

Modeling Subsurface Soil Moisture Based on Hyperspectral Data - First Results of a Multilateral Field Campaign

SINA KELLER¹, FELIX M. RIESE¹, NIKLAS ALLROGGEN², CONRAD JACKISCH³ & STEFAN HINZ¹

Abstract: Soil moisture dynamics and its spatial distribution represent important data of a landscape. Until today, a precise and spatially continuous measurement of these is difficult. Current remote sensing approaches focus on an upscaling of measured hyperspectral and hydrological data from laboratory setups by means of hydrological models. In this paper, we follow an alternative approach on a pedon-scale. We present a multilateral dataset from an irrigation campaign in August 2017 employing a comprehensive hydrological, geophysical, and hyperspectral sensor setup. In a first regression modeling approach, we apply machine learning methods to estimate subsurface soil moisture based on hyperspectral data and hydrological reference data. The derived results reveal the potential of estimating field-measured soil moisture. They provide the basis for further studies of modeling approaches based on this unique dataset.

1 Introduction

Soil moisture is a key variable of a landscape. Knowledge about its dynamics and spatial distributions is crucial for many application areas from agricultural planning to disaster mitigation. Characterizing soil moisture of field sites usually is based on sparse or incomplete experimental data. Common approaches to measure the spatial-temporal dynamics of soil moisture rely either on point-wise measurements e.g. with probes relying on estimations of the relative electrical permittivity or soil water tension, or on procedures such as satellite-based remote sensing techniques. The point-wise measurements provide a high temporal but sparse spatial resolution. Satellite-based techniques are characterized by a high spatial, heterogeneous, and poor temporal resolution. This gap between spatial and temporal resolution is well-known (RUBIN & HUBBARD 2005; ROBINSON et al. 2008).

To close this gap, researchers increasingly apply terrestrial remote sensing techniques in particular with passive sensors since the 1970s (HUETE & USTIN 2004). One possible solution opens up with recent developments in close-range hyperspectral remote sensing. It provides a high spatial and spectral resolution for analyzing the dynamics of surface signatures. To retrieve spectral signatures from the recorded hyperspectral data presents a major task in this context (cf. SALISBURY & DARIA 1992).

¹ Karlsruhe Institute of Technology (KIT), Institute of Photogrammetry and Remote Sensing, Englerstraße 7, D-76131 Karlsruhe, e-mail: [Sina.Keller, Felix.Riese, Stefan.Hinz]@kit.edu

² University of Potsdam, Institute of Earth and Environmental Science, Karl-Liebknecht-Straße 24-25, D-14476 Potsdam-Golm, e-mail: Niklas.Allroggen@geo.uni-potsdam.de

³ Karlsruhe Institute of Technology (KIT), Institute of Water and River Basin Management, Otto-Amman-Platz 1, D-76131 Karlsruhe, e-mail: Conrad.Jackisch@kit.edu

Hyperspectral sensors measure only the reflectance of the top soil layer. However, the soil surface consists of mixed features. They overlay in the reflectance spectra which results in interference. Mixed reflections often impede inference on the state and the process of the investigated object (cf. PLAZA et al. 2003; CLARK et al. 2003). Hence, such effects constitute the modeling of subsurface states and processes based on hyperspectral data as a remaining and challenging task. Many studies which deal with the estimation of soil moisture from spectral signatures reduce the overlaying mixture aspects by examining only bare soil. These studies are mentioned in the subsequent section. They conduct these experiments primarily in a laboratory setting or combine these settings with a field experiment.

From a model-based perspective, two approaches are commonly applied to estimate soil-moisture contents based on hyperspectral data: First, the widely used spectral indices which essentially engineer features by combining spectral bands (FABRE et al. 2015; OLTRA-CARRIÓ et al. 2015; VEREECKEN et al. 2014); second, the emerging data-driven machine learning approaches which increasingly are applied in remote sensing (STAMENKOVIC et al. 2013; ALI et al. 2015). The latter are justified by the potentials of machine learning to process large data, in case of satellite-based data, and handle non-linear regression problems such as modeling soil-moisture dynamics. The majority of these machine learning approaches integrate supervised learning, e.g. Random Forest (RF), Support Vector Regression (SVR), or artificial neural networks.

Despite of all progress in the combination of hyperspectral data and machine learning approaches, the uncertainty to derive estimations from the spectral surface states on subsurface conditions remains. This uncertainty of state measurements even exists in the domain of point-based pedo-hydrological monitoring (JACKISCH et al. 2018). Techniques which can quantify water distribution in the soil continuum gradually emerge (KOESTEL AND LARSBO 2014; SCHLÜTER et al. 2016, JACKISCH et al. 2017). From a process-perspective, the soil surface is known as central area related to the partitioning and redistributing of precipitation before the latter infiltrates into the soil matrix and flows in subsurface paths (among others BROOKS et al. 2015, JARVIS 2007, NIMMO 2016).

Referring to this characterization, we seek to address the complex modeling of soil moisture dynamics based on spectral data and by conducting a multi-sensor field experiment:

- by creating defined surface conditions during irrigation and subsequent drying of the study area and
- by explicitly monitoring the development of the thermal conditions in the subsurface and ground at the atmospheric boundary in one dimension with diverse sensor systems.

In addition to the multi-sensor field experiment, we present a framework which involves standard supervised learning approaches for estimating subsurface soil moisture dynamics solely with hyperspectral data in order to analyze the potential of remote sensing as a stand-alone tool. After presenting the field experiment in Section 2, we explain in detail the measured dataset in Section 3. Subsequently in Section 4, we describe the applied machine learning approaches to estimate soil moisture values and dynamics. Thereupon, we present the derived results and discuss these results in Section 5. The paper concludes in Section 6 and we give a summary of prospective studies.

