved as sine an cosine function)
  - surface elevation (ele_dem)
  - porosity
  - soil temperature (temp)
  - slope
  - depth (z)

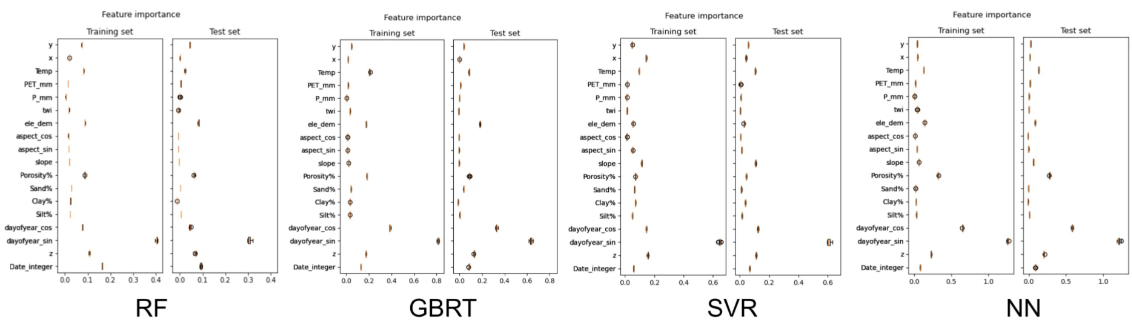


Figure 7: *Feature importances for training and test data set of every model and seed 12,000*

Pred. Soil Moisture Maps

- Shallower depth is persistently drier than the deeper layers.
- The drier zone (shallower depth) is better predicted by RF as it is better suited for data scarce situations (visual inspection).

- GBRT and SVR perform comparably to RF in deeper zones, while their performance lags behind RF in shallow depths.

- SM in area close to the stream (figure 2) is generally underestimated (with respected to the SM sensors), while SM farther north of the stream is generally overestimated.

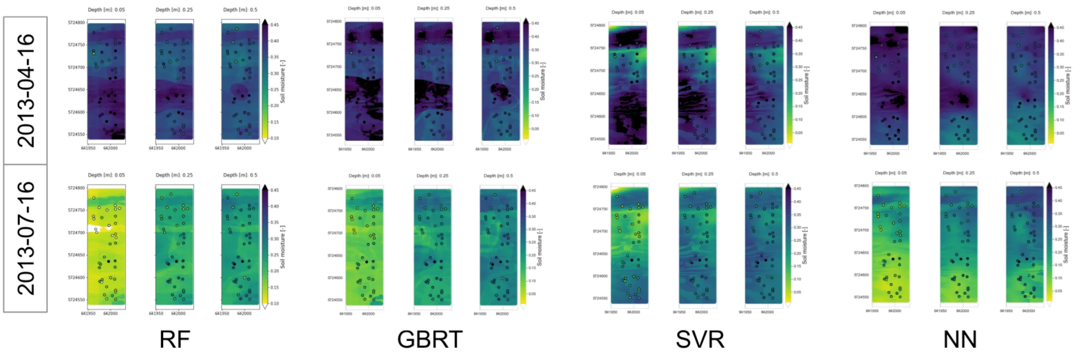


Figure 8: *Predicted soil moisture for two dates and every model for seed 12,000.*

SUMMARY AND CONCLUSION

Conclusion

- ML can be used to predict soil moisture spatially (Boog et al. 2021) but temporal prediction can prove difficult since the available temporal features such as climatic variables do not adequately describe the temporal fluctuations in soil moisture.

- Performance varies with models and initial train/test split of the data (seeds), this highlights the importance of choosing representative splits or averaging over considering multiple splits

- RF and GBRT turned out as promising and easy to parameterise first choices for exploration of the potential of ML techniques in data for spatio-temporal soil moisture mapping.

- While ML aglorithms should not be the key determinant for the performance of the model, they do require expert knowledge with respect to parameter tuning and may require high computational resources.

Table 1: *Take Home Table*

Feature	**RF/GBRT**	**SVR/NN**
Spatial performance	Perform well in both dry conditions as well as close to the stream.	Not suitable for shallow dry zones but capture the area close to the stream better.
Temporal performance	Underestimate SM in wet conditions, perform well in dry conditions	Large range of residuals in winter to spring, still better in wet conditions and deeper layers
Take aways	(relatively) straightforward implementation even with defaults	Complicated to train: convergence trouble (SVR), parameterization (NN)

Acknowledgements

The authors acknowledge the Helmholtz Centre for Environmental Research - UFZ for funding and support. The scientific results have (in part) been computed at the High-Performance Computing (HPC) Cluster EVE, a joint effort of both the Helmholtz Centre for Environmental Research - UFZ (http://www.ufz.de/ (http://www.ufz.de/)) and the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig (http://www.idiv-biodiversity.de/ (http://www.idiv-biodiversity.de/)).

Appendix

- No correlation between climatic conditions (P, PET) and SM at any depth. This is surprising as we expected soil moisture to respond to rainfall at least at the shallow most depths.

- Some (weak) correlation between PET at deeper depths. This is also surprising since we expected shallow most layer to respond to PET as well.

- Expected: Negative correlation between soil temperature and soil moisture at the shallow most depth. We surprisingly don't see this at deeper zones even though the relationship should still hold.

