

A Comparative Analysis of Remote Sensing Soil Moisture Datasets Fusion Methods: Novel LSTM Approach Versus Widely Used Triple Collocation Technique

Haojin Zhao , Carsten Montzka , *Senior Member, IEEE*, Harry Vereecken , and Harrie-Jan Hendricks Franssen 

Abstract—Microwave remote sensing technology has emerged to provide valuable products to monitor and assess soil moisture content at regional or global scales. However, each soil moisture product exhibits different advantages and shortcomings. Data fusion could help improve accuracy by merging information from different sources. In this research, a traditional triple collocation (TC) based method and a novel long short term memory network (LSTM) are used to merge soil moisture products from the soil moisture active passive mission, Advanced Microwave Scanning Radiometer 2 (AMSR2), and The Advanced SCATterometer for a study area located in western Europe. This research reveals that the LSTM outperforms the traditional TC based method for data fusion. The study identifies that both climate forcing and physiographic attributes significantly influence the spatial and temporal variations observed in the LSTM merging scheme. Consequently, the study highlights the considerable potential of the LSTM method for large-scale integration of remote sensing soil moisture data.

Index Terms—Data fusion, long short term memory model, remote sensing, satellites, soil moisture, triple collocation.

I. INTRODUCTION

SOIL moisture is a fundamental parameter with broad implications for agricultural production, hydrological processes, and climate studies [1], [2]. Precise and timely estimation of soil moisture at regional and global scales is essential for effective management of water resources [3]. Here, microwave remote sensing offers an effective way to retrieve soil moisture information due to its sensitivity to the dielectric properties of the soil [4], [5].

Microwave remote sensing operates at different frequencies, namely L-band (1–2 GHz), C-band (4–8 GHz), and X-band (8–12 GHz), and has been widely employed in soil moisture retrievals. Each frequency exhibiting unique interactions with the soil surface and subsurface [6]. Satellites based on X-band and C-band include Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E) [7], and the successor Advanced Microwave Scanning Radiometer 2 (AMSR2) [8], and the Advanced Scatterometer (ASCAT) from Meteorological Operational platform series (Metop-A, Metop-B, and Metop-C) [9], [10]. Over the past decade, L-band has been commonly considered favorable for large-scale soil moisture monitoring with its reduced sensitivity to vegetation cover and atmosphere interference. In addition, it has a greater signal penetration depth into the soil profile, allowing for a more comprehensive assessment of soil moisture conditions [11], [12], [13]. Satellites operating at L-bands include the soil moisture and ocean salinity (SMOS) [14] and the soil moisture active passive (SMAP) [15].

The remote sensing soil moisture products derived from multiple instruments exhibit notable disparities, not only caused by different microwave frequencies but also from orbits, polarization methods, and retrieval algorithms [16], [17], [18]. These discrepancies emphasize each soil moisture product's distinct strengths and limitations and potentially result in inconsistent findings. Therefore, there is a need to combine remote sensing data to reduce uncertainties from any individual datasets and provide a more accurate estimation of soil moisture.

The arithmetic mean is the common method to merge information from multiple satellites, and its effectiveness has been demonstrated in various applications [19], [20], [21]. Other more sophisticated linear weighting methods could assign the weights to each dataset based on their prediction errors,

Received 28 October 2023; revised 14 March 2024 and 24 May 2024; accepted 3 September 2024. Date of publication 6 September 2024; date of current version 25 September 2024. The work of Haojin Zhao was supported in part by China Scholarship Council under Grant 201806010358, in part by the Dragon 5 cooperation between the European Space Agency (ESA), and in part by the Ministry of Science and Technology of the China under Grant 59316. The work of C. Montzka was supported by the German Ministry of Economic Affairs and Climate (BMWK) through the German Aerospace Center for the AssimEO project under Grant 50EE1914A. The work of Harrie-Jan Hendricks Franssen was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—under Grant SFB 1502/1-2022–450058266. (Corresponding author: Haojin Zhao.)

Haojin Zhao is with the Institute of Bio- and Geosciences, Agrosphere, Forschungszentrum Jülich, 52425 Jülich, Germany, also with the Centre for High-Performance Scientific Computing in Terrestrial Systems, HPSC TerrSys, 52425 Jülich, Germany, and also with the Faculty of Georesources and Materials Engineering, RWTH Aachen University, 52062 Aachen, Germany (e-mail: h.zhao@fz-juelich.de).

Carsten Montzka is with the Institute of Bio- and Geosciences, Agrosphere, Forschungszentrum Jülich, 52425 Jülich, Germany (e-mail: c.montzka@fz-juelich.de).

Harry Vereecken and Harrie-Jan Hendricks Franssen are with the Institute of Bio- and Geosciences, Agrosphere, Forschungszentrum Jülich, 52425 Jülich, Germany, and also with the Centre for High-Performance Scientific Computing in Terrestrial Systems, HPSC TerrSys, 52425 Jülich, Germany (e-mail: h.vereecken@fz-juelich.de; h.hendricks-franssen@fz-juelich.de).

Digital Object Identifier 10.1109/JSTARS.2024.3455549

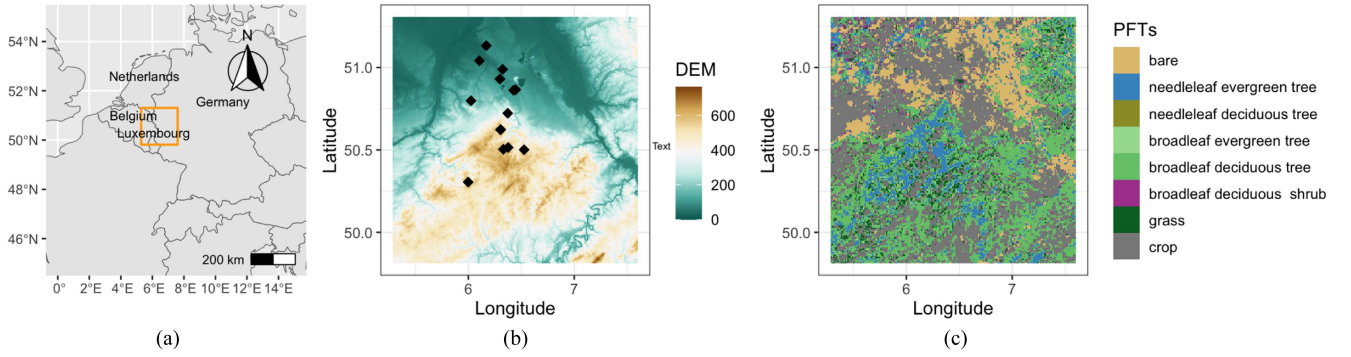


Fig. 1. (a) Location of the research area in Europe, (b) digital elevation model (DEM: m) and (c) plant functional types (PFTs) over the research area. The diamonds indicate the locations of the CRNS sites.