2 Field Campaign

To correspond as much as possible to real-world conditions, we performed our field experiment on a grassland site and differ thereby from the bare-soil laboratory experiments. The field campaign was conducted between the 15th and 21st August 2017 in Linkenheim-Hochstetten, Germany. The grassland site on loamy sand was chosen for its behavioral hydrological properties and its accessibility.

The experiment was designed in eight plots covering an area of one square meter each (Fig. 1). The plots in row 1 are intensively monitored and the plots in row 2 functioned as replica with fewer observations. The pedo-hydrological states and dynamics of the plots were monitored using several time domain reflectometry (TDR, IMKO Pico32) probes and tensiometers (METER T5). In addition, a Bromide tracer (3 g/L KBr) was used in the irrigation water, which was resampled two hours after the end of irrigation in percussion drilled core samples with 8 cm diameter. These cores were sampled in 2 cm increments to about 40 cm depth for Bromide analysis; for methodological details see JACKISCH 2017.

The irrigation was facilitated by a drip irrigation platform (ALLROGGEN et al. 2017), fed from a common reservoir. Outflow was monitored by a mechanical meter. To maximize the number of functional replica, the irrigation scheme is based on pulses of five liter, except for the last pulse in column D with 15 L. Moreover, the irrigation intensities increased (see height of irrigation bar in Fig. 1) in steps of approximately 0.3 L/min, 0.8 L/min, 1.2 L/min, and 2 L/min. To provide data on the dynamics of subsurface conditions, we employed ground penetrating radar (GPR) measurements to record repeated shot gather and common offset profiles across the experimental plots (Fig. 1).

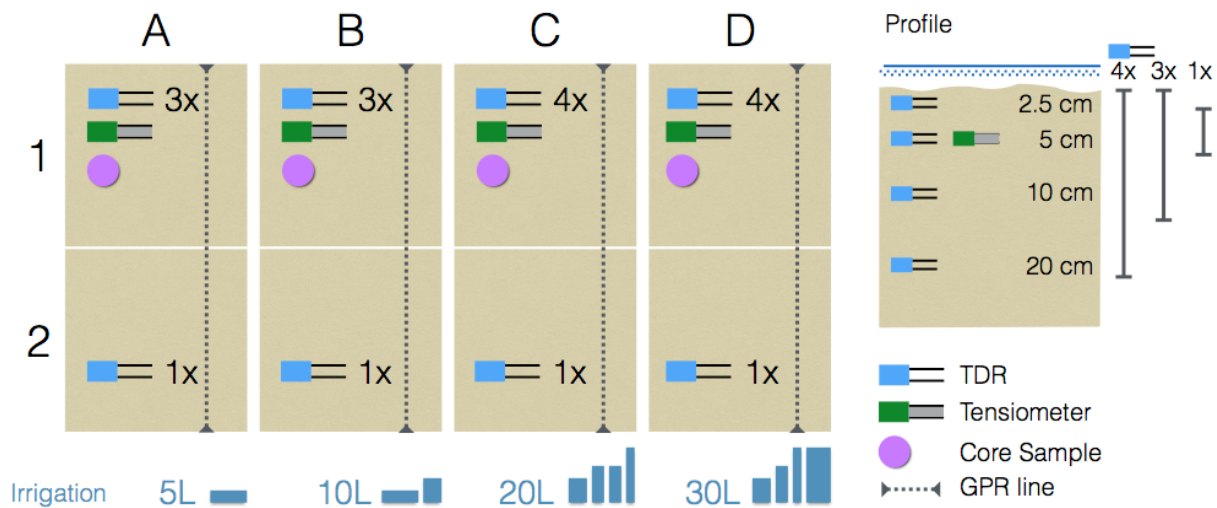


Fig. 1: Sketch of the experimental plots and the instrumentation. Each plot has a size of 1m². The irrigation pulses have been 5L each with increasing intensity. The last pulse at D1 and D2 were 15L.

Due to the high dependency of the electromagnetic wave propagation velocity to soil moisture variations, the GPR is frequently used for mapping spatial variations of soil moisture (e.g. HUISMAN et al. 2003). In principle, GPR transmits an electromagnetic wave and records the

signal at one or several receivers. The recorded signals are used to calculate the GPR propagation velocity and estimate the soil moisture content.

As remote sensing techniques, we installed a hyperspectral snapshot camera, a hyperspectral spectroradiometer, and an infrared (IR) camera mounted on a stage at 10 m distance to avoid interference with the GPR. A hyperspectral snapshot camera (Cubert) presents in this paper the primary remote sensing sensor system. It recorded frames of the plot areas. Additionally, we measured hyperspectral, point-based spectra (ASD). Thus, we are able to compare the hyperspectral snapshot data and the hyperspectral point-measured data (cf. Fig. 2). With the long-wave infrared (LWIR) measurements (FLIR), we extend the spectral range. Table 1 outlines the descriptions of the applied remote sensing sensors.

Table 1: Description of the installed remote sensing sensor systems

Sensor System	Spectrum	Channels	Measurement Frequency	Spatial Resolution
Cubert UHD 285 (hyperspectral snapshot)	450 - 950 nm	125	20 / hour	50 × 50
ASD FieldSpec 4 (spectroradiometer)	350 - 2500 nm	2151	1 / hour	1 × 1
FLIR Tau2 640 (LWIR)	7.5 - 13.5 μm	1	10 / hour	640 × 512

Fig. 2 provides a summary of the used sensors system and measured data during the field campaign. The linkage between soil-moisture values as reference data and hyperspectral images as input data is used in this paper to address the estimation of subsurface soil moisture with supervised learning approaches.

		Input Data for Modeling				
		Hyperspectral snapshot	Radiospectrometer	IR camera	GPR	Soil moisture
Reference Data	Hyperspectral snapshot				?	X
	Radiospectrometer	✓			?	X
	IR camera					X
	GPR	X	X	X		X
	Soil moisture	✓	X	X	X	

Fig. 2: Summary of the measured sensor data split into reference data (rows) and input data for modeling (columns). Green cells with “✓” characterize the data used in this paper. Cells labeled with “X” characterize the linkages of reference and input data which we will approach in prospective studies. Cells labeled with “?” symbolize linkages which we intent to study, if they can be pursued.

