Are the Current Expectations for SAR Remote Sensing of Soil Moisture Using Machine Learning Overoptimistic?

Liujun Zhu, *Member, IEEE*, Junjie Dai, Junliang Jin, Shanshui Yuan, Ziwei Xiong, and Jeffrey P. Walker, *Fellow, IEEE*

***Abstract*— High-resolution surface soil moisture is essential for advancing various applications. The increased synthetic aperture radar (SAR) missions over the past decade present an opportunity to obtain large-scale, high-resolution soil moisture data. Machine learning methods are increasingly used for this purpose, but they generally suffered from the availability of ground-based observations. The real performance in view of a global product is still unclear. Consequently, commonly used machine learning methods were evaluated in this study in simulated global mapping scenarios with few training data, using a global dataset of 209 318 samples from 1021 locations worldwide, and a unique regional dataset with intensive ground and airborne-derived soil moisture from L-band passive microwave observations. Three evaluation scenarios based on the global dataset were involved, with ≤5% samples used for training. The target accuracy of 0.06 m^3/m^3 was only met in the dependent evaluation scenario, where the training and testing samples were randomly split. In the temporal evaluation scenario and spatial evaluation scenario, where training and testing samples came from different time periods or locations, the best models achieved median root-mean-square errors (RMSEs) of only 0.078 and 0.089 m^3/m^3, respectively. The evaluation on the regional dataset showed consistently worse accuracy statistics (RMSE > 0.1 m^3/m^3 and R < 0.41). Moreover, all methods failed to capture the spatial patterns of soil moisture, compared to airborne-derived passive soil moisture maps. These findings, therefore, suggest that current expectations for SAR-based soil moisture estimation using machine learning may be overoptimistic, requiring more robust approaches for scenarios with sparse ground measurements.**

***Index Terms*— Data-scarce scenario, machine learning, Sentinel-1, soil moisture.**

I. Introduction

ACCURATE soil moisture data have been used to support a wide range of applications, e.g., precipitation estimation [1], flood prediction [2], drought monitoring [3], and crop yield assessment [4]. Despite great efforts on modeling or ground observation of soil moisture [5], [6], spaceborne remote sensing is the only viable avenue to capture global soil moisture data with both high spatial and temporal detail. Among the various remote sensing techniques, synthetic aperture radar (SAR) is currently the most promising technique to measure high-resolution (<1 km) soil moisture over large spatial scales operationally [7], [8].

Over the past decades, numerous approaches have been developed for SAR remote sensing of soil moisture, typically involving the inversion of SAR observations or the decomposition of surface scattering using scattering models [9]. Since the scattering models often struggle to accurately represent real surface scattering with a fixed parameterization solution [10] and the available SAR measurements are fewer than the unknowns [11], the SAR retrievals have not achieved the same level of accuracy as their passive counterpart. Despite the challenges, near-operational soil moisture monitoring from Sentinel-1 has been made available using the change detection [12], [13], [14] and time series [15], [16] methods, by assuming that the temporal dynamic of soil moisture is much faster than that of other surface parameters. Moreover, the potential uncertainties of forward scattering modeling, assumptions, and data preprocessing steps have been partly removed using an ensemble of time series retrievals [15], [17]. However, further improvements of these methods, including accuracy, feasibility, and stability, are both necessary and challenging, requiring additional development of scattering models and preprocessing platforms of big SAR data.

Data-driven approaches have emerged as a more promising alternative for large-scale soil moisture retrieval, without the need for complex physical modeling [18]. Conventional machine learning methods have confirmed the capability to retrieve soil moisture content from remote sense data [19]. Commonly used methods include the multilayer perceptron (MLP), support vector regression (SVR), random forest (RF), and eXtreme Gradient Boosting (XGBoost) methods, with a great number of successful applications from regional [20], [21] to global [22], [23] scales. Since it is convenient to integrate data from different sensors and/or sources, surface terrain features, multispectral data, microwave data, soil proprieties, and/or weather data have been commonly used jointly for improved accuracy [24], [25]. Some studies have

Received 9 October 2024; revised 4 January 2025; accepted 20 January 2025. Date of publication 24 January 2025; date of current version 7 February 2025. This work was supported in part by the National Key Research and Development Program under Grant 2023YFC3209800 and in part by the National Natural Science Foundation of China under Grant 42101374 and Grant 42371369. *(Corresponding author: Junliang Jin.)*

Liujun Zhu is with the Yangtze Institute for Conservation and Development, Hohai University, Nanjing 210098, China, and also with the Department of Civil Engineering, Monash University, Clayton, VIC 3800, Australia (e-mail: liujun.zhu@hhu.edu.cn).

Junjie Dai, Junliang Jin, and Shanshui Yuan are with the Yangtze Institute for Conservation and Development, Hohai University, Nanjing 210098, China (e-mail: jljin@hhu.edu.cn).

Ziwei Xiong and Jeffrey P. Walker are with the Department of Civil Engineering, Monash University, Clayton, VIC 3800, Australia (e-mail: Ziwei.Xiong@monash.edu; Jeff.Walker@monash.edu).

Digital Object Identifier 10.1109/TGRS.2025.3533927

conducted comprehensive evaluations and comparisons of multiple machine learning methods [24], [26], with ensemble methods, e.g., RF and XGBoost, generally achieving better results.

The recent progress in deep learning enables more powerful models for soil moisture retrieval. Convolutional neural networks (CNNs), long short-term memory networks (LSTMs), and transformers are among the most used architectures, valued for their proficiency in feature extraction and learning. In particular, the CNN algorithm can autonomously extract features relevant to soil moisture such as soil texture, vegetation cover, and surface roughness from sequences of SAR images, thereby improving prediction accuracy [27]. In contrast, the LSTM algorithm is highly effective in analyzing temporal sequences, as it can capture both long- and short-term dependencies crucial for predicting soil moisture dynamics influenced by seasonal and climatic variations [28]. Transformers, which have become the most successful architecture for time series analysis, have also been successfully applied to soil moisture estimation and prediction [29], [30].

Despite the great success, the effectiveness of these algorithms is heavily dependent on the size and quality of the dataset. Generally, more training data leads to better results, and the closer the distribution of data between the training and prediction sets (i.e., less domain shift), the better the model performs. Accordingly, most studies focused on scenarios with ample in situ measurements, where the training set is typically >50% of the entire dataset [23]. However, the availability of ground truth measurements for training a robust model is often limited by the high cost of data sampling and human constraints, being typically insufficient for large-scale applications. The International Soil Moisture Network (ISMN) [31] has made incredible efforts to aggregate and harmonize soil moisture measurements collected from different ground-based monitoring stations worldwide. However, the available in situ measurements still have a large discrepancy in space and time, with a limited number of stations operating in a continuous manner. While soil moisture products with a spatial resolution of 1 km have been developed using machine learning methods [22], [23], the effectiveness of these methods may be exaggerated in areas and/or periods with limited ground samples. In other words, the generalization capability of existing machine learning approaches has not been fully evaluated in scenarios with domain shifts.

Consequently, four widely used machine learning algorithms, i.e., MLP, SVR, RF, and XGBoost, have been evaluated here in four scenarios characterized by sparse training data. Although deep learning methods are generally more powerful, they were not included in this study because they require a much larger training set than conventional machine learning methods and are more sensitive to scenarios with limited training samples [18]. Four evaluation scenarios were established using all of the ISMN data from 2016 to 2021, along with intensive ground soil moisture measurements collected from six airborne campaigns conducted in the Yanco agricultural area, Australia. The first three scenarios focused on global applications using the ISMN dataset, evaluating the capability of machine learning methods: without domain shift, as well as their transferability across time and space with domain shifts. The fourth scenario was focused on a local-scale application in the Yanco area, Australia, addressing whether a sparse local soil moisture network is sufficient for regional soil moisture retrieval. These scenarios covered the full range of potential machine learning applications in soil moisture retrieval, providing experimental insights into the accuracy for areas or periods lacking in situ measurements.

II. Data and Preprocessing

A. Soil Moisture

1) Global Soil Moisture From the ISMN Dataset: The soil moisture records of 2016–2021 archived in the ISMN database were used for training and evaluation at 1 km, including 25 networks across the world (refer to the Appendix). Notably, the OzNet soil moisture monitoring network located in Australia was excluded from the global analysis and reserved for the regional scenario. Due to variations in sensor numbers, measurement depths, and recording frequencies at different stations, a number of processing steps were required. First, the measurements from shallow soil layers ($\leq$5 cm) on the same calendar date were averaged to obtain the daily near-surface soil moisture at the point. These daily averages within the same EASE-2.0 1-km grid cell were further averaged to serve as the spatial estimate of ground truth for model training and evaluation. Fig. 1 summarizes the network configurations, locations, the number of grid cells, and the number of available samples for each network. In total, 1021 grid cells with 1-km resolution were included, with the majority located in the Continental United States (CONUS) and Europe. Only the daily averaged soil moisture data coinciding with Sentinel-1 acquisitions were retained, resulting in 207 826 samples.

2) Regional Soil Moisture From the Yanco Area: The regional dataset was collected from the Yanco agricultural area, New South Wales, Australia. The Yanco area is one of the core calibration/validation sites of the SMAP mission, being a semi-arid climate with an annual average rainfall of approximately 400 mm [32]. Thirty-nine soil moisture monitoring stations were established in this area, forming the main part of the OzNet network [32]. Among them, 14 stations were installed in 2003, with their measurements included in the ISMN dataset, while the remaining stations were specifically set up for a series of airborne campaigns since 2009 [33]. Moreover, some stations were upgraded in 2019 to measure soil moisture at deeper layers. The measurement time intervals at these sites were either 20 or 30 min. To be consistent with the ISMN dataset preprocessing, daily averaged soil moisture for the 0–5-cm layer was first extracted and then upscaled to a 1-km resolution. A total of 4106 samples with coincident Sentinel-1 observations were achieved for the period of 2016–2021.