commonly evaluated against reference datasets. Kim et al. [22], [23] developed a method to merge two satellite soil moisture products by maximizing their correlation with reference datasets from ERA-Interim (the European Centre for Medium-Range Weather Forecasts Reanalysis-Interim) and MERRA-Land (Modern-Era Retrospective Analysis for Research and Applications Land) and showed the improvement performance for the merged dataset when comparing with in situ observations. However, they found that the reference dataset can impact the fusion process as the set of weights is governed by the correlation among the parent datasets and the used reference. Triple collocation (TC) is a mathematical technique to assess the error statistics of three or more products without relying on ground truth measurements [24], [25], [26]. For example, Yilmaz et al. [27] applied the TC method to merge modeled and remote sensing soil moisture data, and subsequently, the method gained widely usage [28], [29], [30], [31], [32], [33], [34], [35]. In general, the linear weighting methods is easy to understand, straightforward to use, and can be applied to various datasets.

Machine learning (ML) technique has recently been implemented to merge remote sensing information from various instruments because of its remarkable learning abilities. The ML algorithms are capable of handling large and noisy datasets and extracting spatial and temporal features from multiple data sources, making it a valuable approach to enhance the data fusion accuracy and efficiency [36], [37]. The success of ML in soil moisture fusion has been demonstrated in previous studies [37], [38], [39], [40], including increased correlations and decreased ubRMSE.

This study focuses on merging soil moisture retrievals obtained from three distinct satellite sources using the traditional triple collocation (TC) method and the long short-term memory (LSTM) technique. The primary aim is to assess the possibility of LSTM as a data fusion approach and compare its performance against the TC-based merging scheme. In addition, we aim to explore the potential benefits of the LSTM merging scheme in terms of temporal and spatial soil moisture estimation. Furthermore, we investigate the impact of different parameter setups in the LSTM scheme across various scenarios to gain insights into the respective contributions of these parameters to the soil moisture merging process.

II. STUDY AREA AND DATA

A. Study Area

The study area is situated in western Europe, encompassing regions within Germany, Belgium, Luxembourg, and the Netherlands and covering an area of $300 \text{ km} \times 300 \text{ km}$ (see Fig. 1). It consists of a diverse range of land-use types and distinct geographical features, including main crops in the flat terrain and deciduous and coniferous forests in the upland regions [41]. The study area exhibits a temperate marine climate. The annual precipitation patterns vary across the region and are influenced by terrain elevation, with the northern part experiencing an average precipitation of 600–900 mm and the southern part receiving higher amounts, reaching up to 1600 mm annually [42].

B. Remote Sensing Soil Moisture Products

In this study, three different soil moisture products are employed for the data fusion scheme, including the SMAP Enhanced Level 3 Radiometer Soil Moisture (SPL3SMP_E), Version 5 [43], the AMSR2/GCOM-W1 surface soil moisture L3 product [44], and ASCAT Surface Soil Moisture Climate Data Record 7 Extension 12.5 km sampling (H119) [45]. In addition, the multimission passive soil moisture product Version 6.1 from European Space Agency Climate Change Initiative (ESA CCI) [46], [47], [48] serves as the reference dataset to evaluate the merged soil moisture.

The SMAP L-band SPL3SMP_E product, first released on December 2016, provides a global daily composite soil moisture estimate at a spatial resolution of 9 km using the Equal-Area Scalable Earth Grid (EASE-Grid) [49]. It is generated with the Backus-Gilbert optimal interpolation technique, which extracts information from antenna temperatures and converts them to brightness temperatures [49]. The final soil moisture estimates are then derived through the Dual Channel Algorithm (DCA).

AMSR2 is deployed on the Global Change Observation Mission - Water 1 (GCOM-W1) satellite, using the Land Parameter Retrieval Model (LPRM) and providing daily measurements of soil moisture at 25 km resolution [50]. The AMSR2 data archive consists of two distinct products: ascending (daytime) and descending (nighttime) acquisitions. It offers two available

bands, namely C-band and X-band, for soil moisture measurements [51]. Both the descending and ascending C-band (6.93 GHz) soil moisture products from AMSR2 are utilized in the analysis.

The ASCAT H119 product is generated by data from the Advanced Scatterometer (ASCAT) instrument on board the MetOp-A, MetOp-B, and MetOp-C satellites [52]. ASCAT measures the backscattered radar signals from the Earth's surface, which are utilized to estimate soil moisture. The H119 product provides long-term soil moisture monitoring capabilities, with data available starting from January 2007 [10]. It has a sampling resolution of 12.5 km on the Discrete Global Grid (DGG) and a temporal resolution of approximately 1-3 days [10]. The soil moisture information derived from ASCAT is expressed in the degree of saturation (%), which is subsequently converted to volumetric soil moisture content (cm^3/cm^3) by multiplication with the soil porosity data retrieved from the Global Land Data Assimilation System (also within the ASCAT datasets). Only measurements with a soil moisture noise designation below 20% are retained for analysis.

The ESA CCI SM v 6.1 Passive product is retrieved using the Land Parameter Retrieval Model (LPRA) [53] algorithm. This particular product is derived from Level2 data obtained from various satellite missions, including SMAP, SMOS, AMSR2, AMSR-E, among others. The merged passive product provides a global coverage of soil moisture at the spatial resolution of 0.25° and temporal resolution of 1 day. Note that the ESA CCI dataset is mainly used as a reference for visual comparison due to its widespread use in hydrological studies [29], [54], [55], [56].