3 Data Processing & Analysis

3.1 Hydrological Processes

The different irrigation schemes and respective soil water redistribution result in stepwise increases of soil moisture as shown in Fig. 3. A conversion of the changes in soil moisture to the recovery of absolute water input outlines the boundaries of certainty of the experiment and sensor-based soil water monitoring. Fig. 3 cumulates the total water recovery at the sensors of the four plots (A1, B1, C1, and D1). With a perfect recovery, each step would be exactly 5 mm and the recession curves would be very gentle as in A1 and C1 of Fig. 3. The deviation can be explained an occurrence of free water, a lateral diffusive redistribution from the plot, and a non-uniform flow field. Peaking maxima and steep recession curves suggest partially quick lateral redistribution of free water.

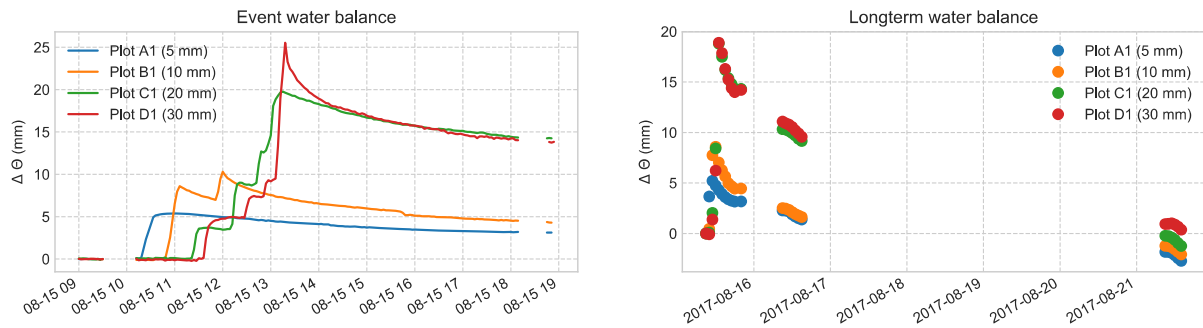


Fig. 3: Left: Water balance in the plots A1 to D1 on the day of the irrigation. Right: Water balance over the course of the experiment.

As complementary reference, we derived soil water retention properties for the test site based on undisturbed ring samples (Fig. 4). When plotting the observed pairs of soil moisture and matric potential, deviations from the laboratory-derived drying curve can be attributed to the existence of free water. As such, these periods may be least suitable to assume locally homogeneous and well-mixed states.

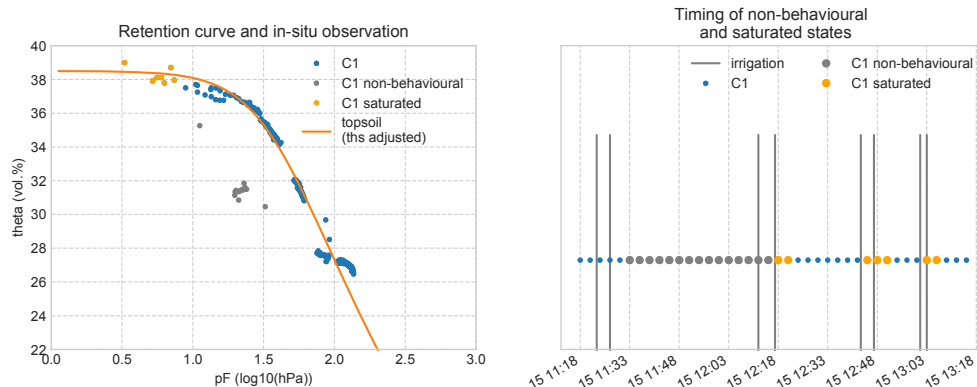


Fig. 4: Exemplary laboratory and in-situ soil water retention to identify non-behavioral states of plot C1. These mostly occur immediately after the irrigations.

As a reference for the overall event water distribution into the subsurface, the concentration profiles of Bromide tracer are considered. For the light irrigation with 5 L in 17 min (A1), infiltrating water reaches almost 0.2 m depth with a peak around 0.15 m (Fig. 5, left). Interestingly, this local peak persists over the different plots. Moreover, relative and absolute surface concentrations do not deviate much from the different irrigations. According to the cumulative recovery plots in Fig. 5 (right), it can be concluded that high irrigation intensities do not necessarily result in stronger surface signals. Furthermore, approximately 45% of the irrigated tracer mass represents the maximum of the relative recovery in the top 4 cm. At this level, we expect to deduce the surface signals.

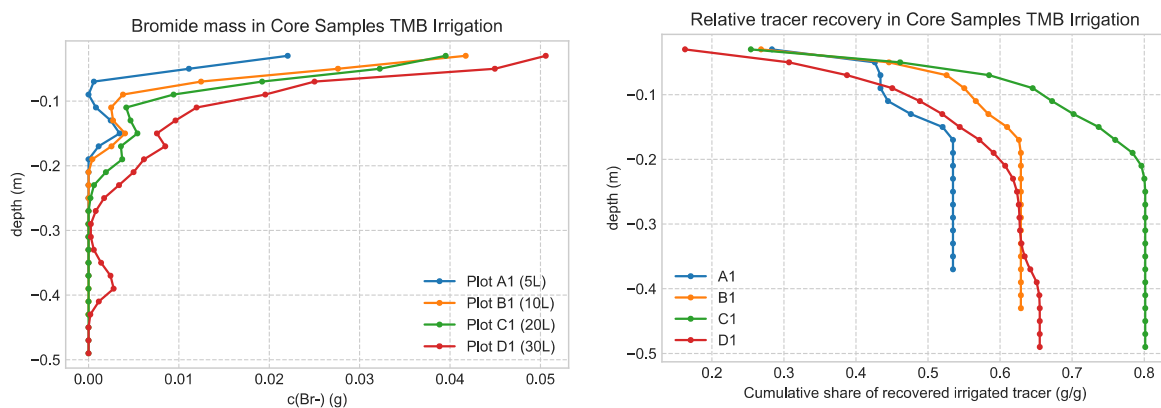


Fig. 5: Bromide tracer recovery.