A total of six airborne field campaigns were also conducted since the launch of Sentinel-1 (April 3, 2014) in the Yanco area, aimed at developing, calibrating, and validating state-of-the-art microwave remote sensing concepts and algorithms for soil moisture. These campaigns include the Soil Moisture

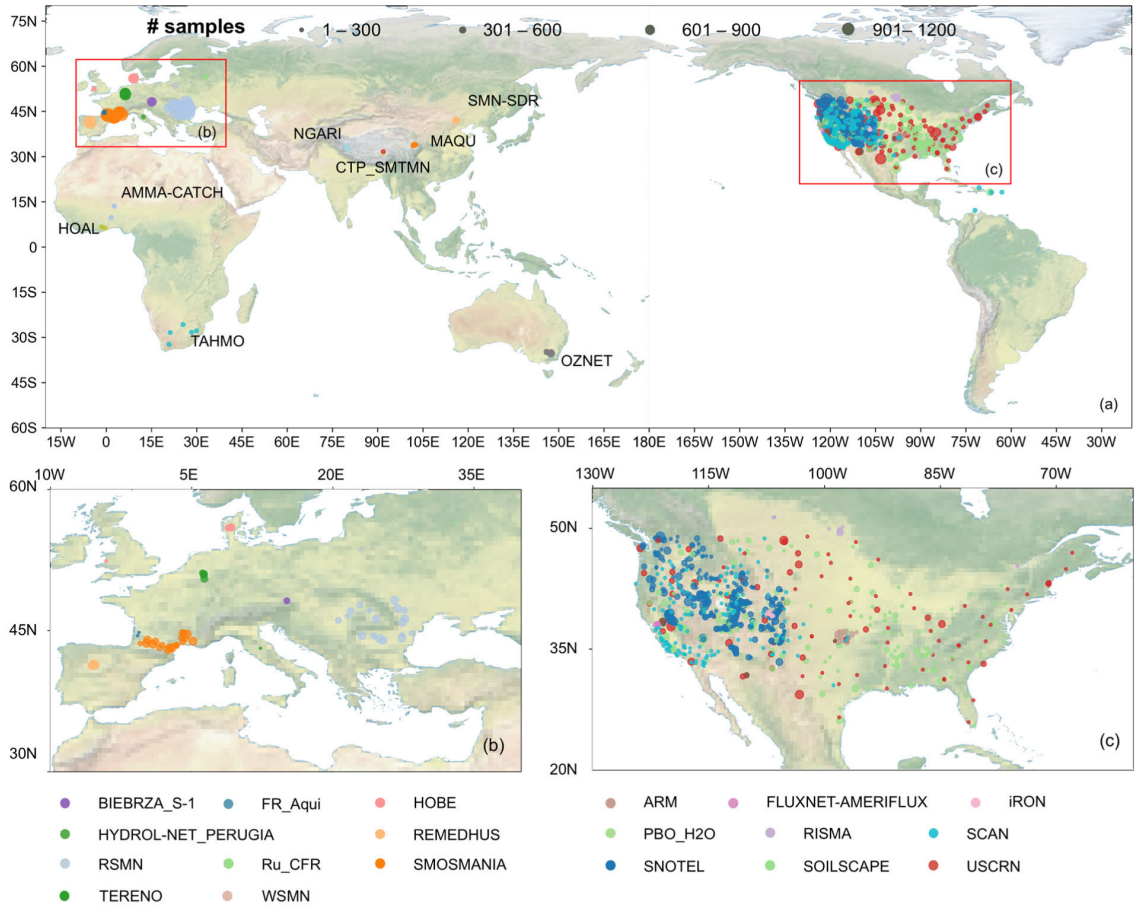

Fig. 1. Spatial distribution of available samples. Dots of the same color indicate that these sites belong to the same network, with each dot being a 1-km grid cell. The size of dots indicates the number of samples.

Active Passive Experiment-4 and -5 (SMAPEx-4 and -5, Ye et al. [32]) conducted in 2015, the P-band Radiometer Inferred Soil Moisture (PRISM-1 and -2) in 2019 and 2021, and the Active and Passive P-band and L-band Experiment-1 and -2 (APPLEx-1 and -2) conducted in 2022. A summary of each campaign is presented in Table I. While the focus areas of each campaign were slightly different, they were all located in the flight area of SMAPEx-4 and -5, covering 71 × 89 km (see Fig. 2). Despite the different motivations, intensive ground soil moisture measurements for the 0–5-cm layer were collected during each campaign in at least three focus areas using the Hydra-probe Data Acquisition System (HDAS) [34], with uniform grid spacings of either 250 or 50 m (see Table I and Fig. 2). These measurements were upscaled to the same 1-km grids by taking the simple average, resulting in a total of 2435 samples.

In addition to the ground soil moisture sampling, L-band brightness temperature data was simultaneously collected during each campaign at a resolution of $\leq$1 km. High-accuracy soil moisture was then estimated using a calibrated Tau–Omega model, which was initially employed for the calibration and validation of various active/passive soil moisture retrieval and/or downscaling methods [32]. Only the soil moisture retrievals from the SMAPEx-4 and -5 were available with an overall RMSE of $\sim$0.06 m^3/m^3 on the Sentinel-1 acquisition dates [32]. The relatively higher retrieval error was primarily due to the overestimation of high soil moisture values ($>$0.3 m^3/m^3), with the spatial patterns of the ground measurements being well captured. For each campaign, only two Sentinel-1 acquisitions were collected, given its 12-day revisit interval in the Yanco area. Consequently, the four corresponding L-band soil moisture maps were used for inter-comparison, with a particular focus on analyzing the spatial patterns.

B. Predictor Attributes

The selection of predictor attributes (features) is crucial to the performance of machine learning methods. Rigorous feature selection can definitely improve results, but this process becomes challenging when only a limited number of training samples are available. Therefore, the predictor attributes in this study were chosen based on the radar scattering process and the key drivers of soil moisture in hydrologic processes, aligning with the approach of many existing studies [24], [25]. These features were categorized into two groups: dynamic and static predictors, depending on their temporal variability. The data used in this study include the Sentinel-1 SAR data, Sentinel-2 optical data, SMAP radiometer data, normalized difference vegetation index (NDVI), and a few other auxiliary variables such as soil texture and terrain features.

TABLE I
SUMMARY OF THE SIX AIRBORNE CAMPAIGNS MADE IN THE YANCO AGRICULTURE AREA AUSTRALIA SINCE THE LAUNCH OF SENTINEL-1

Experiment Name	Date	Motivation	Focus areas for in-situ soil moisture	Airborne data used in this study
SMAPEx-4	May 1–22, 2015	Calibration/validation of the SMAP active/passive concept	Six 3×3 km focus areas, with a sampling grid of 250 m	Soil moisture retrieved by the L-band airborne radiometer at 1 km
SMAPEx-5	Sep. 7–27, 2015			
PRISM-19	Sep.28 – Oct.19, 2019	Demonstration of using a P band radiometer for measuring soil moisture of top ~15 cm	Three 3×3 km focus areas, with sampling grids of either 50 m or 250 m	
PRISM-21	Mar. 8–26, 2021			
APPLEx-1	Jan. 27–Feb 26, 2022	Demonstration of an active/passive L and P band concept for soil moisture profile estimation	Four 3×3 km focus areas, with a sampling grid of 50 m. The APPLEx-2 only covered two focus area because of standing water in two focus areas.	
APPLEx-2	Oct. 12–Nov.12 2022			

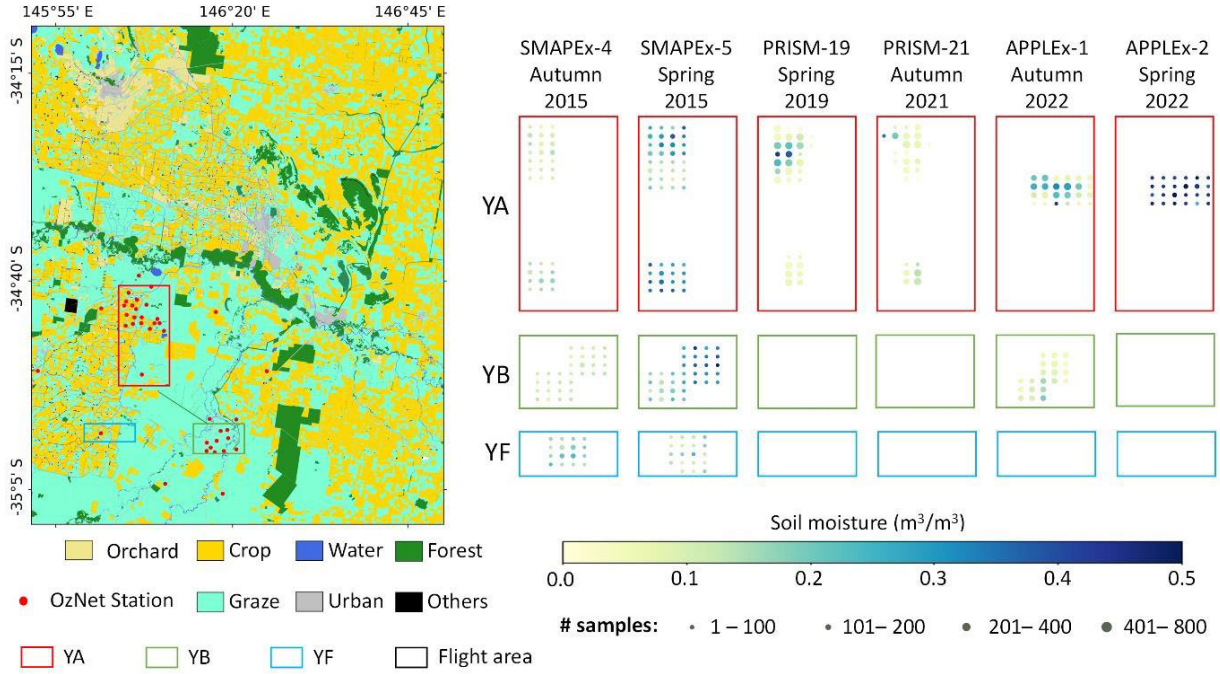

Fig. 2. Flight area of SMAPEx-4 and -5 (left panel) and the soil moisture measurements collected in the six airborne campaigns made in the Yanco agriculture area (right panel) since the launch of Sentinel-1. Each dot shows the average soil moisture of a 1-km grid cell, with the size being the number of HDAS measurements made within the grid cell.

1) Dynamic Predictors: Sentinel-1 interferometric wide (IW) ground range detected (GRD) VV and VH polarizations served as the primary dynamic predictors, as their values are directly related to soil permittivity and, consequently, soil moisture. All GRD acquisitions archived in Google Earth Engine (GEE) were utilized. The GEE Sentinel-1 GRD dataset had been preprocessed to a 10-m pixel size (sigma naught in dB) using a standard SAR processing workflow, which included thermal noise removal, radiometric calibration, terrain correction, and conversion to the logarithmic scale. The GEE provided easy access to data for each 1-km grid cell through a simple aggregation process. Specifically, pixels overlapping within a 1-km grid cell were averaged, with the overlapping areas used as weights. Accordingly, the speckle noise was substantially reduced in this step, being equivalent to using a multilook of ~200 × 50. In addition, the local incidence angle (LIA) and day of year (DOY) of each acquisition were included as predictor attributes. The LIA helps the models learn the angular dependence of VV and VH polarizations [11], while DOY allows the models to account for seasonal changes that impact soil moisture due to phenological and meteorological shifts.

To represent the effect of vegetation, the NDVI from the Terra and Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) was used, specifically the MYD13Q1 and MOD13Q1 products (V6). These NDVI products have an identical spatial resolution of 250 m and a temporal resolution of 16 days. The NDVI for the Sentinel-1 acquisition dates was first interpolated from the eight-day composites of the two MODIS products using a spline function, with this interpolated NDVI, and then resampled to the 1-km grid cells using the same aggregation method described earlier. The temperature

at 2 m and total precipitation were extracted from the ERA5-Land products, providing hourly data with a grid resolution of 0.1°. Given the complexity of how precipitation and temperature affect soil moisture [35], the seven-day average values preceding the DOY were used to represent the average meteorological conditions for simplicity. In addition, surface soil moisture from the SMAP Level-3 enhanced passive soil moisture product Version 5 [36] was included as reliable prior knowledge; although having a resolution of 27 km, this was applied at the sampling interval of 9 km. The SMAP mission, launched by NASA in 2015, uses a combination of L-band radar (active) and radiometer (passive) technologies to provide global measurements of soil moisture and freeze-thaw states. Since SMAP soil moisture has a revisit of 2–3 days, it was temporally interpolated to match the Sentinel-1 acquisitions using a spline function. Similarly, SMAP soil moisture, ERA5 temperature, and precipitation were resampled to the 1-km grid cells using the same aggregation method.