It is worth noting that both SMAP and AMSR2 are passive sensors. However, the soil moisture products are largely influenced by other factors, e.g., the microwave frequencies, orbits, retrieval, and ancillary data such as soil texture, land cover classes. Several studies [30], [57], [58], [59] have validated the SMAP and AMSR2 soil moisture data at regional and global scales and revealed that they have distinct soil moisture trends and spatial patterns. Moreover, Montzka et al. [60] validated the SMAP, AMSR2 over the same research region using both standard validation methods and the TC approach and revealed structural quality differences between SMAP and AMSR2 retrievals. Consequently, we assume that the SMAP and AMSR2 are independent products in the subsequent research.

Given the different spatial resolutions of these remote sensing soil moisture products, it is necessary to rescale them prior to the application of merging schemes. In this study, we rescale the datasets to a $500 \text{ m} \times 500 \text{ m}$ resolution using the nearest neighbor method.

C. In Situ Soil Moisture Observations

Thirteen cosmic-ray neutron sensing (CRNS) stations [61] from the Terrestrial Environmental Observations (TERENO) network [62] is used for measuring soil moisture at field scale. Fig. 1 shows the locations of CRNS stations. The diameter of the CRNS footprint varies between 260 to 480 m [63]. Previous studies [63], [64] proved that the CRNS effective penetration

depth follows an exponential function, ranging from 12 to 76 cm depending on the soil moisture. The remote sensing soil moisture retrievals are typically assumed to be uniform for the penetration depth of the sensor and nonuniformity as function of depth is not considered. Studies [65], [66] have demonstrated that the RMSE of our CRNS measurements is generally below $0.03 \text{ cm}^3/\text{cm}^3$ despite the diverse characteristics of the sites (e.g., land use, biomass, meteorological conditions, soil texture). In general, both CRNS measurements and satellite soil moisture products exhibit similar sensitivity to the soil moisture within the first few centimeters. The CRNS method for measuring soil moisture content has been employed for validation of remote sensing datasets on soil moisture, and its reliability has been demonstrated in a number of studies [60], [67], [68], [69]. This indicates that the CRNS is suitable for validating remote sensing soil moisture retrievals as well as being targeted in the LSTM training. When used for validating the remote sensing soil moisture datasets, the CRNS are resampled onto the satellite grid using a nearest neighbor search. Original data processing methods and detailed information regarding the 13 stations can be downloaded and found in previous studies [70], [71].

III. METHODS

The remote sensing soil moisture data merging scheme is depicted in Fig. 2.

A. TC-Based Merging Scheme

The TC analysis was initially introduced and applied in meteorology by Stoffelen [72], which enables the estimation of systematic and random errors in data obtained from three distinct sources without requiring a reference dataset. The methodology assumes a linear relationship between the soil moisture observations and the actual ground truth value θ [29]

$$SM_i = \beta_i \theta + \alpha_i + \varepsilon_i \quad (1)$$

where $i \in [x, y, z]$ is one of the three collocated soil moisture datasets, α_i and β_i represent the additive and multiplicative biases of dataset i , respectively, while ε_i denotes the zero-mean random error. Accordingly, the covariances between the two soil moisture datasets would be given by

$$\begin{aligned} \text{Cov}(SM_x, SM_y) &= E(SM_x SM_y) - E(SM_x) E(SM_y) \\ &= \beta_x \beta_y \sigma_\theta^2 + \beta_y \text{Cov}(\theta, \varepsilon_y) \\ &\quad + \beta_x \text{Cov}(\theta, \varepsilon_x) + \text{Cov}(\varepsilon_x, \varepsilon_y). \end{aligned} \quad (2)$$

Considering the following assumptions [73], orthogonality of errors $\text{Cov}(\theta, \varepsilon_x) = 0$, dependence of errors $\text{Cov}(\varepsilon_x, \varepsilon_y) = 0$, and an expected error of zero $E(\varepsilon_x) = 0$, (2) can be reduced to

$$\begin{aligned} Q_{xy} &= \text{Cov}(SM_x, SM_y) \\ &= \begin{cases} \beta_x \beta_y \sigma_\theta^2 & \text{for } x \neq y \\ \beta_x^2 \sigma_\theta^2 + \sigma_{\varepsilon_x}^2 & \text{for } x = y. \end{cases} \end{aligned} \quad (3)$$

Given that we have seven unknowns ($\beta_x, \beta_y, \beta_z, \sigma_{\varepsilon_x}^2, \sigma_{\varepsilon_y}^2, \sigma_{\varepsilon_z}^2, \sigma_\theta^2$) in a system of six equations, by introducing a new variable