During the drying process of the experiment, the surface states of the different plots return to antecedent state, with continued drying after five days. The established state variance between the plots converges only slowly (Fig. 3, right).

Based on the hydrological findings, we assume that the irrigated water established a non-uniform flow field into the subsurface and reached a depth range between 0.2 m to 0.4 m. The post-irrigation recession is not uniquely defined by the irrigation amount alone. Different mobilizations in the subsurface may result in similar surface soil moisture dynamics. This effect can be seen in Fig. 3 (left) at 20 mm and 30 mm irrigation. With respect to functional coherence, we have shown that datapoints after initial irrigation and briefly after the successive irrigation pulses should be excluded from references for the spectral signals (see Section 4).

3.2 Ground Penetrating Radar Data Processing

GPR data are collected during the first two days of the experiment to provide information on the subsurface heterogeneity and the spatial soil moisture patterns. We used a pulseEKKO PRO GPR system equipped with a pair of 1 GHz shielded antennas. To ensure accurate and repeatable positioning, the GPR antennas were attached to wooden guides and their position was recorded by a self-tracking total station while we recorded a series of common offset profiles and shot gathers (not shown here). In total, we recorded 87 GPR profiles at irregular time intervals.

The GPR data processing follows a standard processing scheme including a bandpass filter, time-based amplitude scaling, first-arrival based zero-time correction, gridding to a regular 0.01 m spacing and a Kirchhoff migration approach using a constant velocity of 0.1 m/ns.

In Fig. 6, we present a processed common offset GPR profile originating from plot B. The data show direct signals until a traveltime of 8 ns. Later signals are reflected at subsurface structures. For interpreting these data in terms of soil moisture changes, we identify traveltime differences to a reflector over the course of the experiment by applying a correlation based automatic picking algorithm starting with a seed traveltime of 15 ns (corresponding to a depth of 0.6 m). The results of this approach (exemplary shown as a red line in Fig. 6) show traveltime variations in the order of 0.5 ns. For a hydrological interpretation, we apply a petrophysical translation of traveltime differences to soil moisture variations based on the complex refractive index model (CRIM) as presented by ALLROGGEN et al. (2015) using a realistic estimate for the initial soil velocity of 0.1 m/ns.

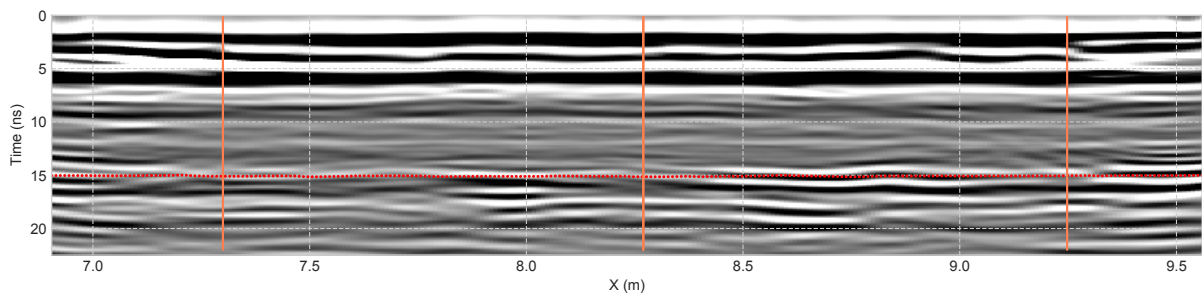


Fig. 6: Exemplary common offset GPR profile from plot B recorded at 10:57 in the morning of the first day of the experiment. Red dots mark the traveltimes selected by our correlation analysis. Orange lines mark the boundaries of the irrigation plots

In Fig. 7, we present exemplary results for the retrieved spatial soil moisture variation for plots A and B. We notice a good agreement of the total detected changes for the irrigated water (5 mm for plot A and 10 mm for plot B) in the time shortly after the end of the irrigation and a gradual drying afterwards. It should be noted that several structures in the signal persist over the course of the experiment. With regard to remote sensing application this method provides spatial details on the interpretation of soil moisture dynamics. However, one has to notice the integration depth of 0.6 m. Thus, surface moisture cannot be resolved. For making these signals possible, we recorded shot gathers during the field experiment, which by analyzing the direct arrivals can provide information on the shallower soil moisture variations. With regard to point-based soil moisture measurements, the deviations in event water recovery (Fig. 4) can also be attributed to limited representativity of single TDR probes in the diverse state field.