2) Static Predictors: The static variables included terrain features, soil texture, location, and land cover. Three terrain features were derived from the Shuttle Radar Topography Mission Digital Elevation Data Version 3 (SRTM DEM V3), i.e., elevation, slope, and aspect. They not only change the real incidence angle of radar observations but also substantially affect soil moisture variations. The SRTM DEM has a spatial resolution of 30 m, so the three terrain features were initially calculated at this 30-m resolution and then aggregated to the 1-km grid. Soil textures were extracted from the global gridded soil information products (SoilGrids 250 m) [37]. Three key soil variables, being sand, clay, and bulk density for the top 5-cm layer, were selected and aggregated from 250 m to 1 km as they reflect soil porosity, water holding capacity, and dielectric constant relationships. In addition, location information, including latitude and longitude, was included to enable the models to leverage spatial analysis and contextual understanding [38].

The inclusion of land cover types allows the model to learn land cover-specific characteristics. However, instead of relying on existing sophisticated land cover products, this study utilized surface reflectance data from Sentinel-2 directly, as the land cover information can be automatically learned by the model [38]. The multispectral instrument aboard the twin Sentinel-2 satellites has 13 spectral bands, ranging from visible to shortwave infrared, with a spatial and temporal resolution of 10–60 m and five days, respectively. To minimize the impact of clouds on the optical data, the median values of one-year of Sentinel-2 acquisitions preceding the Sentinel-1 acquisition DOYs were used. This included six spectral reflectance bands: B2 (490 nm), B3 (560 nm), B4 (665 nm), B8 (842 nm), B11 (1610 nm), and B12 (2190 nm). The reflectance data were also aggregated to the 1-km grid.

All the variables were normalized to a range of 0–1 using the min–max normalization method, with the maximum and minimum values listed in Table II. Notably, aspect, longitude, and DOY are cyclic variables, meaning that the end of one cycle seamlessly transitions to the start of another. Consequently, simple linear scaling or normalization does not adequately capture this cyclical relationship. Therefore, sine and cosine functions were applied to map these variables into a 2-D space, preserving their cyclical properties.

TABLE II
25 INPUT VARIABLES OF THE FOUR MACHINE LEARNING METHODS AND THE MAXIMUM AND MINIMUM VALUES USED IN DATA NORMALIZATION

Variable name	Min.	Max.	#Variables
VV [dB]	-30	5	1
VH [dB]	-35	0	1
Incidence angle [°]	29.1	46	1
Sentinel-1 DoY*	1	365	2
Sand [%]	0	100	1
Clay [%]	0	100	1
Bulk density [g/cm^3]	1	1.8	1
Elevation [m]	0	5500	1
Slope [°]	0	40	1
Aspect [°]*	0	360	2
Latitude	0	90	1
Longitude*	0	180	2
NDVI	-0.2	1	1
Temperature [°C]	-10	-35	1
Precipitation [m/d]	0	0.3	1
Reflectance at 490 nm, 560 nm, 665 nm, 842 nm, 1610, and 2190 nm	0	1	6
Soil moisture at 9 km [m^3/m^3]	0	0.6	1

*: mapped into a circular representation in a two-dimensional space.

III. METHOD

A. Machine Learning Methods

In this study, four machine learning approaches were evaluated (i.e., MLP, SVR, RF, and XGBoost), representing a broad range of conventional machine learning techniques commonly used in the remote sensing of soil moisture. They are well described in the textbook or literature [39], so only a brief introduction is provided here, with a focus on their hyperparameters. The MLP is inspired by the neural networks of the human brain. It consists of interconnected neurons, organized into layers: an input layer, one or more hidden layers, and an output layer. Each neuron processes input signals through weighted connections and is followed by an activation function to produce a nonlinear output, which then serves as the input for the next layer. MLPs have demonstrated strong capabilities in recognizing patterns and making predictions from large datasets. In this study, a naïve MLP model was used, with five main hyperparameters considered: the number of hidden layers and neurons in each layer, the learning rate, the activation function, and the strength of the L2 regularization term. The commonly used values for these hyperparameters are listed in Table III.

SVR is a type of support vector machine (SVM) used for regression tasks. Its goal is to find a function that deviates from the actual observed values by no more than a specified margin while maintaining the flattest possible function. Three key hyperparameters are typically required in SVR: the kernel type, epsilon, and the C parameter. The epsilon parameter defines a margin of tolerance within which no penalty is

TABLE III
HYPERPARAMETERS OF MACHINE LEARNING MODELS, U(A, B) MEANS THAT THE VALUE WAS RANDOMLY SELECTED FROM A UNIFORM DISTRIBUTION RANGING FROM A TO B

Model	Hyperparameter	Candidates
MLP	# hidden layers	1,2,3,4,5
	# neurons in each layer	16,32,64,128,256
	Activation	'relu', 'tanh'
	Strength of the L2 regularization term (α)	0.0001, 0.001, 0.01
	Learning rate	0.001, 0.01, 0.1
SVR	*C*	0.1, 0.3, 1, 3, 10
	Epsilon	0.01, 0.03, 0.1, 0.2, 0.3, 0.5
	Kernel type	'linear', 'rbf', 'poly', 'sigmoid'
RF	# trees	U(10, 100)
	Maximum depth	'Auto', 10, 20, 30
	# minimum samples for node split	2
	# minimum samples for leaf	1, 2, 4, 8
	# maximum features	'sqrt', 'log2', # features (25)
XGBoost	# trees	U(100, 300)
	Maximum depth	3, 4, 5
	Learning rate	0.3, 0.1, 0.03, 0.01, 0.003, 0.001

assigned to errors, allowing the model some flexibility. The *C* parameter is crucial for balancing the tradeoff between minimizing the error on the training data and ensuring that the model generalizes well to new data by maintaining smoothness. The candidate values for these hyperparameters are also listed in Table III.

The RF algorithm is an ensemble method widely used for classification, regression, and anomaly detection. It operates by building a few decision trees during training and then averaging the predictions of the individual trees as the output. The four key hyperparameters for RF describe the structure and numbers of trees. In addition, bootstrap sampling was employed in this study to increase the diversity of trees, which enhances the robustness of the model. The candidate values for these hyperparameters are provided in Table III. In particular, the "auto" setting for maximum depth means that all nodes are expanded until all leaves are pure.

XGBoost is a powerful and efficient implementation of the gradient boosting framework. Similar to the RF algorithm, XGBoost is an ensemble of trees but built sequentially. Each new tree is trained to correct the errors made by the previous trees, thereby gradually improving the overall model performance. This sequential training process often allows XGBoost to outperform RF in many applications, with higher efficiency. The key hyperparameters for XGBoost include the number of trees, the maximum depth of the trees, and the learning rate (see Table III). These hyperparameters are crucial for balancing model complexity and training speed, ensuring optimal performance.

B. Evaluation Scenarios

The performance of the four methods was evaluated in four scenarios with few training samples, being dependent validation, temporal validation, spatial validation, and region validation in the Yanco area. In the dependent validation, a progressive training approach was used, incrementally increasing the training set from 0.5% to 5% of the whole sample. This is a considerably small training size compared to most existing studies, where tenfold or fivefold cross validation was made [20], [21]. The remaining samples were used to assess the models' performance. This process enables to evaluate the models' effectiveness with varying training sizes, assuming sample availability from the research area. Since both the training and testing samples were randomly selected from the same dataset, the distributions of the training and testing sets were generally well-matched, making this the simplest scenario for model validation. The dependent scenario evaluation provides insights into the models' effectiveness under ideal conditions with no domain shift.

In the temporal validation scenario, the effect of the temporal size of the training set on model accuracy was examined. Grid cells with more than 21 samples from 2021 were selected, as well as samples from 2016 to 2020, resulting in 528 grid cells. Similar to the progressive training approach used in the dependent scenario, the number of training samples from each of the 528 grid cells was incrementally increased from 1 to 21 in steps of 2, with all training samples taken from 2021. The 114 793 samples from 2016 to 2020 were used as the testing set, ensuring temporal independence from the training set. This approach aimed to provide insights into whether a limited number of samples from a short time period could support accurate future data retrieval or aid in the reconstruction of historical periods. Notably, 11 training samples from each of the 528 grid cells account for ∼5% of the testing set, being equivalent to the 5% of the dependent scenario. This scenario introduced a temporal gap between the training and testing data, making it a more challenging task compared to the dependent scenario.

In the spatial validation scenario, the impact of the spatial distribution of training samples on model performance was assessed. The experiment started by randomly selecting one of the 1007 grid cells from the global dataset as the initial training set and progressively increased the number of selected grid cells to 57 in steps of 8. The samples from the remaining grid cells were used as the testing set. Given the substantial variation in climate, landscape, soil, and vegetation characteristics across different grid cells, models trained on a limited number of grid cells are less likely to capture the complex relationships between soil moisture and other surface characteristics. This scenario is thus the most challenging but provides insights into the accuracy of global products over areas without ground measurements.

The first three validation scenarios offer comprehensive evaluations from different perspectives. However, real-world applications often focus on a small region with a sparse in situ network. Accordingly, the fourth scenario used the OzNet stations in the Yanco area as the training set, while the in situ measurements collected during six airborne campaigns (see Table I) served as the testing set. In addition, airborne-derived soil moisture from L-band passive microwave data at 1-km resolution was used for intercomparison, with a focus on the spatial pattern of soil moisture maps.