- Since the correlation is weak to none, the lag_correlation analysis is not applicable.

- Lack of straightforward relationships between these individual characteristics and SM motivated us to attempt to draw a relationship using ML approaches.

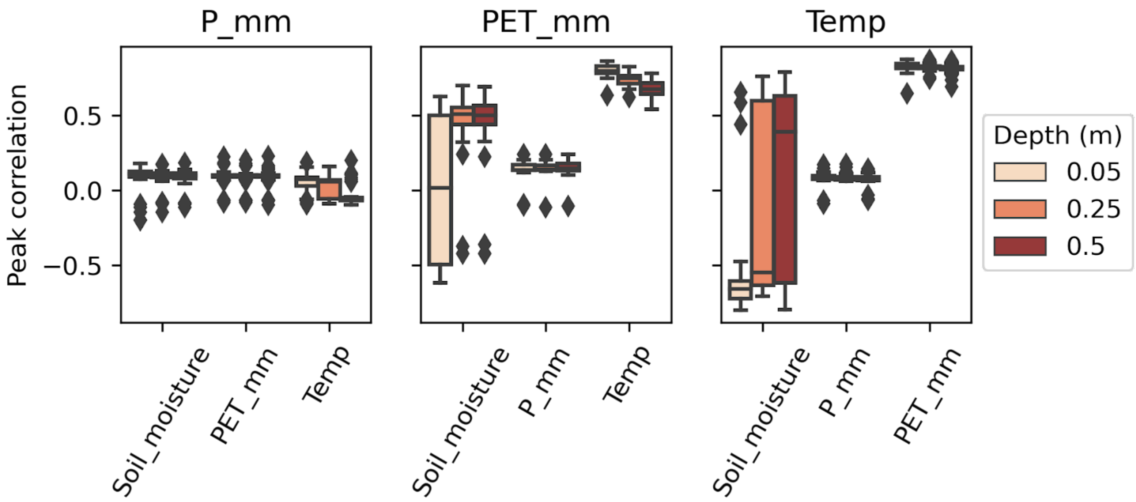


Figure 9: *Correlation analysis of temporal features.*

AUTHOR INFORMATION

Swamini Khurana: swamini.khurana@gmail.com, swamini.khurana@natgeo.su.se

Researcher, Department of Physical Geography, Stockholm University, Sweden

Doctoral researcher, Department of Environmental Microbiology, Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany

Timo Houben: timo.houben@ufz.de

Doctoral researcher, Department of Computational Hydrosystems, Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany

Pia Ebeling: pia.ebeling@ufz.de

Doctoral researcher, Department of Hydrogeology, Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany

Julia Schmid: julia.schmid@ufz.de

Doctoral researcher, Department of Ecological Modeling, Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany

Lennart Schmidt: lennart.schmidt@ufz.de

Scientist, Department of Monitoring and Exploration Technologies, Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany

Mohit Anand: mohit.anand@ufz.de

Doctoral researcher, Department of Computational Hydrosystems, Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany

Johannes Boog: johannes.boog@ufz.de

Postdoctoral researcher, Department of Environmental Informatics, Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany

ABSTRACT

Soil moisture influences several natural processes such as water and energy fluxes across the land-atmosphere interface as well as nutrient availability to plants and microbes. Therefore, accurate estimation of soil moisture is important for land management practitioners, meteorologists and hydrologists. Soil itself is an inherently spatially heterogeneous medium and soil moisture is non-linearly linked with numerous environmental factors such as climatic conditions, terrain as well as soil characteristics. This makes estimating soil moisture at the field scale using sparsely distributed point measurements a challenging task. In our work, we explored data-driven machine learning (ML) techniques to estimate soil moisture for a small agricultural catchment in time and space.

We used existing data from the TERENO site at Schäfertal in central Germany (size of 2 ha). The data includes depth specific (at 0.05, 0.25 and 0.5 m depths) daily mean soil moisture records at 30 locations from Sept 2012 to Sept 2013 as well as climate, soil properties, and terrain attributes. We identified auxiliary features and the field data density required to accurately estimate soil moisture using different algorithms: Support Vector Regression (SVR), Random Forests (RF), and Neural Networks (NN). We compared their performance with conventional geostatistical tools such as ordinary kriging. Preliminary results suggest that soil texture and temperature, and geographic features (depth, slope, elevation) are promising predictors of soil moisture. This study lays the foundation to suggest the features and data density required to estimate soil moisture at other sites with similar environmental settings. Tools that can be used by diverse practitioners may be developed based on our results and methodology.

REFERENCES

Johannes Boog, Pia Ebeling, Timo Houben, Swamini Khurana, Julia Schmid, Lennart Schmidt (2021): Using machine learning to spatially estimate soil moisture in an agricultural field. Presentation at GEOMAR Data-Science-Symposium 2021.

Edoardo Martini, Ute Wollschläger, Simon Kögler, Thorsten Behrens, Peter Dietrich, Frido Reinstorf, Karsten Schmidt, Markus Weiler, Ulrike Werban and Steffen Zacharias (2015): Spatial and Temporal Dynamics of Hillslope-Scale Soil Moisture Patterns: Characteristic States and Transition Mechanisms, Vadose Zone Journal, Vol. 14, No. 4, Wiley, doi:10.2136/vzj2014.10.0150.

LINK TO A SURVEY

Enter your survey URL here

https://example.com

Submit

Delete