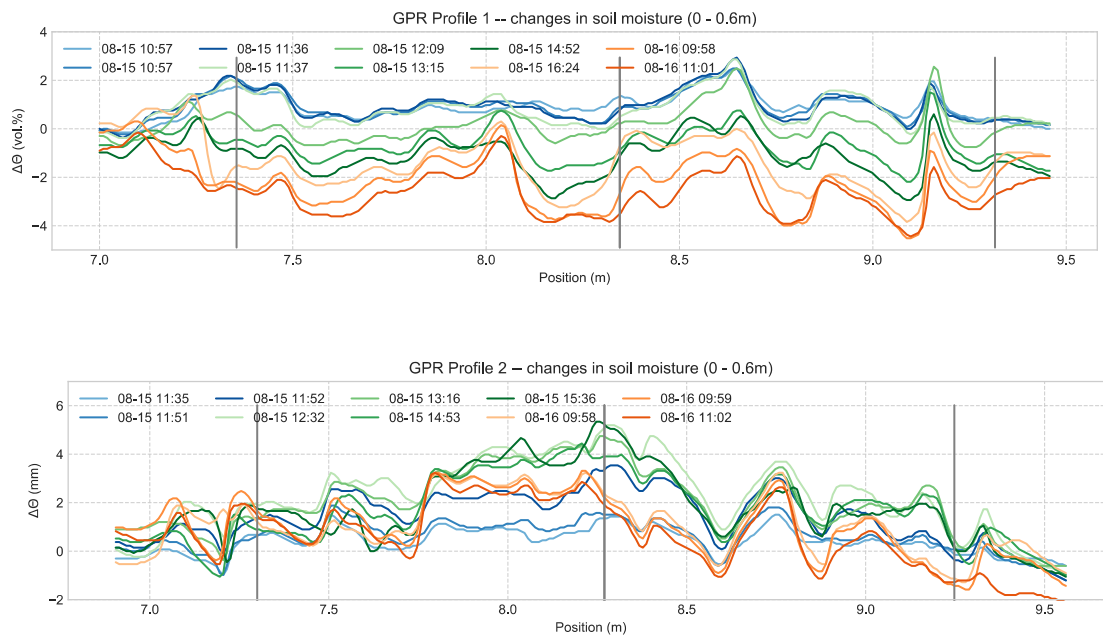


Fig. 7: Changes in soil moisture along common offset profile inferred from changes in apparent two-way travelttime to reflector.

3.3 Remote Sensing Data Processing

Since the hyperspectral snapshot sensor provides the highest temporal resolution, the major analysis in the remote sensing part of this paper is performed on the snapshot data. To apply machine learning (ML) approaches, the feature selection presents a prerequisite for the latter.

In this paper, we conduct two different feature selection procedures. First, the average spectra of each plot and measurement are calculated with respect to the approximated position of the TDR probes (Fig. 8, left). The wooden guides (cf. Fig. 8, left) are masked and ignored in the calculation of the average spectrum. The spectralon (cf. Fig. 8, left) functions as white reference of the spectrum in every image. Second, we dismiss five bands at the beginning and five bands at the end of the spectrum to remove occurring sensor artefacts. On the basis of the comparison in Fig. 8 (right) between a spectroradiometer and a snapshot, these artefacts are identifiable. The spectroradiometer data functions as a cross check for the snapshot data. The shapes of both spectra are similar, but the amplitudes of both distributions differ clearly. The observed offset indicates differences in the calibration and measurement angle.

Two aspects have to be considered: First, the estimation in the following sections relies on a specific sensor, in this case the hyperspectral snapshot sensor. Second, with respect to the similar spectra shapes, we assume that both hyperspectral sensors measured the same reflectance of the surface soil moisture states.

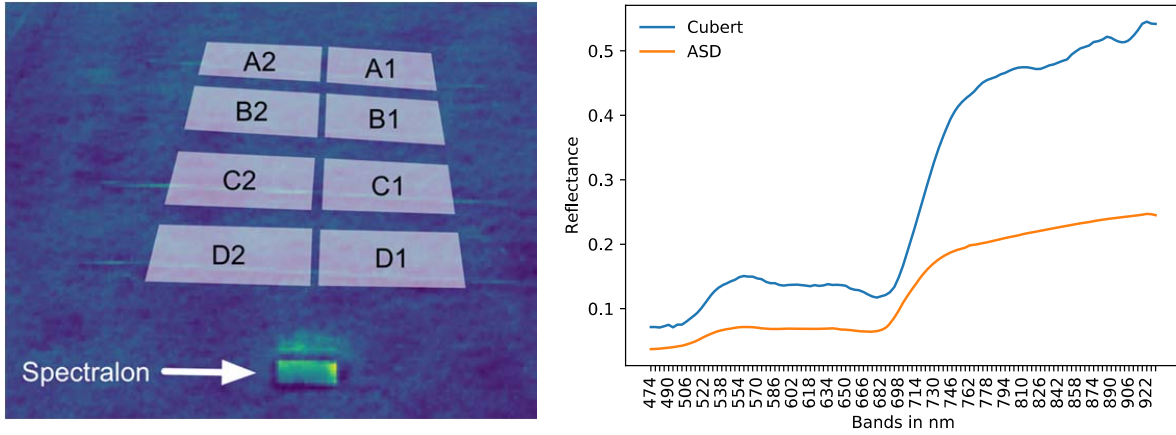


Fig. 8: Left: An example of a hyperspectral snapshot. The 50×50 image is pan-sharped to 1000×1000 due to visualization purposes. In context of the modeling approach, the raw images with the 50×50 resolution are used. The dark areas are masked and, based on the white grid areas, the average spectra are calculated. Right: Exemplary spectra of the spectroradiometer and the snapshot sensors recorded at the same plot and at the same time.

In addition to the hyperspectral data acquisition, the infrared (IR) data are averaged to calculate an average temperature value. No further pre-processing is applied on the IR data. After a pre-analysis of the data, we decided to exclusively focus on the estimation quality of the hyperspectral data importance in the regression. Within the entire multilateral dataset, the IR as well as the hyperspectral spectroradiometer measurement are included and can be analyzed in future studies.

In the following, we refer to the measured soil moisture value (5 cm) as reference data combined with the simultaneous recorded, pre-processed as well as averaged hyperspectral data as a single datapoint. In total, the dataset consists of 1940 datapoints.

4 Regression Methods

The supervised learning approach is performed with the hyperspectral image data as input vector and the soil moisture data as desired output value. For the soil-moisture estimation, we apply common regression methods i.e. Random Forest (RF) (cf. BREIMAN 2001), extremely randomized trees (ET) (cf. GEURTS et al. 2006) as an extension of the RF algorithm, as well as Support Vector Regression (SVR) (cf. VAPNIK 1995). These regression methods represent model-independent approaches.