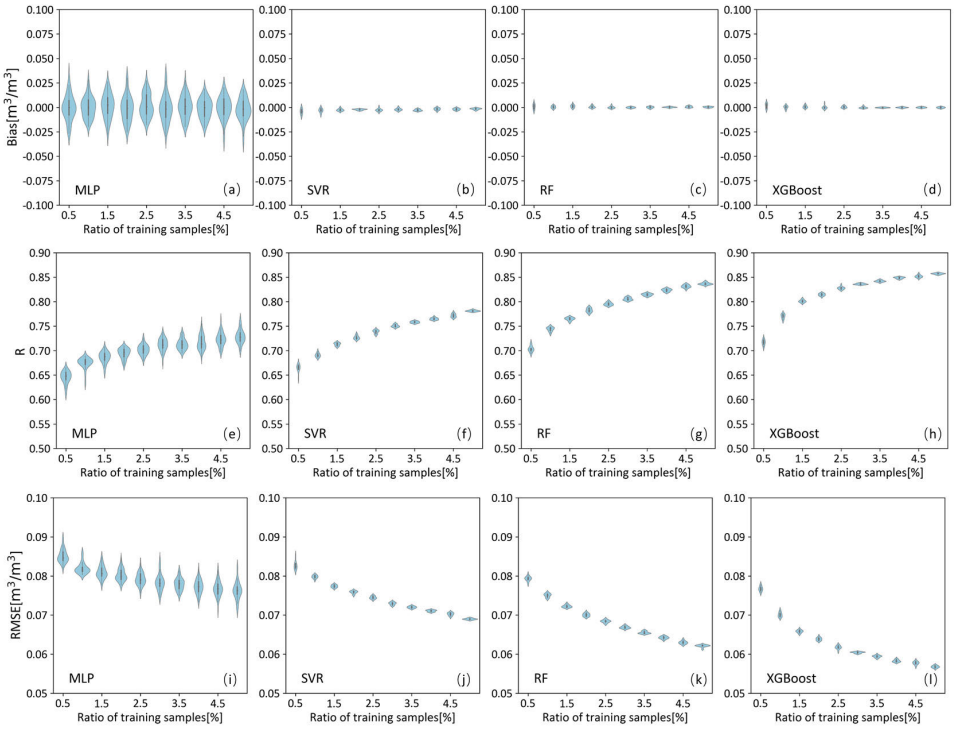

Fig. 3. Performance of MLP, SVR, RF, and XGBoost (left to right) in the dependent scenario with varying small numbers of training samples. The 0.5% in the x-axis means ∼1000 random selected training samples. The violin plot of each case shows the statistics of the 50 random implementations.

C. Hyperparameters Tuning and Accuracy Metrics

The selection of hyperparameters is crucial to the performance of all four methods. While various search methods can be used, ensuring a fair comparison of the four methods in scenarios with limited training samples is challenging because search methods based solely on the training set can introduce large uncertainty in the comparison. To address this issue, a unique search scheme was proposed in this study. For each machine learning method and evaluation subscenario, 100 models were trained on the training set using 100 random combinations of hyperparameters and then evaluated using the testing set. The models and the corresponding hyperparameter combinations with the lowest root-mean-square error (RMSE) were kept. The evaluation accuracy is thus an approximation of the accuracy one could expect when employing the best possible hyperparameter tuning.

The selection of training samples or grid cells substantially impacts model performance and comparison. To account for this, all subscenarios were repeated 50 times using randomly selected training sets. Since hyperparameters are also influenced by the training set, optimal hyperparameters were obtained for each of the 50 implementations. This iterative approach ensured a more robust evaluation of the models across different training set configurations. Three accuracy metrics were used: bias, Pearson correlation coefficient (R), and RMSE.

IV. Results

A. Evaluation in the Dependent Scenario

Fig. 3 illustrates the performance of the four methods in the dependent scenario. The violin plots for MLP show that as the ratio of training samples increased, the spread of bias slightly narrowed, but the variations remained relatively large and stable. This suggests that the fluctuations caused by the different compositions of training samples did not substantially diminish with more training data. In contrast, the violin plots for SVR, RF, and XGBoost were narrowly centered around 0 across all training sample ratios, indicating that the biases of these three methods were relatively insensitive to both the size and composition of the random training samples. This makes SVR, RF, and XGBoost more stable in terms of providing unbiased predictions.

The XGBoost outperformed the other methods in terms of R and RMSE, with a minimal spread of all three metrics as the training sample ratio increased. The median RMSE reached the radar accuracy target of 0.06 m^3/m^3 [40] when the training ratio was >3%. Although not fully investigated, the XGBoost performance is expected to further improve with more training samples beyond 5%. The RF model closely followed XGBoost, demonstrating a consistent dependence on the size of the training sets, though with slightly more variability. The SVR and MLP methods also benefited significantly ($p < 0.01$) from an incremental increase of 0.5% in training samples, but their improvements were smaller than for the RF and XGBoost counterparts. Moreover, the median RMSE for the SVR and MLP methods started from higher initial values, being 0.082 and 0.085 m^3/m^3, respectively. Both methods required a much larger training set (>5%) to reach the requirement of 0.06 m^3/m^3. Similar to the patterns observed in bias, the violin plots for MLP showed the widest spread among all the models, indicating much higher variability in performance due to different compositions of training samples. This variability is likely due to the more complex architecture

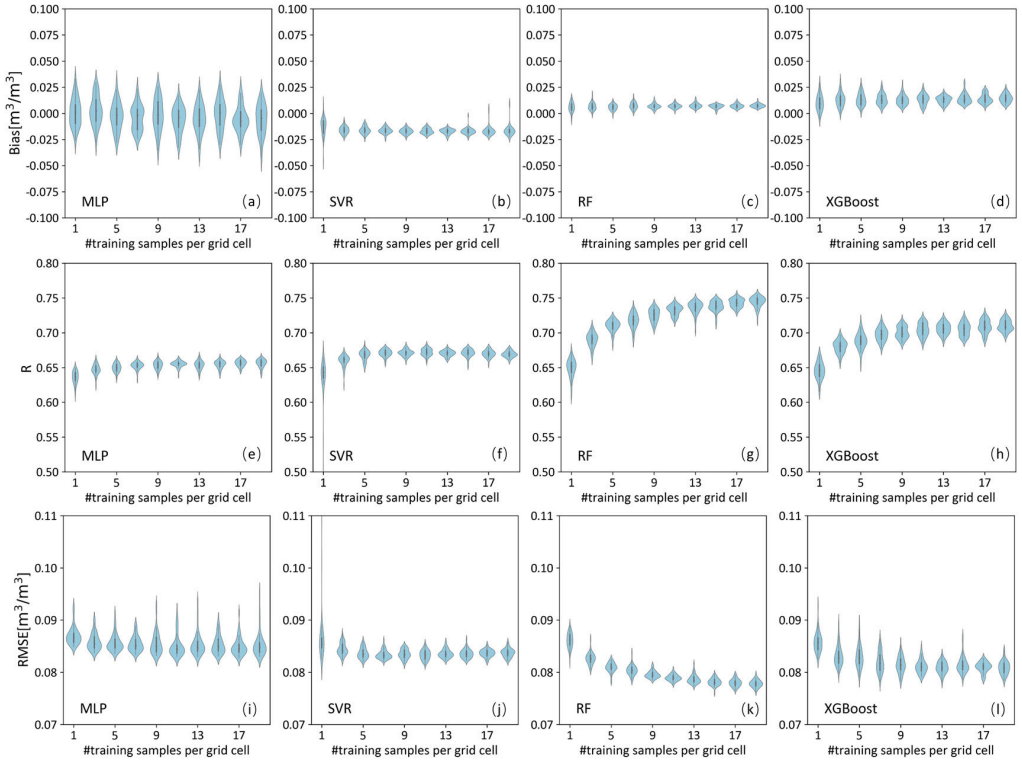

Fig. 4. Performance of MLP, SVR, RF, and XGBoost (left to right) in the temporal validation scenario with varying small numbers of training samples per grid cell. One in the x-axis means 528 training samples from 528 grid cells in the calendar year of 2021, with the size being ~0.47% of the testing size. The testing samples were from the same grid cells but an independent period of 2016–2020.

of MLP, which, although simpler than deep learning models, still requires more data and fine-tuning to achieve stable predictions.

Fig. 4 illustrates the performance of the four machine learning methods in the temporal validation scenario. The MLP model exhibited consistent variation across different training sizes, with median biases remaining close to zero. Most random implementations showed a bias between −0.025 and 0.025 m^3/m^3, being consistent with the results observed in the dependent scenario (see Fig. 3). In contrast, the other three methods had relatively larger variations in bias compared to the dependent scenario. Despite these differences, a consistently small bias was maintained with marginal variation as the number of training samples increased. The SVR method had a median bias of around −0.016 m^3/m^3, while the RF and XGBoost methods achieved small positive median biases of approximately 0.007 and 0.013 m^3/m^3, respectively. These results suggest that while temporal validation introduces slightly more variability, all methods still performed with low bias as the training set size increased.

The performance of all four methods improved in terms of R and RMSE as more training samples were used, but the benefits were much smaller than those observed in the dependent scenario. The improvements in median RMSE ranged from 0.002 to 0.008 m^3/m^3 as the number of training samples per grid cell increased from 1 to 21. Notably, most of the improvements came from increasing the number of grid-specific samples from 1 to 5. As expected, the performance of all methods in the temporal validation scenario was worse than in the dependent scenario, despite the training size being nearly twice as large. The RF model achieved the best results, with a median RMSE of 0.086–0.078 m^3/m^3 and R values between 0.652 and 0.746, followed by the XGBoost and SVR methods. The MLP method performed the worst, with a median RMSE of 0.087–0.085 m^3/m^3. None of the methods met the target of 0.06 m^3/m^3 in this scenario. These results suggest that the empirical relationships learned by the models from 2021 generalized poorly to the years 2016–2020. Moreover, additional training samples from 2021 contributed little to improving performance. Though the variability caused by different combinations of training samples was larger than in the dependent scenario, it remained marginal relative to the absolute values. This aligns with the observation that adding more training samples provides minimal improvement.

B. Evaluation in the Spatial Validation Scenario

Fig. 5 shows the performance of the models in the spatial validation scenario. The median biases for all methods remained close to 0 m^3/m^3 across all subscenarios; however, the variations in bias caused by different combinations of training grid cells were substantially larger than in the other two scenarios, especially when only 1–9 training grid cells were used. This indicates substantial overfitting for all models when limited stations were available, though this issue was somewhat mitigated by increasing the number of training grid cells. Similar patterns were observed in the R and RMSE results. The variance of the MLP and SVR models decreased more rapidly as the number of training grid cells increased, suggesting that they were more sensitive to the training data's

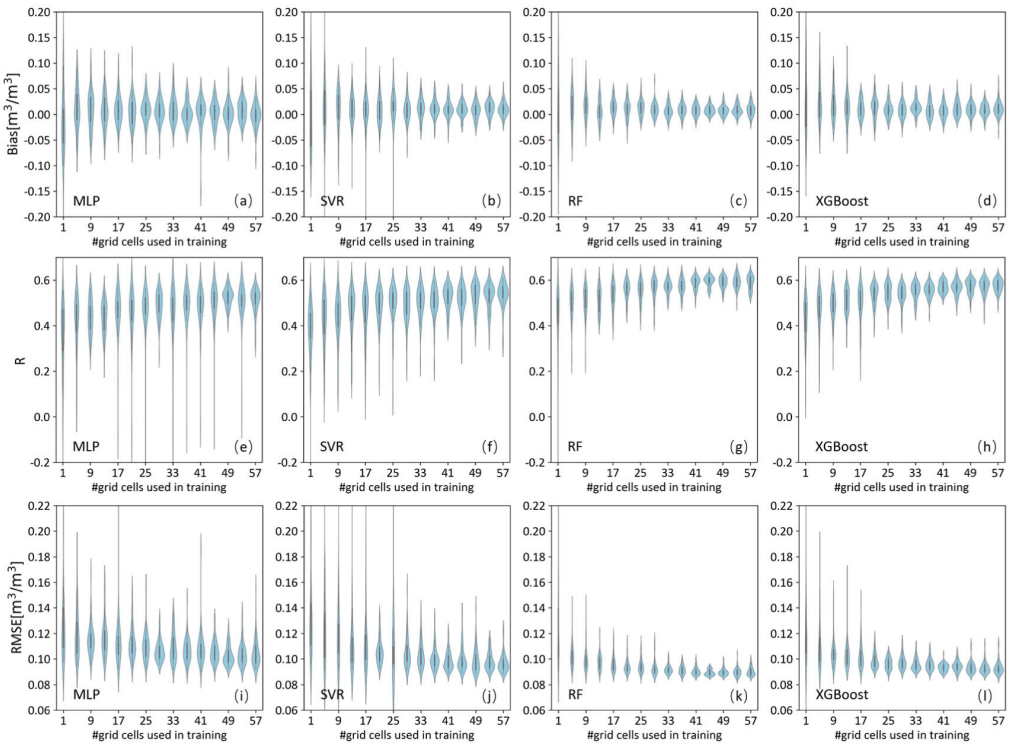

Fig. 5. Performance of MLP, SVR, RF, and XGBoost (left to right) in the spatial validation scenario with increasing numbers of grid cells used in training.

spatial distribution. The RF and XGBoost models maintain a consistent advantage in terms of lower variance, indicating that they were more robust in handling spatial variability in the dataset, offering more reliable performance with fewer grid cells.