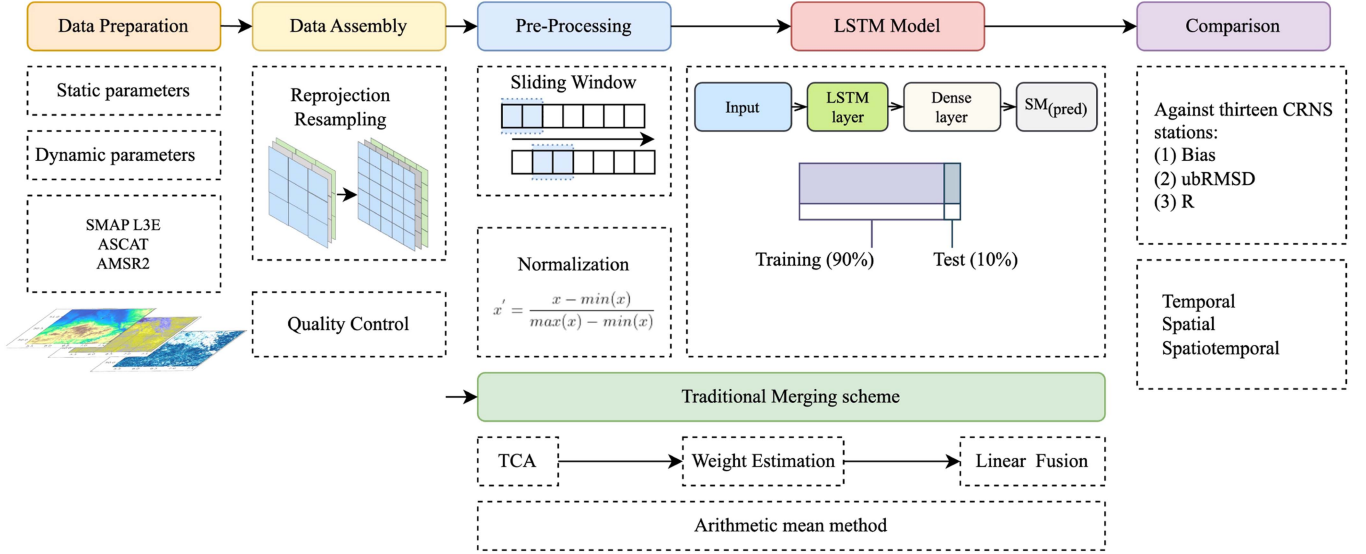


Fig. 2. Flow diagram of the study, comprising five major steps.

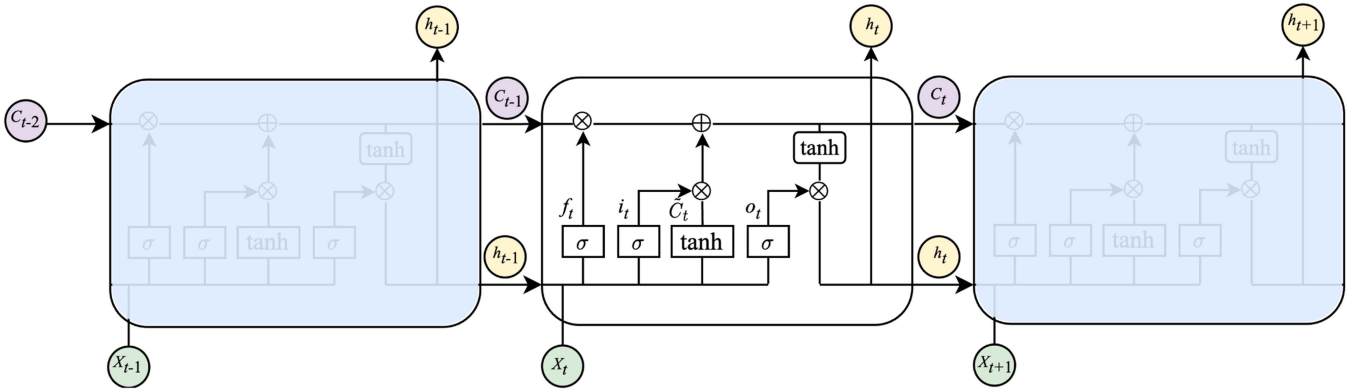


Fig. 3. Schematic diagram of the LSTM neural network structure. f , i , o represent the values of forget gate, input gate, and output gate, respectively. The symbols C , X , and h refer to the cell state, current input and hidden state, respectively. The symbols σ and $\tanh$ represent the sigmoid and hyperbolic tangent activation functions, $\otimes$ means multiplication by weights.

$\lambda_x = \beta_x \sigma_\theta$, the equation could be written as

$$Q_{xy} = \text{Cov}(SM_x, SM_y) = \begin{cases} \lambda_x \lambda_y & \text{for } x \neq y \\ \lambda_x^2 + \sigma_{\varepsilon_x}^2 & \text{for } x = y. \end{cases} \quad (4)$$

Then the error variance values for the three independent datasets can be calculated based on the following equations:

$$\sigma_\varepsilon = \begin{cases} \sqrt{Q_{xx} - \frac{Q_{xy}Q_{xz}}{Q_{yz}}} \\ \sqrt{Q_{yy} - \frac{Q_{xy}Q_{yz}}{Q_{xz}}} \\ \sqrt{Q_{zz} - \frac{Q_{xz}Q_{yz}}{Q_{xy}}} \end{cases} \quad (5)$$

The linear weight fusion method based on least square error can be employed to calculate the weight value for each pixel in each

soil moisture product, and for each grid cell

$$\begin{cases} w_1 \\ w_2 \\ w_3 \end{cases} = \begin{cases} \frac{1/\sigma_{\varepsilon_x}^2}{1/\sigma_{\varepsilon_x}^2 + 1/\sigma_{\varepsilon_y}^2 + 1/\sigma_{\varepsilon_z}^2} \\ \frac{1/\sigma_{\varepsilon_y}^2}{1/\sigma_{\varepsilon_x}^2 + 1/\sigma_{\varepsilon_y}^2 + 1/\sigma_{\varepsilon_z}^2} \\ \frac{1/\sigma_{\varepsilon_z}^2}{1/\sigma_{\varepsilon_x}^2 + 1/\sigma_{\varepsilon_y}^2 + 1/\sigma_{\varepsilon_z}^2} \end{cases}, w_1 + w_2 + w_3 = 1. \quad (6)$$

By using the calculated weight for three remote sensing soil moisture products, the merged soil moisture value at a given grid cell is obtained by

$$SM_{TC} = w_1 * SM_x + w_2 * SM_y + w_3 * SM_z. \quad (7)$$

B. LSTM

The LSTM model, which belongs to the recurrent neural network (RNN) architectures class, has demonstrated notable performances in capturing patterns within sequential data [74]. Unlike conventional RNNs, LSTM models effectively address

the inherent difficulties of modeling long sequential data. The solution employed by the LSTM model involves using gating mechanisms to regulate the memory retention process [75].