The RF and the ET regressors are implemented in the Python package *scikit-learn* (PEDREGOSA et al. 2011). We perform the regression with the default settings of hyperparameters, defining the number of estimators in both methods. Hyperparameters are parameters set before the estimator's training process. The regression result converges with 1000 estimators. The RF and ET regressors rank the importance of the input variables after the training (cf. Section 5). The

support vector regressor (SVR) is based on the package *libsvm* (CHANG 2011) and is also implemented in Python's *scikit-learn*. The SVR hyperparameters are tuned with a grid-search approach.

The dataset is split randomly into a train subset of 382 datapoints and a test subset of 1558 datapoints. For a reasonable modeling of continuous soil moisture variables, the distributions of both subsets have to be representative of the soil moisture distribution in terms of distribution range and, if possible, in terms of distribution shape. As outlined in Section 3.1, the soil moisture distribution differs between the different plots. Thus, we have to consider these effects in context of the soil moisture modeling based on hyperspectral data. Pre-emptively, we exclude combinations between two plots (e.g. A1 and B1) for the training and validation.

Providing the regressor with data from all plots enables the machine learning approaches to identify spectral bands which are correlated to the soil-moisture values. Fig. 9 provides histograms of the soil moisture distribution in the test and training subsets. All datapoints have been cross-checked with the hydrological findings introduced in Section 3.1. In coherence with the hydrological interpretation of the observations, we exclude datapoints which are measured after initial irrigation and briefly after the successive irrigation pulses.

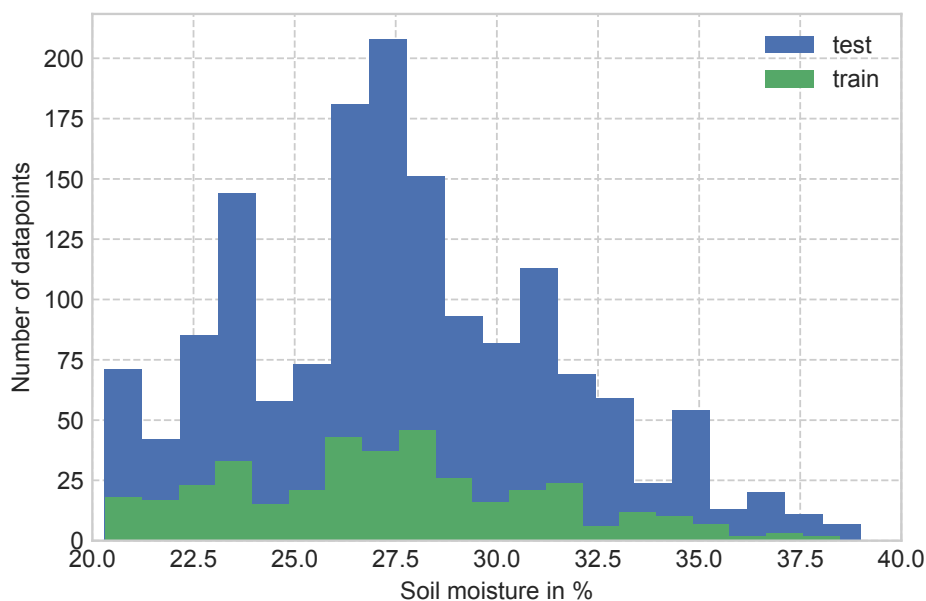


Fig. 9: Soil-moisture distribution of the test (blue) and train (green) subset.

The regression parameters are learned from conjugate single input and reference datapoints. Thus, time is not used in these approaches, neither for training nor for estimation. By this, it is possible to estimate soil moisture also from one single hyperspectral image, like in the satellite-case. The modeling process is outlined in Fig. 10.

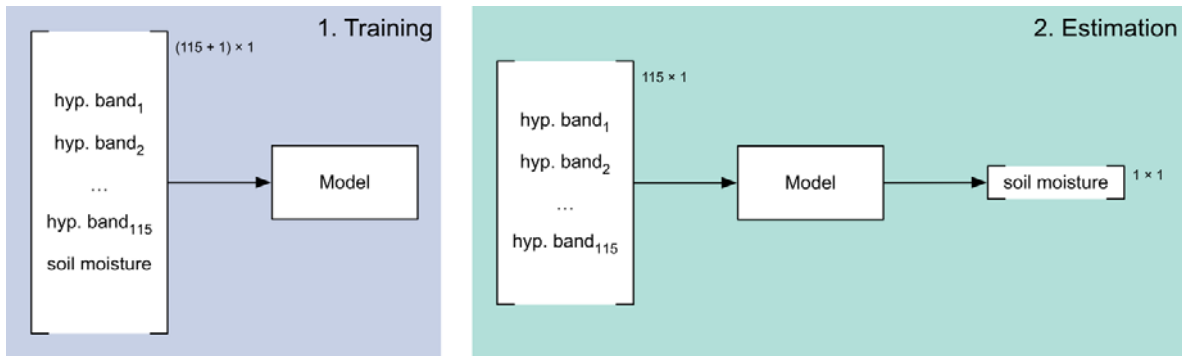


Fig. 10: Regression schema for estimating soil moisture

5 Results & Discussion

The regression results are summarized in Fig. 11 and Table 2. The SVR performs badly with a coefficient of determination R^2 of 56%. RF and ET outperform the SVR to solve the presented regression problem. RF reaches a R^2 of 73% and a root mean squared error (RMSE) of 2.00 (%-soil moisture). The best regression results are achieved by the ET with a R^2 score of 78% and a RMSE of 1.81 (%-soil moisture).