All methods achieved significantly worse performance ($p < 0.01$ in the t-test) compared to their counterparts in the dependent and temporal validation scenarios. The RF and XGBoost methods outperformed the SVR and MLP methods ($p < 0.01$), but their median RMSEs were as large as 0.089 and 0.092 m^3/m^3, respectively, when trained with 57 grid cells (∼5% of the total grid cells). Furthermore, the improvement of adding four extra training grid cells diminished, with the benefit in RMSE dropping from ∼0.080 to just 0.001 m^3/m^3 as the number of training grid cells increased. These results suggest that none of the methods were able to meet the target RMSE of 0.06 m^3/m^3 in scenarios with limited training stations. Moreover, even with a larger number of training grid cells, the models struggled to achieve satisfactory generalization, indicating that the methods are unlikely to perform well over areas with a considerable number of training stations. This highlights the challenge of achieving robust soil moisture predictions in spatially heterogeneous environments with limited training data.

C. Evaluation in the Regional Scenario

Fig. 6 shows the performance of the four methods in the focus areas of the six airborne campaigns made in the Yanco agriculture area of Australia. Consistent with the results from the previous three scenarios, the RF and XGBoost methods outperformed the MLP and SVR methods. However, the RMSE for the RF and XGBoost methods was still high, at 0.112 and 0.113 m^3/m^3, being nearly double the target RMSE of 0.06 m^3/m^3. The MLP and SVR methods had a comparable RMSE of 0.117 and 0.121 m^3/m^3, respectively, though the SVR methods yielded a much lower R value of 0.156. All the machine learning models exhibited a tendency to regress toward the mean, overestimating low soil moisture values and underestimating high values. While this scenario differed from the dependent, temporal, and spatial validation scenarios, it most closely resembled the spatial validation scenario, as most testing samples came from different grid cells of the OzNet stations. These results further demonstrated the challenge of achieving accurate predictions in spatially diverse areas with limited training data.

Fig. 7 provides an intercomparison of the soil moisture maps retrieved by the four methods as well as the airborne passive soil moisture. The soil moisture was relatively low in the early stage of SMAPEx-4 (the first two weeks of May 2015) but increased abruptly after a rainfall event on May 16, 2015. This was well captured by airborne passive soil moisture maps, with the high areas of high soil moisture content on May 17, 2015, being areas of intensive rainfall that occurred before the airborne observation. The dry conditions observed on May 5, 2015 were generally captured by all four machine learning methods, though they exhibited an average overestimation of approximately 0.040–0.057 m^3/m^3. The Sentinel-1 acquisition on May 17, 2015 only covered half of the research area. In this case, all machine learning methods underestimated soil moisture and failed to capture two prominent wet patches in the southwest quadrant. Moreover, clear boundaries of 9 km grids were observed in the soil moisture maps produced by the MLP and SVR methods, suggesting that the SMAP soil moisture data applied on the 9-km grid had a negative and

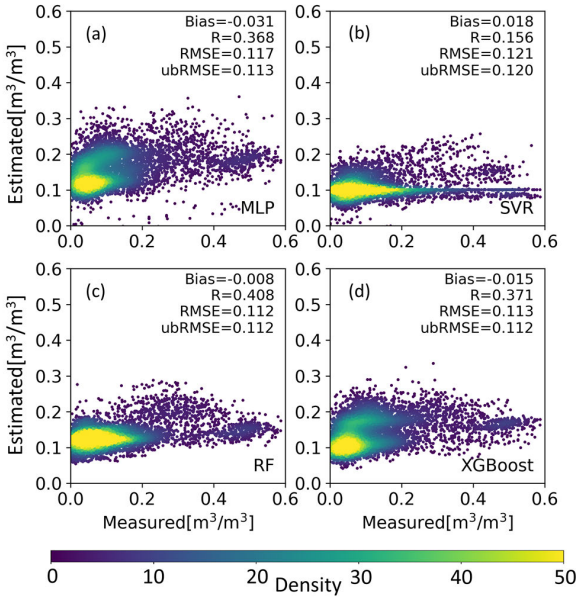

Fig. 6. Measured versus estimated soil moisture for the four machine learning methods on the regional dataset. The training samples were taken from the OzNet in situ stations, while the testing samples in this figure were the intensive ground soil moisture measurements collected in the six airborne campaigns (see Fig. 2).

overly dominant influence on these methods. In contrast, the RF and XGBoost methods produced similar soil moisture patterns, with soil moisture values in forested areas being higher than in other land-use types (see Fig. 2).

At the start of SMAPEx-5, soil moisture levels exceeded 0.4 m^3/m^3 across the entire flight area due to heavy rainfall in the first week of September 2015, followed by a drying process over the next three weeks. The drying rate varied spatially, creating substantial heterogeneity by September 15, 2015. The retrieved average soil moisture values from the MLP, SVR, RF, and XGBoost methods on that date were 0.203, 0.110, 0.177, and 0.164 m^3/m^3, respectively—substantially lower than the airborne soil moisture average of 0.273 m^3/m^3. Moreover, the spatial heterogeneity was largely underestimated by all the methods, with a 9-km grid cell pattern clearly observed in the SVR method result. By September 26, 2015, soil moisture had generally decreased to below 0.1 m^3/m^3, except for areas with irrigated farms, which maintained higher moisture levels. While the retrieved soil moisture from all four methods reflected the drying trend, their values were consistently higher than those from the airborne observations, especially for the MLP method. Apart from the SVR method, the spatial patterns of retrieved soil moisture from the other three methods showed little change between September 15 and 26, continuing to overestimate moisture in forested areas.

V. Discussion

This study evaluated the capability of the MLP, SVR, RF, and XGBoost methods in estimating soil moisture, with a focus on scenarios with limited training data. The RF and XGBoost methods consistently outperformed the SVR and MLP methods across all scenarios, aligning with findings from previous studies, e.g., [41]. Moreover, the machine learning models performed best in the dependent validation scenario and worst in the spatial validation scenario.

The accuracy target of 0.06 m^3/m^3 was only met by the RF and XGBoost methods in the dependent scenarios, despite the use of optimal hyperparameters selected through the testing set. Given that real-world applications are typically more challenging than the dependent scenario, such as the regional evaluation conducted in the Yanco agricultural area, conventional machine learning methods are unlikely to provide reliable soil moisture estimates in regions with sparse data. Similarly, existing global soil moisture products based on machine learning may contain substantial uncertainties in data-scarce areas, requiring further evaluation and refinement.

Many existing studies have reported more favorable estimation errors, typically in the range of 0.02–0.06 m^3/m^3 [41], [42], [43]. This discrepancy can partly be attributed to the difference in training sample size, as fewer training samples mean that the model is less likely to represent unseen data accurately. While the RF and XGBoost methods performed well, even the SVR and MLP methods are expected to meet the 0.06-m^3/m^3 target if more than 50% of the samples were used in the dependent scenario (see Fig. 3). However, the benefit of additional training samples was substantially diminished in the temporal and spatial validation scenarios (see Figs. 4 and 5). In the temporal validation scenario, additional training samples were from the same domain and did not improve the model's generalization to unseen periods. Similarly, adding a few extra training stations in the spatial validation scenario did not improve the performance for unseen sites. The success of existing applications of machine learning methods may thus be overoptimistic in view of producing a global high-resolution product, where generalization across diverse, unseen regions and periods is critical.

The varying ways of presenting results are another source for explaining the difference in findings between this study and others. It is challenging to compare R and coefficient of determination, or RMSE and mean absolute error, with many studies reporting averages of accuracy metrics across multiple stations [15]. Moreover, all machine learning methods are highly sensitive to hyperparameters, and using improper hyperparameters can lead to weak baselines. In this study, optimal hyperparameters were achieved by using the testing data, meaning that the reported accuracy represents the best possible outcome. However, this level of performance is unlikely to be achieved in real-world applications.

Reporting biases, along with the use of weak baselines, have also contributed to confusion in the evaluation and comparison of methods, hindering the advancement and application of machine learning techniques for soil moisture estimation. As deep learning is playing an increasingly important role in the remote sensing of soil moisture, there is an urgent need for a standardized evaluation framework. This should include common datasets, consistent training/validation/testing scenarios, metrics, and baseline algorithms to ensure comparability across studies. However, preparing a standard dataset for soil moisture is much more challenging than for other remote sensing applications (e.g., landcover classification) due to the larger temporal and spatial variation of soil moisture