As illustrated in Fig. 3, a single LSTM layer comprises a sequential arrangement of cell states, each corresponding to a specific time step t . The gating mechanism is incorporated within each cell state, encompassing a forget gate for determining the exclusion of information from the preceding time step, an input gate for selecting and storing new information from the current time step, and an output gate for determining the information to be outputted from the current time step. During the training phase, the LSTM model incorporates four weight matrices, namely W_f , W_i , W_o , and W_c , along with four bias vectors denoted as b_f , b_i , b_o , and b_c . These weight matrices and bias vectors pertain to the forget, input, output, and candidate gates. These set of weights are updated through an iterative learning process using the stochastic gradient descent during training. In the case where there are multiple input variables, the LSTM model processes each variable independently. The weight matrices (W_f , W_i , W_o , and W_c) generated for all input variables at a given time step are concatenated and used as input for the next time step. Thus, the LSTM model can capture the relationships between different input variables”

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (8)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (9)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (10)$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \tilde{C}_t \quad (11)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (12)$$

$$h_t = o_t \otimes \tanh(C_t). \quad (13)$$

1) *Input and Target Data:* For the LSTM model, the training targets are the CRNS observations from 13 stations. The input data consists of remote sensing soil moisture observations from SMAP L3E, AMSR2, and ASCAT. Additional relevant parameters are categorized into static attributes and dynamic parameters. The static attributes include sand, clay, and silt percentages, bulk density, and altitude from the digital elevation model (DEM). Soil parameters are derived from Soil-Grids [76], a global digital soil mapping system based on soil profile observations from the WoSIS database [77]. Temporal atmospheric forcing data are obtained from regional reanalysis COSMO-REA6 [78], which includes precipitation, air temperature, global radiation, air humidity, and wind speed time series. The time series of leaf area index (LAI) data is extracted from MCD15A2H Version 6 Moderate Resolution Imaging Spectroradiometer (MODIS) Level 4 [79] and this dataset provides an 8-day composite LAI with a pixel size of 500 m.

2) *Experimental Setups:* The training and validating period is from 1 January 2015 to 31 December 2017, with a training-test ratio set at 0.1. In order to facilitate the model’s ability to recognize temporal relationships and understand patterns over multiple time steps (as illustrated in Fig. 3), a sliding window approach is employed during data processing, creating overlapping subsequences of input data with a window size of 10 days. The

TABLE I
SUMMARY OF PERFORMANCE METRICS IN THE LSTM MODEL TRAINING PERIOD (2015 TO 2017)

	RMSE: cm ³ /cm ³	
	Training (90%)	Testing (10%)
LSTM	0.055	0.045
LSTM_DFSF	0.058	0.052
LSTM_DT	0.056	0.045
LSTM_ST	0.063	0.064

architecture of the LSTM model comprises a single LSTM layer followed by a dense layer. The LSTM layer is configured with a hidden state feature size of 256. The input data are represented as a matrix with dimensions equal to number of features $\times$ number of time steps.

The LSTM model is trained with soil moisture retrievals, with and without static and dynamic features to simulate the target soil moisture measured by CRNS at a single grid pixel. Then the trained model is applied to 13 CRNS stations as well as other grid cells for another independent year, specifically from 10 January to 31 December 2018. Four scenarios are considered in the analysis: LSTM, which includes both static and dynamic data in the input; LSTM_DFSF, where both static and dynamic data are excluded; LSTM_DT and LSTM_ST, representing scenarios where only dynamic or static data are present, respectively.

The performance of the LSTM model is evaluated over the training and testing period. A summary of the model performance in the training period (2015–2017) is provided in Table I. An agreement is observed between the CRNS measurements and the predicted soil moisture in both the training and testing datasets in most scenarios. A smaller RMSE is observed in the testing set than in the training dataset as we seek to enhance the testing accuracy and avoid overfitting. The observed smaller RMSE in the testing dataset can be attributed, at least in part, to the smaller size of the testing set. A smaller dataset can for example exhibit less variation in the data, which may result in capturing less extreme values that would otherwise contribute to a higher RMSE. Furthermore, an early stopping mechanism is incorporated into the model training process to prevent overfitting of the training data and to enhance the generation of reliable results for unseen data.

C. Performance Metrics

The validation is conducted for satellite soil moisture products, as well as for the merged dataset. At each CRNS site, the bias, RMSE, unbiased root mean square error (ubRMSE), and Pearson correlation coefficient (r) are calculated. θ_i represents the corresponding satellite retrieval or merged dataset, θ_{crns} represents the CRNS soil moisture measurements, and $E[\cdot]$ is the expectation operator. σ_i and σ_{crns} are the time-variances of the estimated (retrieval or merged dataset) and observed soil moisture for the pixel, respectively

$$\text{Bias} = E(\theta_i) - E(\theta_{\text{crns}}) \quad (14)$$

$$\text{RMSE} = \sqrt{E[(\theta_i - (\theta_{\text{crns}}))]^2} \quad (15)$$

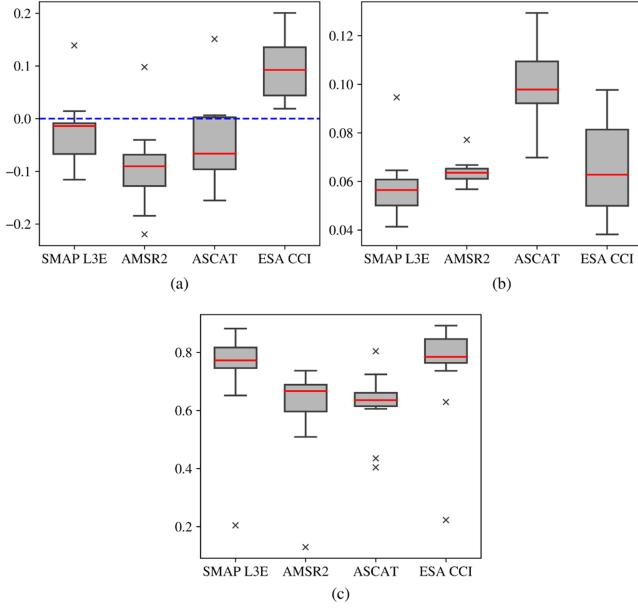


Fig. 4. Box plots of statistical metrics for comparison between remote sensing soil moisture and CRNS soil moisture at 13 stations. The metrics are calculated for the period from January 10, 2018 to December 31, 2018. (a) Bias. (b) ubRMSD. (c) R.

$$\text{ubRMSE} = \sqrt{E[(\theta_i - E(\theta_i)) - (\theta_{\text{crns}} - E(\theta_{\text{crns}}))^2]} \quad (16)$$

$$R = \frac{\sqrt{E[(\theta_i - E(\theta_i)) \cdot (\theta_{\text{crns}} - E(\theta_{\text{crns}}))]}{\sigma_i \sigma_{\text{crns}}} \quad (17)$$