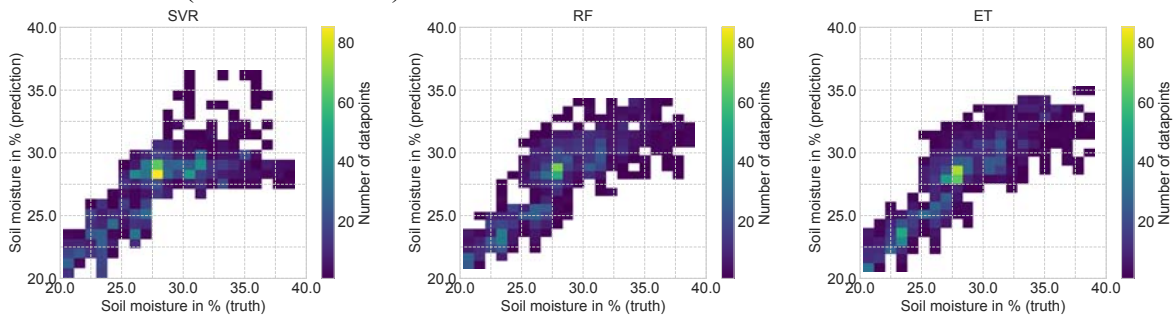


Fig. 11: Regression results for the SVR, RF, and ET.

Table 2: Regression results with the coefficient of determination R^2 and the RMSE.

	R^2 in %	RMSE in %-soil moisture
SVR	56.38	2.58
RF	73.69	2.00
ET	78.41	1.81

Fig. 12 shows the feature importance of the hyperspectral bands based on the ET regressor as well as the mean spectrum of the dataset with its respective standard deviation. In contrast to widely used spectral indices, the supervised learning approaches RF and ET weight the input vector (hyperspectral data) by means of their importance for the estimation quality. Considering the feature importance, we recognize that the spectral bands between 720 nm and 850 nm are significantly more important than the remaining ones. While the variance of the mean spectrum is relatively large in the upper third of the band spectrum, the feature importance distribution shows a peak between 720 nm and 850 nm. The bands between 850 nm and 930 nm possess a minor feature importance and a large variance. This finding indicates that these band exhibit noise, e.g. occurring due to weather and sensor conditions. Furthermore, the RF and ET separate the reflectance of vegetation and soil moisture without further information on the underlying processes is inserted.

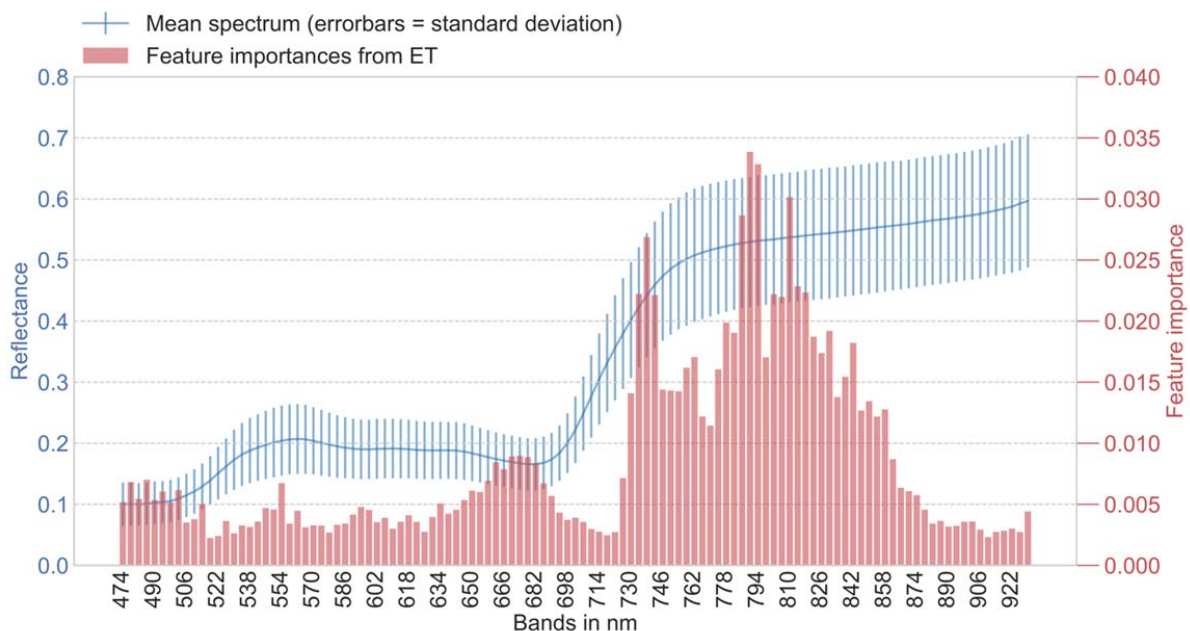


Fig. 12: Mean spectrum of the dataset with its standard deviation as error bars (blue line), feature importance of the ET regressor (red bar plot).

6 Conclusion & Outlook

In this paper, we investigated the potential of hyperspectral data for surface soil moisture modeling in grass-land like areas. We acquired a multi-sensor dataset measured in August 2017 in Linkenheim-Hochstetten, Germany. The resulting dataset consists of data from a suite of different hydrological, geophysical, and remote sensing systems. We presented the respective sensor data as an overview. In extension to previous remote sensing studies focusing on bare soil

experiments in the context of modeling surface soil moisture, we transferred the data acquisition to a grass side.

As input data hyperspectral and hydrologically reasonable data are used. The applied standard machine learning approaches only rely on the described input data: only artefacts were excluded, but no pre-processing was conducted. The derived first results reveal that an estimation of surface soil moisture can be achieved based on hyperspectral data in combination with machine learning approaches. An additional implementation including e.g. tuning of regressors will be conducted in future studies. Furthermore, we demonstrated that an experiment at a complex environment, e.g. grass soil, is feasible. We showed that the proposed supervised methods are able to handle this complexity, if the proposed setting of hyperspectral input and desired soil moisture data is given.

Coming in short to the hydrological perspective, further research will be conducted in modeling the excluded datapoints (measurements immediately at the irrigation point). One promising aspect to tackle this challenge might be to consider the infrared data additionally in the context of machine learning. The GPR data show a level of spatial heterogeneity, which was not modeled in this study. Nevertheless, when studying spatial variations this method provides the possibility for validating and training spatial classification approaches.