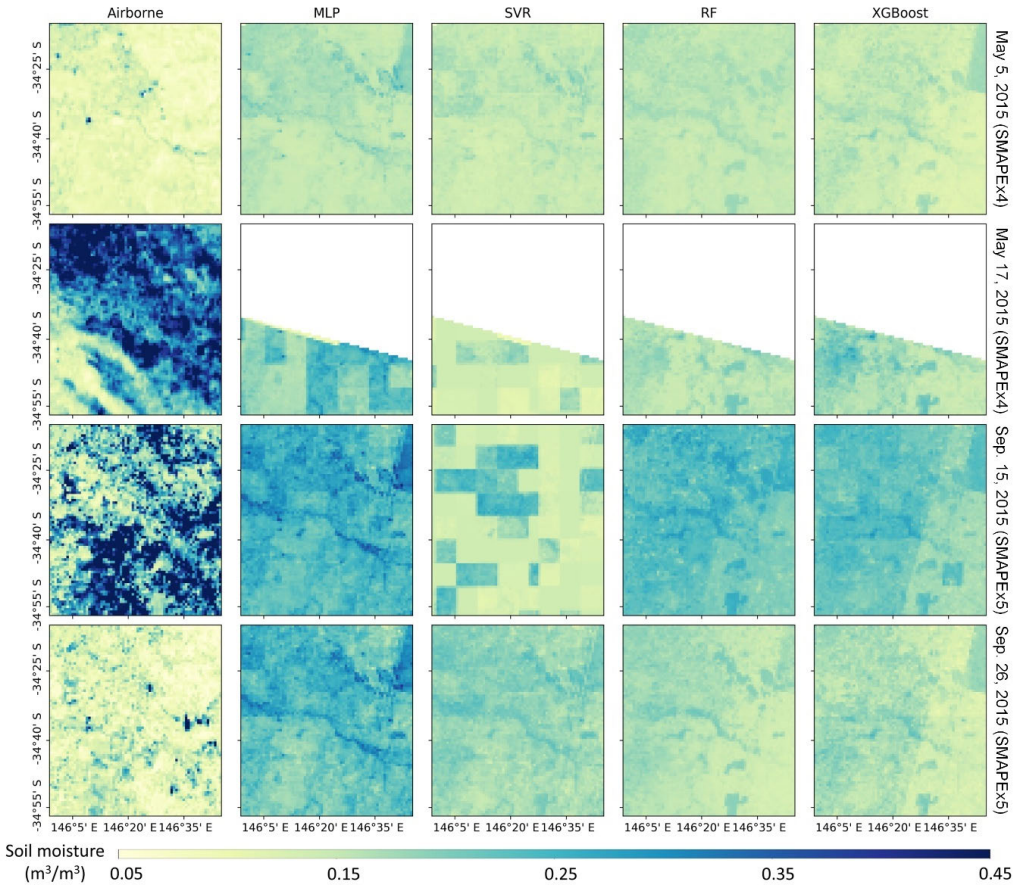

Fig. 7. Comparison of the soil moisture maps retrieved from the airborne passive microwave data against those from the MLP, SVR, RF, and XGBoost methods (left to right). Sentinel-1 is not available in the blank areas of the second row.

and the higher cost for ground truth data collection. In this study, all ISMN stations were used to build a large dataset for global evaluation of daily averaged soil moisture at a 1-km resolution. Most ISMN stations are sparsely distributed in space making it challenging to prepare the truth at specific grid sizes. While the spatial representativeness error can be an acceptable value (e.g., $\sim$0.02 m^3/m^3 in Southern Italy [13]), treating point measurements as the truth of a 1-km grid cell can introduce large uncertainties. A further evaluation of the four machine learning methods using input variables prepared at 50 m showed a marginal difference compared to evaluations made at 1 km (see Fig. 8 in Appendix). Although soil moisture at 50 m can still have large discrepancies with in situ measurements, the small differences observed between 50-m and 1-km resolutions suggest that spatial representativeness errors should result in unbiased overall accuracy statistics for global evaluation. In other words, the performance may be overestimated in some periods or areas and underestimated in others but balanced out to provide a reliable global assessment.

The main motivation of this study was to evaluate machine learning methods for producing a global soil moisture product. Given the limited spatial coverage of descending and ascending Sentinel-1 orbits, a merged daily product was more practical for the SAR data. Therefore, daily averaged in situ soil moisture values were used instead of values collected near the Sentinel-1 acquisition time, introducing a limitation of the prepared dataset. Fortunately, the daily average soil moisture closely matched the values collected near the Sentinel-1 acquisition times. Less than 1% of the samples exhibited differences exceeding the sensors' general calibration accuracy of 0.04 m^3/m^3. Moreover, the overall root-mean-square difference between the daily averages and the values near the acquisition time was 0.008 m^2/m^2, which is relatively small when compared to the target RMSE of 0.06 m^2/m^2.

The use of all available ISMN stations raises another concern; some stations are located in densely vegetated areas where C-band SAR may be insensitive to or negatively correlated with surface soil moisture. However, soil moisture can still be estimated with acceptable accuracy in such situations using machine learning methods, as auxiliary data (e.g., NDVI) can play a dominant role with the backscatter contributing indirectly. While further investigations into the importance of specific input variables are possible, the results can vary over time and space, and such analyses are beyond the scope of this study.

The regional validation scenario is an example of real applications that focus on a small area with a low-density in situ network. Unfortunately, none of the methods in this study were able to meet the target accuracy of 0.06 m^3/m^3 for the research area. Moreover, the spatial patterns of the retrieved soil moisture maps deviated a lot from those captured

TABLE IV
NUMBER OF SAMPLES FROM THE 25 SOIL MOISTURE NETWORKS USED IN THE GLOBAL EVALUATIONS

Network	#Stations	#Samples	Country	Network	#Stations	#Samples	Country
HOAL [51]	3	1342	Austria	Ru_CFR	2	67	Russia
AMMA-CATCH [52]	6	547	Benin, Niger, Mali	REMEDHUS[61]	20	13506	Spain
RISMA [53]	13	2016	Canada	SMN-SDR[62, 63]	34	1714	Sudan
NGARI, NAQU, MAQU [54]	39	2531	China	WSMN[64]	2	30	UK
CTP_SMTMN [55]	1	44	China	SNOTEL[65]	372	81992	USA
TAHMO	2	288	Gulf of Guinea	iRON[66]	9	1172	USA
HOBE [56]	28	9372	Denmark	SOILSCAPE[67]	18	1811	USA
FR_Aqui [57]	2	214	France	ARM[68]	14	3479	USA
SMOSMANIA [58]	19	13926	France	PBO_H2O[69]	141	6490	USA
TERENO [59]	5	2645	Germany	USCRN[70]	91	18459	USA
HYDROL-NET_PERUGIA	1	63	Italy	SCAN[71]	161	32284	USA
BIEBRZA_S-1 [60]	2	421	Poland	FLUXNET-AMERIFLUX	4	849	USA
RSMN	18	12564	Romania				

by the airborne-derived soil moisture from passive microwave observations (see Fig. 7). In many applications, the spatial pattern can be just as critical as the absolute values [44], [45]. However, direct evaluation of spatial patterns is challenging, as such ground truth of an area is generally not available. Consequently, many studies resort to comparing their results as a time series at individual stations or with other soil moisture products at a coarse grid resolution [46], [47], typically being tens of kilometers.

The existing airborne-derived soil moisture maps from passive microwave observations [32], [48], [49] present one of the few reliable sources for evaluating spatial patterns at higher resolutions ($\leq$1 km). The mismatched spatial patterns in this study suggest that incorporating spatial constraints or spatial dependencies, which can be extracted by advanced machine learning models, such as the CNN or transformer methods, is essential for improving accuracy. In addition, the role of coarse input variables, such as the SMAP soil moisture, must be handled carefully, as they can introduce erroneous spatial discontinuities, especially at the boundaries of the coarse pixels (see Fig. 7).

Further improvements can be achieved by utilizing unlabeled data and incorporating prior knowledge, considering the poor performance and reduced contribution of adding additional training samples in the spatial and temporal validation scenarios tested here (see Figs. 4 and 5). Fortunately, there are a great number of semi-supervised methods and transfer learning algorithms [50] that offer promising solutions for enhancing soil moisture estimation in data-scarce regions. For example, a recent study demonstrated that pretrained models at a 9-km resolution based on SMAP soil moisture products can be successfully transferred to a 1-km resolution using fewer in situ measurements [18]. In the same temporal evaluation scenario, this cross-resolution transfer approach achieved a median RMSE of 0.068 m^3/m^3 when 11 training samples per grid cell were used, significantly ($p < 0.01$) outperforming the RF method, which had a median RMSE of 0.079 m^3/m^3. Leveraging knowledge from hydrologic models and/or land surface models could also facilitate the use of deep learning techniques, as these models can learn most trainable parameters without the need for extensive in situ measurements. Such approaches are expected to enable the development of more sophisticated machine learning models that can be better generalized across regions and conditions where the ground data are limited.

VI. CONCLUSION

This study provides a comprehensive evaluation of four machine learning methods utilizing SAR remote sensing observations for soil moisture mapping. The study used a global dataset of $\sim$200 000 samples alongside a regional dataset, being the first evaluation focused on scenarios with scarce samples. The main findings include: 1) conventional machine learning methods can meet the target accuracy of 0.06 m^3/m^3 in a dependent scenario using few training samples; 2) all of the methods failed to estimate the soil moisture at an acceptable accuracy in the spatial and temporal validation scenarios, with the RMSE $>$0.09 m^3/m^3 in all subscenarios; 3) more

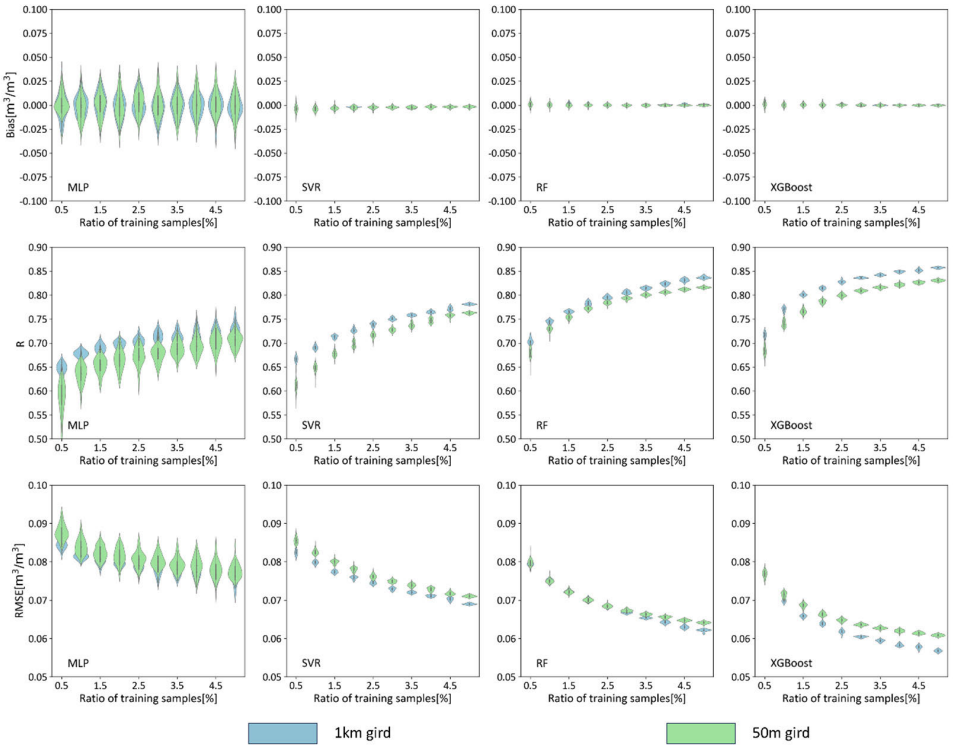

Fig. 8. Same as Fig. 3 but for both 1-km and 50-m grids.

training samples or training stations can substantially improve the accuracy in a dependent scenario but were not critical for the other scenarios; and 4) evaluation on the regional dataset showed consistently poorer results, with the RMSE being >0.1 m^3/m^3. These findings resulted in the conclusion that expectations for SAR remote sensing of soil moisture using machine learning are currently overoptimistic. More robust methods for data-scare scenarios and a comparable evaluation scheme without reporting bias and weak baselines are urgently required to reach the target of producing the first global soil moisture product from SAR.