IV. RESULTS AND DISCUSSIONS

A. Evaluation of Remote Sensing Products

Fig. 4 summarizes the error metrics for remote sensing soil moisture products compared to CRNS observations at 13 stations between 11 January 2018 to 31 December 2018. The SMAP L3E has the best performance, with median bias of $-0.014 \text{ cm}^3/\text{cm}^3$, median ubRMSD of $0.057 \text{ cm}^3/\text{cm}^3$, and median R of 0.77. For AMSR2 and ASCAT, the median ubRMSD is $0.064 \text{ cm}^3/\text{cm}^3$ and $0.098 \text{ cm}^3/\text{cm}^3$, respectively, while median R are 0.67 and 0.64. The SMAP L3E, AMSR2, and ASCAT show negative bias compared to CRNS, which could be attributed to the vertical depth mismatch between remote sensing products and CRNS observations. However, the ESA CCI passive dataset gives a wet bias of approximately $0.09 \text{ cm}^3/\text{cm}^3$, the finding is consistent with previous studies [80], [81]. The large difference between in situ observations and ASCAT and AMSR2 might be attributed to the soil porosity data obtained from LDAS, which converts soil saturation into volumetric soil moisture. The key to improving the accuracy of volumetric soil moisture measurements obtained from AMSR2 and ASCAT systems lies in acquiring high quality soil porosity data characterized at high spatial resolution. The time series of satellite-based soil moisture retrievals and corresponding ground soil moisture averaged over 13 sites are presented in Fig. 5. Here, we present the average in order to reduce the influence of spatial heterogeneity, which

might be introduced by scale mismatches between remote sensing products and in situ measurements. SMAP L3E, ASCAT, and ESA CCI passive exhibit the ability to represent the seasonal variations approximately. Nevertheless, the soil moisture range observed in the AMSR2 data exhibits significant damping.

B. Comparison of LSTM Performance With Different Setups

Fig. 6 shows the spatial patterns of the merged soil moisture derived from different LSTM setups. It is found that the LSTM merged soil moisture demonstrates an integration of the distinctive attributes found in both LSTM_ST and LSTM_DT, on the contrary, the LSTM_DFSF exhibits the lowest spatial variability. In order to evaluate the accuracy of the different LSTM setups and compare the influence of different features, we compare evaluation metrics for different LSTM setups with respect to reproducing measured soil moisture at CRNS stations in Fig. 7. It is found that the LSTM gives the best performances, with relative low median bias of $-0.013 \text{ cm}^3/\text{cm}^3$, median ubRMSD of $0.052 \text{ cm}^3/\text{cm}^3$, and high median R values of 0.77. Conversely, the LSTM_DT model exhibits the poorest performance, characterized by a larger positive median bias of $0.136 \text{ cm}^3/\text{cm}^3$ and median ubRMSD value of $0.106 \text{ cm}^3/\text{cm}^3$, along with the lowest median R value of 0.55. In terms of all evaluation metrics, LSTM_ST demonstrates superior performance compared to LSTM_DFSF, indicating that the soil moisture at finer resolution is mostly influenced by the static parameters. Previous studies demonstrated that the small scale variability of soil moisture is primarily influenced by topography and soil textural parameters. Meteorological variables, on the other hand, exhibit a significant impact on soil moisture variability mostly at regional or watershed scales [38], [82]. The poor performance of the LSTM_DFSF scenario compared to the LSTM scenario also illustrates that both dynamic parameters and static features play an important role while predicting soil moisture at finer spatial scales.

To study the temporal dynamics of the soil moisture, the in situ data and merged soil moisture datasets derived from different LSTM setups are compared in Fig. 8. The LSTM and LSTM_ST merged datasets exhibit consistent agreement with observed dry-down and wetting-up patterns, demonstrating their ability to capture precipitation intensities accurately. The LSTM_DFSF and LSTM_DT merged datasets can hardly capture the temporal soil moisture variations, especially during soil moisture dry down events, which usually happens during a period following the precipitation events. In contrast to LSTM_DFSF, LSTM_DT exhibits a more pronounced overestimation, arising from the duplication of precipitation data present in both the dynamic features and the remote sensing soil moisture datasets. This demonstrates that the LSTM model learns the features of various soil types, and captures the soil dry down patterns over time.

C. Evaluation of Merged Soil Moisture by TC

The relative weights of SMAP L3E, AMSR2, and ASCAT calculated by the TC based method are shown in Fig. 9. The weights are assigned to each dataset as function of the magnitudes of their error variances. Lower error variances will be