For prospective studies, we intend to publish the aggregated dataset.

7 Acknowledgements

We thank our colleagues Torsten Schuler, Jürgen Hauth, and Armin Hofmann from the Institute of Technology and Management in Construction at the Karlsruhe Institute of Technology for their technical support. We also thank Wolfgang Groß from the Fraunhofer Institute of Optronics, System Technologies and Exploitation for lending us a spectroradiometer and the research group “Catchments as Organized Systems” for collaborative discussions.

8 References

- ALI, I., GREIFENEDER, F., STAMENKOVIC, J., NEUMANN, M. & NOTARNICOLA, C., 2015: Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data. *Remote Sensing*, **7**, 16398-16421.
- ALLROGGEN, N., VAN SCHAİK, N.L.M.B. & TRONICKE, J., 2015: 4D ground-penetrating radar during a plot scale dye tracer experiment. *Journal of Applied Geophysics*, **118**, 139-144.
- ALLROGGEN, N., JACKISCH, C. & TRONICKE, J., 2017: Four-dimensional gridding of time-lapse GPR data, presented at the 2017 9th International Workshop on Advanced Ground Penetrating Radar (IWAGPR), 1-4.
- BREIMAN, L., 2001: Random forests. *Machine Learning*, **45**(1), 5-32.
- BROOKS, P. D., CHOROVER, J., FAN, Y., GODSEY, S. E., MAXWELL, R. M., MCNAMARA, J. P. & TAGUE, C. (2015). Hydrological partitioning in the critical zone: Recent advances and opportunities for developing transferable understanding of water cycle dynamics. *Water Resources Research*, **51**(9), 6973-6987.

- CHANG, C.C., 2011: LIBSVM: A library for support vector machine. *ACM Transactions on Intelligent Systems and Technology*, **2**(3), 1-27.
- CLARK, R.N., SWAYZE, G.A., LIVO, K.E., KOKALY, R.F., SUTLEY, S.J., DALTON, J.B. & GENT, C.A., 2003: Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems. *Journal of Geophysical Research: Planets*, **108**(E12).
- FABRE, S., BRIOTTET, X. & LESAIGNOUX, A., 2015: Estimation of Soil Moisture Content from the Spectral Reflectance of Bare Soils in the 0.4–2.5 μm Domain. *Sensors*, **15**, 3262-3281.
- GEURT, P., ERNST, D. & WEHENKEL, L., 2006: Extremely randomized trees. *Machine Learning*, **63**(1), 3-42.
- HUETE, X. & USTIN, Y., 2004: Manual of Remote Sensing, Remote Sensing for Natural Resources Management and Environmental Monitoring, *Manual of Remote Sensing*, **4**.
- HUISMAN, J. A., HUBBARD, S. S., REDMAN, J. D. & ANNAN, A. P., 2003: Measuring Soil Water Content with Ground Penetrating Radar: A Review. *Vadose Zone Journal*, **2**(4), 476-491.
- JACKISCH, C., ANGERMANN, L., ALLROGGEN, N., SPRENGER, M., BLUME, T., TRONICKE, J. & ZEHE, E., 2017: Form and function in hillslope hydrology: in situ imaging and characterization of flow-relevant structures. *Hydrol. Earth Syst. Sci.*, **21**(7), 3749-3775.
- JACKISCH, C., GRAEFF, T., GERMER, K. & DURNER, W., 2018: Soil moisture and matric potential – what do we measure? *Vadose Zone Journal*, in submission.
- JARVIS, N. J., 2007: A review of non-equilibrium water flow and solute transport in soil macropores: principles, controlling factors and consequences for water quality. *European Journal of Soil Science*, **58**(3), 523-546. <http://doi.org/10.1111/j.1365-2389.2007.00915.x>
- KOESTEL, J. & LARSBO, M., 2014: Imaging and quantification of preferential solute transport in soil macropores. *Water Resources Research*, **50**(5), 4357-4378.
- NIMMO, J. R., 2016: Quantitative Framework for Preferential Flow Initiation and Partitioning. *Vadose Zone Journal*, **15**(2).
- OLTRA-CARRIÓ, R., BAUP, F., FABRE, S., FIEUZAL, R. & BRIOTTET, X., 2015: Improvement of Soil Moisture Retrieval from Hyperspectral VNIR-SWIR Data Using Clay Content Information: From Laboratory to Field Experiments. *Remote Sensing* **7**, 3184-3205.
- PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M. & DUCHESNAY, E., 2011: Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, **12**, 2825-2830.
- PLAZA, A., MARTÍNEZ, P., PLAZA, J., & PÉREZ, R., 2003: Spatial/spectral analysis of hyperspectral image data. *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, 298-307.
- ROBINSON, D. A., CAMPBELL, C. S., HOPMANS, J. W., HORNBUCKLE, B. K., JONES, S. B., KNIGHT, R. & WENDROTH, O., 2008: Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review. *Vadose Zone Journal*, **7**(1), 358-389.
- SALISBURY, J.W. & D'ARIA, D. M., 1992: Emissivity of terrestrial materials in the 8–14 μm atmospheric window. *Remote sensing of Environment*, **42**(2), 83-106.
- SCHLÜTER, S., BERG, S., RÜCKER, M., ARMSTRONG, R.T., VOGEL, H.-J., HILFER, R. & WILDENSCHILD, D., 2016: Pore-scale displacement mechanisms as a source of hysteresis

for two-phase flow in porous media. *Water Resources Research*, **52**(3), 2194-2205.
<http://doi.org/10.1002/2015WR018254>

VAPNIK, V., 1995: *The nature of statistical learning theory*. Springer science & business media.

VERECKEN, H., HUISMAN, J.A., PACHEPSKY, Y., MONTZKA, C., VAN DER KRUK, J., BOGENA, H. & VANDERBORGHT, J., 2014: On the spatio-temporal dynamics of soil moisture at the field scale. *Journal of Hydrology*, **516**, 76-96.