APPENDIX

See Table IV and Fig. 8.

REFERENCES

[1] A. Manoj J, R. K. Guntu, and A. Agarwal, "Spatiotemporal dependence of soil moisture and precipitation over India," *J. Hydrol.*, vol. 610, Jul. 2022, Art. no. 127898.

[2] V. Prakash and V. Mishra, "Soil moisture and streamflow data assimilation for streamflow prediction in the narmada river basin," *J. Hydrometeorol.*, vol. 24, no. 8, pp. 1377–1392, Aug. 2023.

[3] M. Cao, M. Chen, J. Liu, and Y. Liu, "Assessing the performance of satellite soil moisture on agricultural drought monitoring in the north China plain," *Agricult. Water Manage.*, vol. 263, Apr. 2022, Art. no. 107450.

[4] Y. Zhang, J. P. Walker, and V. R. N. Pauwels, "Assimilation of wheat and soil states for improved yield prediction: The APSIM-EnKF framework," *Agricult. Syst.*, vol. 201, Aug. 2022, Art. no. 103456.

[5] K. A. McColl, S. H. Alemohammad, R. Akbar, A. G. Konings, S. Yueh, and D. Entekhabi, "The global distribution and dynamics of surface soil moisture," *Nature Geosci.*, vol. 10, no. 2, pp. 100–104, Feb. 2017.

[6] M. Rodell et al., "The global land data assimilation system," *Bull. Amer. Meteorol. Soc.*, vol. 85, no. 3, pp. 381–394, Mar. 2004.

[7] H. Lievens et al., "Joint Sentinel-1 and SMAP data assimilation to improve soil moisture estimates," *Geophys. Res. Lett.*, vol. 2017, pp. 6145–6153, Dec. 2017.

[8] J. Peng et al., "A roadmap for high-resolution satellite soil moisture applications—Confronting product characteristics with user requirements," *Remote Sens. Environ.*, vol. 252, Jan. 2021, Art. no. 112162.

[9] K. C. Kornelsen and P. Coulibaly, "Advances in soil moisture retrieval from synthetic aperture radar and hydrological applications," *J. Hydrol.*, vol. 476, pp. 460–489, Jan. 2013.

[10] H. Lievens et al., "Effective roughness modelling as a tool for soil moisture retrieval from C- and L-band SAR," *Hydrol. Earth Syst. Sci.*, vol. 15, no. 1, pp. 151–162, Jan. 2011.

[11] L. Zhu, J. P. Walker, L. Tsang, H. Huang, N. Ye, and C. Rüdiger, "Soil moisture retrieval from time series multi-angular radar data using a dry down constraint," *Remote Sens. Environ.*, vol. 231, Sep. 2019, Art. no. 111237, doi: 10.1016/j.rse.2019.111237.

[12] L. Zhu, R. Si, X. Shen, and J. P. Walker, "An advanced change detection method for time-series soil moisture retrieval from Sentinel-1," *Remote Sens. Environ.*, vol. 279, Sep. 2022, Art. no. 113137.

[13] A. Balenzano et al., "Sentinel-1 soil moisture at 1 km resolution: A validation study," *Remote Sens. Environ.*, vol. 263, Sep. 2021, Art. no. 112554.

[14] B. Bauer-Marschallinger et al., "Toward global soil moisture monitoring with Sentinel-1: Harnessing assets and overcoming obstacles," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 520–539, Jan. 2019.

[15] L. Zhu, S. Yuan, Y. Liu, C. Chen, and J. P. Walker, "Time series soil moisture retrieval from SAR data: Multi-temporal constraints and a global validation," *Remote Sens. Environ.*, vol. 287, Mar. 2023, Art. no. 113466.

[16] S.-B. Kim et al., "Surface soil moisture retrieval using the L-band synthetic aperture radar onboard the soil moisture active—Passive satellite and evaluation at core validation sites," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 1897–1914, Apr. 2017.

[17] L. Zhu, J. P. Walker, and X. Shen, "Stochastic ensemble methods for multi-SAR-mission soil moisture retrieval," *Remote Sens. Environ.*, vol. 251, Dec. 2020, Art. no. 112099.

[18] L. Zhu, J. Dai, Y. Liu, S. Yuan, T. Qin, and J. P. Walker, "A cross-resolution transfer learning approach for soil moisture retrieval from Sentinel-1 using limited training samples," *Remote Sens. Environ.*, vol. 301, Feb. 2024, Art. no. 113944.

[19] S. Ahmad, A. Kalra, and H. Stephen, "Estimating soil moisture using remote sensing data: A machine learning approach," *Adv. Water Resour.*, vol. 33, no. 1, pp. 69–80, Jan. 2010.

[20] J. Stamenkovic, P. Ferrazzoli, L. Guerriero, D. Tuia, and J.-P. Thiran, "Joining a discrete radiative transfer model and a kernel retrieval algorithm for soil moisture estimation from SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3463–3475, Jul. 2015.

[21] J. Wang, F. Wu, J. Shang, Q. Zhou, I. Ahmad, and G. Zhou, "Saline soil moisture mapping using Sentinel-1A synthetic aperture radar data and machine learning algorithms in humid region of China's east coast," *CATENA*, vol. 213, Jun. 2022, Art. no. 106189.

[22] C. Zheng, L. Jia, and T. Zhao, "A 21-year dataset (2000–2020) of gap-free global daily surface soil moisture at 1-km grid resolution," *Sci. Data*, vol. 10, no. 1, p. 139, Mar. 2023.

[23] Y. Zhang et al., "Generation of global 1-km daily soil moisture product from 2000 to 2020 using ensemble learning," *Earth Syst. Sci. Data Discuss.*, vol. 2023, pp. 1–37, Jan. 2023.

[24] T. T. Nguyen et al., "A low-cost approach for soil moisture prediction using multi-sensor data and machine learning algorithm," *Sci. Total Environ.*, vol. 833, Aug. 2022, Art. no. 155066.

[25] M. Karamouz, R. S. Alipour, M. Roohinia, and M. Fereshtehpour, "A remote sensing driven soil moisture estimator: Uncertain downscaling with geostatistically based use of ancillary data," *Water Resour. Res.*, vol. 58, no. 10, p. 2022, Oct. 2022.

[26] S. Datta, P. Das, D. Dutta, and R. K. Giri, "Estimation of surface moisture content using Sentinel-1 C-band SAR data through machine learning models," *J. Indian Soc. Remote Sens.*, vol. 49, no. 4, pp. 887–896, Apr. 2021.

[27] K. Fang, M. Pan, and C. Shen, "The value of SMAP for long-term soil moisture estimation with the help of deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2221–2233, Apr. 2019.

[28] A. Yinglan, G. Wang, P. Hu, X. Lai, B. Xue, and Q. Fang, "Root-zone soil moisture estimation based on remote sensing data and deep learning," *Environ. Res.*, vol. 212, Sep. 2022, Art. no. 113278.

[29] N. Madhukumar, E. Wang, Y. Everingham, and W. Xiang, "Hybrid transformer network for soil moisture estimation in precision irrigation," *IEEE Access*, vol. 12, pp. 48898–48909, 2024.

[30] W. Zheng et al., "GRU—Transformer: A novel hybrid model for predicting soil moisture content in root zones," *Agronomy*, vol. 14, no. 3, p. 432, Feb. 2024.

[31] W. Dorigo et al., "The international soil moisture network: Serving Earth system science for over a decade," *Hydrol. Earth Syst. Sci. Discuss.*, vol. 25, no. 11, pp. 5749–5804, Nov. 2021.

[32] N. Ye et al., "The soil moisture active passive experiments: Validation of the SMAP products in Australia," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 2922–2939, Apr. 2021, doi: 10.1109/TGRS.2020.3007371.

[33] M. S. Yee, J. P. Walker, C. Rüdiger, R. M. Parinussa, T. Koike, and Y. H. Kerr, "A comparison of SMOS and AMSR2 soil moisture using representative sites of the OzNet monitoring network," *Remote Sens. Environ.*, vol. 195, pp. 297–312, Jun. 2017.

[34] O. Merlin, J. Walker, R. Panciera, R. Young, J. Kalma, and E. Kim, "Calibration of a soil moisture sensor in heterogeneous terrain," presented at the Int. Congr. Model. Simul. (MODSIM), 2007.

[35] W. Zhao and Z.-L. Li, "Sensitivity study of soil moisture on the temporal evolution of surface temperature over bare surfaces," *Int. J. Remote Sens.*, vol. 34, nos. 9–10, pp. 3314–3331, May 2013.

[36] P. O'Neill et al., "SMAP enhanced L3 radiometer global and polar grid daily 9 km ease-grid soil moisture version 5," NASA Nat. Snow Ice Data Center, Distrib. Active Arch. Center, Boulder, CO, USA, 2021.

[37] L. Poggio et al., "SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty," *Soil*, vol. 7, no. 1, pp. 217–240, Jun. 2021.

[38] L. Zhu, G. I. Webb, M. Yebra, G. Scortechini, L. Miller, and F. Petitjean, "Live fuel moisture content estimation from MODIS: A deep learning approach," *ISPRS J. Photogramm. Remote Sens.*, vol. 179, pp. 81–91, Sep. 2021.

[39] A. Rani, N. Kumar, J. Kumar, and N. K. Sinha, "Machine learning for soil moisture assessment," in *Deep Learning for Sustainable Agriculture*. Amsterdam, The Netherlands: Elsevier, 2022, pp. 143–168.

[40] P. Lal et al., "A multi-scale algorithm for the NISAR mission high-resolution soil moisture product," *Remote Sens. Environ.*, vol. 295, Sep. 2023, Art. no. 113667.

[41] M. Li and Y. Yan, "Comparative analysis of machine-learning models for soil moisture estimation using high-resolution remote-sensing data," *Land*, vol. 13, no. 8, p. 1331, Aug. 2024.

[42] J. Liu, F. Rahmani, K. Lawson, and C. Shen, "A multiscale deep learning model for soil moisture integrating satellite and in situ data," *Geophys. Res. Lett.*, vol. 49, no. 7, Apr. 2022, Art. no. e2021GL096847.

[43] V. Batchu, G. Nearing, and V. Gulshan, "A deep learning data fusion model using Sentinel-1/2, SoilGrids, SMAP, and GLDAS for soil moisture retrieval," *J. Hydrometeorol.*, vol. 24, no. 10, pp. 1789–1823, Oct. 2023.

[44] A. W. Western, R. B. Grayson, and G. Blöschl, "Scaling of soil moisture: A hydrologic perspective," *Annu. Rev. Earth Planet. Sci.*, vol. 30, no. 1, pp. 149–180, May 2002.

[45] H. Vereecken, J. A. Huisman, H. Bogena, J. Vanderborght, J. A. Vrugt, and J. W. Hopmans, "On the value of soil moisture measurements in vadose zone hydrology: A review," *Water Resour. Res.*, vol. 44, no. 4, pp. 1–21, Apr. 2008.