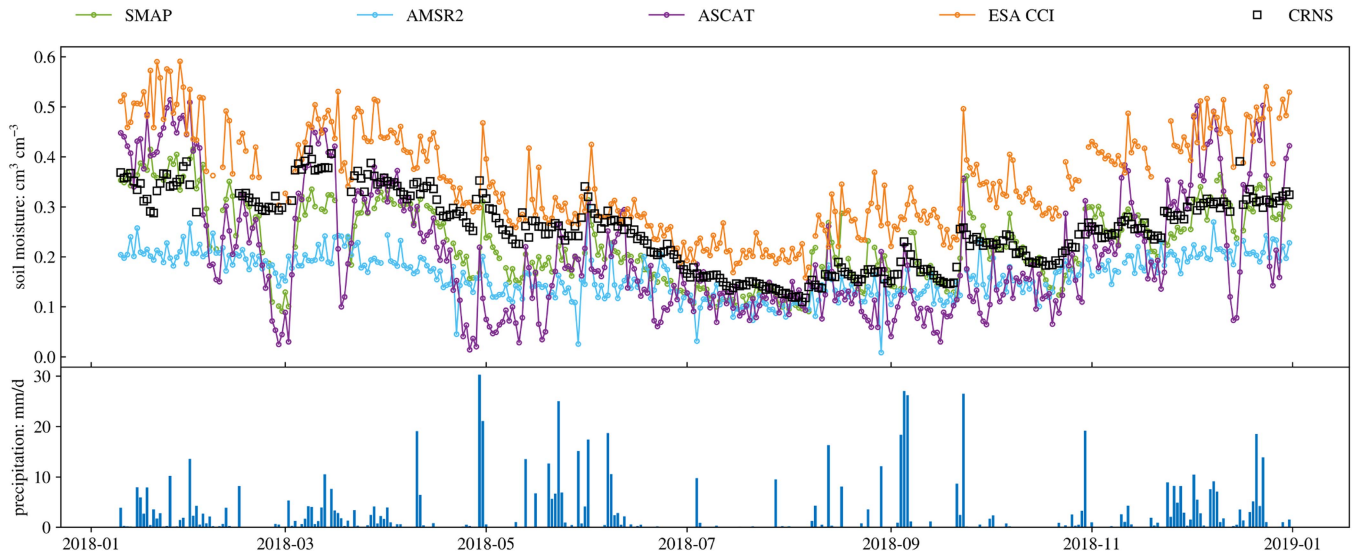


Fig. 5. Time series of soil moisture in situ observations (black square), SMAP L3E (green dots), AMSR (cyan dots), ASCAT (purple dots), and ESA CCI (orange dots) over 13 stations for 10 January 2018–31 December 2018. The bottom graph represents the daily time series of precipitation forcing from COSMO-REA6.

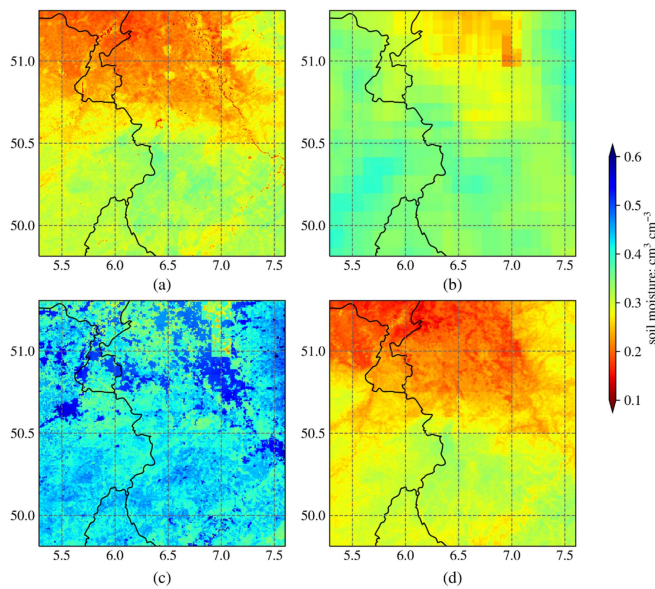


Fig. 6. Comparison of spatial soil moisture maps from different LSTM scenarios. Shown are average soil moisture distributions for the period from January 10, 2018 to December 31, 2018. (a) LSTM. (b) LSTM_DFSF. (c) LSTM_DT. (d) LSTM_ST.

associated with higher weights, whereas higher error variances are associated with lower weights. In general, the ASCAT has low weights (below 0.2), implying that the merged soil moisture product receives larger weights from SMAP L3E and AMSR2. SMAP L3E exhibits high weights in southern regions, which are predominantly characterized by dense forests and complex topographic features. This indicates that the SMAP L3E demonstrates the least variability in error estimates within areas of high vegetation density. This finding is consistent with previous studies, that SMAP could capture the soil moisture under conditions of dense vegetation coverage [49], [58], [83], [84].

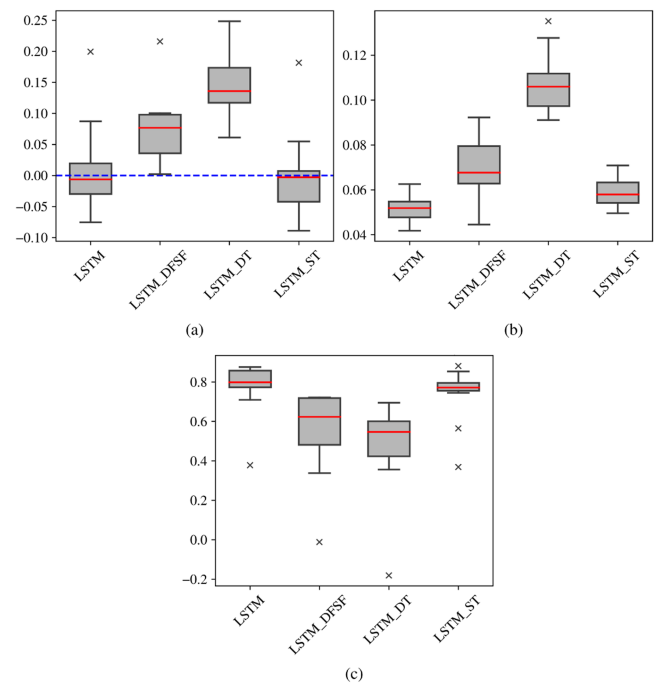


Fig. 7. Box plots of statistical metrics by comparing soil moisture from the LSTM scenario and measured by CRNS evaluated at 13 CRNS stations. The metrics are calculated for the period from January 10, 2018 to December 31, 2018. (a) Bias. (b) ubRMSD. (c) R.

Fig. 10 shows the spatial distribution of soil moisture over the simulation domain. The TC combined dataset is derived from the original products, which in theory has the potential to reduce the random retrieval errors observed in the original products. Our results reveals distinct spatial patterns among the individual products, whereas the merged product exhibits wet and dry patterns similar to the SMAP L3E product. Specifically,