[46] A. Al-Yaari et al., "Assessment and inter-comparison of recently developed/reprocessed microwave satellite soil moisture products using ISMN ground-based measurements," *Remote Sens. Environ.*, vol. 224, pp. 289–303, Apr. 2019.

[47] L. Brocca et al., "Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe," *Remote Sens. Environ.*, vol. 115, no. 12, pp. 3390–3408, Dec. 2011.

[48] R. Panciera et al., "The soil moisture active passive experiments (SMAPEx): Toward soil moisture retrieval from the SMAP mission," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 490–507, Jan. 2014.

[49] H. McNairn et al., "The soil moisture active passive validation experiment 2012 (SMAPVEX12): Prelaunch calibration and validation of the SMAP soil moisture algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2784–2801, May 2015.

[50] Y. Ma, S. Chen, S. Ermon, and D. B. Lobell, "Transfer learning in environmental remote sensing," *Remote Sens. Environ.*, vol. 301, Feb. 2024, Art. no. 113924, doi: 10.1016/j.rse.2023.113924.

[51] G. Blöschl et al., "The hydrological open air laboratory (HOAL) in petzenkirchen: A hypothesis-driven observatory," *Hydrol. Earth Syst. Sci.*, vol. 20, no. 1, pp. 227–255, Jan. 2016, doi: 10.5194/hess-20-227-2016.

[52] T. Lebel et al., "AMMA-CATCH studies in the sahelian region of west-Africa: An overview," *J. Hydrol.*, vol. 375, no. 1, pp. 3–13, Aug. 2009, doi: 10.1016/j.jhydrol.2009.03.020.

[53] E. R. Ojo, P. R. Bullock, J. L'Heureux, J. Powers, H. McNairn, and A. Pacheco, "Calibration and evaluation of a frequency domain reflectometry sensor for real-time soil moisture monitoring," *Vadose Zone J.*, vol. 14, no. 3, pp. 1–12, Mar. 2015, doi: 10.2136/vzj2014.08.0114.

[54] Z. Su et al., "The Tibetan Plateau observatory of Plateau scale soil moisture and soil temperature (tibet-obs) for quantifying uncertainties in coarse resolution satellite and model products," *Hydrol. Earth Syst. Sci.*, vol. 15, no. 7, pp. 2303–2316, Jul. 2011, doi: 10.5194/hess-15-2303-2011.

[55] K. Yang et al., "A multiscale soil moisture and freeze—Thaw monitoring network on the third pole," *Bull. Amer. Meteorological Soc.*, vol. 94, no. 12, pp. 1907–1916, Dec. 2013, doi: 10.1175/bams-d-12-00203.1.

[56] K. H. Jensen and J. C. Refsgaard, "HOBE: The Danish hydrological observatory," *Vadose Zone J.*, vol. 17, no. 1, Jan. 2018, Art. no. 180059, doi: 10.2136/vzj2018.03.0059.

[57] J.-P. Wigneron et al., "The aqui network: Soil moisture sites in the 'les landes' forest and graves vineyards (bordeaux aquitaine region, France)," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 3739–3742, doi: 10.1109/IGARSS.2018.8517392.

[58] J.-C. Calvet, N. Fritz, C. Berne, B. Piguet, W. Maurel, and C. Meurey, "Deriving pedotransfer functions for soil quartz fraction in southern France from reverse modeling," *SOIL*, vol. 2, no. 4, pp. 615–629, Dec. 2016, doi: 10.5194/soil-2-615-2016.

[59] H. R. Bogena et al., "The TERENO-Rur hydrological observatory: A multiscale multi-compartment research platform for the advancement of hydrological science," *Vadose Zone J.*, vol. 17, no. 1, Jan. 2018, Art. no. 180055, doi: 10.2136/vzj2018.03.0055.

[60] K. Dabrowska-Zielinska et al., "Soil moisture in the biebrza wetlands retrieved from Sentinel-1 imagery," *Remote Sens.*, vol. 10, no. 12, p. 1979, Dec. 2018.

[61] Á. González-Zamora, N. Sánchez, M. Pablos, and J. Martínez-Fernández, "CCI soil moisture assessment with SMOS soil moisture and in situ data under different environmental conditions and spatial scales in Spain," *Remote Sens. Environ.*, vol. 225, pp. 469–482, May 2019, doi: 10.1016/j.rse.2018.02.010.

[62] T. Zhao et al., "Soil moisture experiment in the Luan river supporting new satellite mission opportunities," *Remote Sens. Environ.*, vol. 240, Apr. 2020, Art. no. 111680, doi: 10.1016/j.rse.2020.111680.

[63] J. Zheng et al., "Assessment of 24 soil moisture datasets using a new in situ network in the Shandian River Basin of China," *Remote Sens. Environ.*, vol. 271, 2022, Art. no. 112891, doi: 10.1016/j.rse.2022.112891.

[64] G. Petropoulos and J. McCalmont, "An operational in situ soil moisture & soil temperature monitoring network for west wales, U.K.: The WSMN network," *Sensors*, vol. 17, no. 7, p. 1481, Jun. 2017, doi: 10.3390/s17071481.

[65] G. Leavesley et al., "A modeling framework for improved agricultural water-supply forecasting," in *Proc. AGU Fall Meeting Abstr.*, vol. 1, 2008, p. 0497.

[66] E. C. Osenga, J. A. Vano, and J. C. Arnott, "A community-supported weather and soil moisture monitoring database of the roaring fork catchment of the Colorado river headwaters," *Hydrological Processes*, vol. 35, no. 3, p. 14081, Mar. 2021, doi: 10.1002/hyp.14081.

[67] D. I. Shuman et al., "Measurement scheduling for soil moisture sensing: From physical models to optimal control," *Proc. IEEE*, vol. 98, no. 11, pp. 1918–1933, Nov. 2010, doi: 10.1109/JPROC.2010.2052532.

[68] D. Cook. (2024). *Surface Energy Balance System (SEBS) Instrument Handbook*. ARM Climate Res. Facility, Pacific Northwest Nat. Lab., Richland, WA, USA. [Online]. Available: https://www.osti.gov/biblio/1004944

[69] K. M. Larson, E. E. Small, E. D. Gutmann, A. L. Bilich, J. J. Braun, and V. U. Zavorotny, "Use of GPS receivers as a soil moisture network for water cycle studies," *Geophys. Res. Lett.*, vol. 35, no. 24, pp. 1–5, Dec. 2008, doi: 10.1029/2008gl036013.

[70] J. E. Bell et al., "U.S. climate reference network soil moisture and temperature observations," *J. Hydrometeorol.*, vol. 14, no. 3, pp. 977–988, 2013, doi: 10.1175/JHM-D-12-0146.1.

[71] G. L. Schaefer, M. H. Cosh, and T. J. Jackson, "The USDA natural resources conservation service soil climate analysis network (SCAN)," *J. Atmos. Ocean. Technol.*, vol. 24, no. 12, pp. 2073–2077, Dec. 2007, doi: 10.1175/2007JTECHA930.1.

Liujun Zhu (Member, IEEE) received the B.S. degree in geography from Zhejiang Normal University, Jinhua, China, in 2012, the M.Sc. degree in geography information science from Nanjing University, Nanjing, China, in 2015, and the Ph.D. degree in civil engineering from Monash University, Melbourne, VIC, Australia, in 2019, with a thesis on multi-SAR soil moisture retrieval.

From 2017 to 2018, he was a Visiting Scholar with the University of Michigan, Ann Arbor, MI, USA. From 2019 to 2020, he continued his research on radar soil moisture as a Research Fellow with Monash University, where he has held an adjunct position since 2020. In 2021, he was appointed as an Associate Professor with the Yangtze Institute for Conservation and Development, Hohai University, Nanjing, China, where he continues to advance his research. His research interests include microwave remote sensing of soil moisture, machine learning, and its applications in soil moisture studies.

Junjie Dai received the B.S. degree in water engineering from Hohai University, Nanjing, China, in 2022, where he is currently pursuing the master's degree in hydrology and water resources.

His research interests have focused on active microwave remote sensing, machine learning, and their applications in soil moisture retrieval.

Junliang Jin received the bachelor's degree in hydrology and water resources engineering from Zhejiang University, Hangzhou, China, in 2003, and the master's and Ph.D. degrees in hydrology and water resources from Hohai University, Nanjing, China, in 2006 and 2010, respectively.

From 2010 to 2012, he was a Post-Doctoral Researcher with Nanjing Hydraulic Research Institute (NHRI), MWR, Nanjing. From 2010 to 2024, he was a Senior Engineer and a Professor Engineer with NHRI. Since February 2024, he has been a Professor with Hohai University. His research interests include hydrology processes, climate change, remote sensing, soil moisture, evapotranspiration, and hydrological simulation.

Shanshui Yuan received the B.E. degree in hydrology and water resources engineering and the M.E. degree in hydrology from Hohai University, Nanjing, China, in 2009 and 2012, respectively, and the Ph.D. degree in geography from Texas A&M University, College Station, TX, USA, in 2016.

From 2017 to 2020, he worked as a Post-Doctoral Research Associate with The Ohio State University, Columbus, OH, USA. In 2021, he was appointed as a Professor with the Yangtze Institute for Conservation and Development, Hohai University. His research focuses on hydrometeorology, land–atmosphere interactions, and extreme drought/flood events.

Ziwei Xiong received the B.E. (civil) degree (Hons.) from Monash University, Melbourne, VIC, Australia, in 2022, where he is currently pursuing the Ph.D. degree in civil engineering.

In 2023, he was a Visiting Ph.D. Student with the University of Southern California, Los Angeles, CA, USA. His current research interests include active microwave remote sensing and root-zone soil moisture retrieval.

Jeffrey P. Walker (Fellow, IEEE) received the B.E. (civil) (Hons. 1) degree from the University of Newcastle, Callaghan, NSW, Australia, in 1995, and the B.Surveying (University Medal) degree and the Ph.D. degree in water resources engineering from the University of Newcastle in 1995 and 1999, respectively. His Ph.D. thesis was among the early pioneering research on estimation of root-zone soil moisture from assimilation of remotely sensed surface soil moisture observations.

After his Ph.D. degree, he joined the NASA Goddard Space Flight Center, Greenbelt, MD, USA, to implement his soil moisture work globally. In 2001, he moved to the Department of Civil and Environmental Engineering, The University of Melbourne, Melbourne, VIC, Australia, as a Lecturer, where he continued his soil moisture work, including the development of the only Australian airborne capability for simulating new satellite missions for soil moisture. In 2010, he was appointed as a Professor with the Department of Civil Engineering and Environmental Engineering, Monash University, Melbourne, where he has continued this research. He is contributing to soil moisture satellite missions at NASA, ESA, and JAXA, as a Science Team Member of the Soil Moisture Active Passive (SMAP) mission and a Cal/Val Team Member of the Soil Moisture and Ocean Salinity (SMOS) and Global Change Observation Mission–Water (GCOM-W), respectively.

Dr. Walker is a Laureate Fellow of the Australian Research Council.