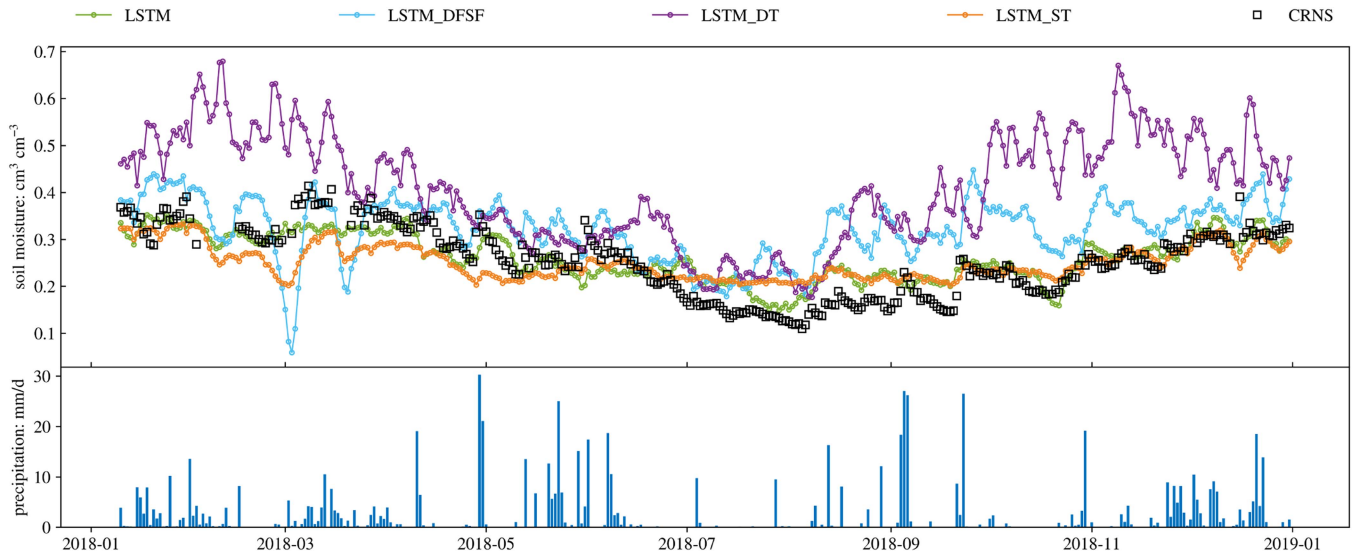


Fig. 8. Time series of soil moisture in situ observations (black square), LSTM (green dots), LSTM_DFSF (cyan dots), LSTM_DT (purple dots), and LSTM_ST (orange dots) over 13 stations for 10 January 2018–31 December 2018. The bottom graph represents the daily time series of precipitation forcing from COSMO-REA6.

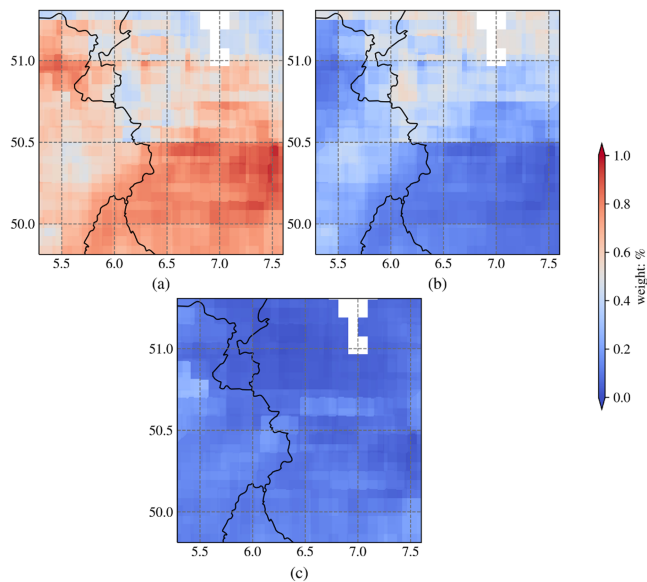


Fig. 9. Spatial distribution of the weights of (a) SMAP L3E; (b) AMSR2; and (c) ASCAT estimated using triple collocation for data fusion of soil moisture.

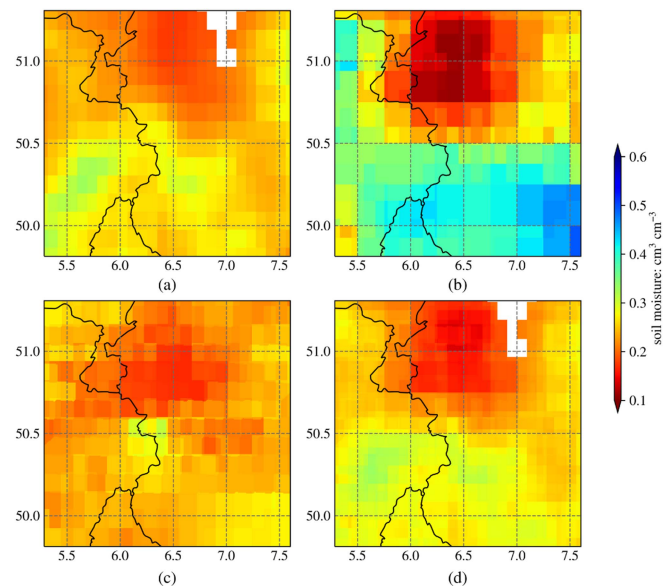


Fig. 10. Spatial soil moisture distributions derived from original and merged dataset. Shown are average soil moisture distributions for the period from January 10, 2018 to December 31, 2018. (a) SMAP L3E. (b) AMSR2. (c) ASCAT. (d) TC merged.

the figure shows high levels of soil moisture in southwestern areas accompanied by a general decrease in wetness in central northern regions.

D. Comparison of Merged Soil Moisture Datasets From TC Based, Mean, and LSTM Scenario

The boxplot (see Fig. 11) summarizes the comparison between the LSTM merged product, TC merged product, and the simple arithmetic average dataset (equal weights weight for each parent dataset). It shows that the LSTM merged product

and TC merged product have similar performance in terms of median ubRMSD and R (ubRMSE = $0.052 \text{ cm}^3/\text{cm}^3$ and $0.052 \text{ cm}^3/\text{cm}^3$, $R = 0.79$ and 0.80 , respectively). However, the TC merged product has a larger bias than the LSTM merged product ($-0.041 \text{ cm}^3/\text{cm}^3$ compared to $-0.006 \text{ cm}^3/\text{cm}^3$). The arithmetic average product shows the worst performance, with the largest bias of $-0.058 \text{ cm}^3/\text{cm}^3$, a median ubRMSE of $0.058 \text{ cm}^3/\text{cm}^3$, and a median R of 0.76. The three merged soil moisture datasets show better performance than the ESA CCI

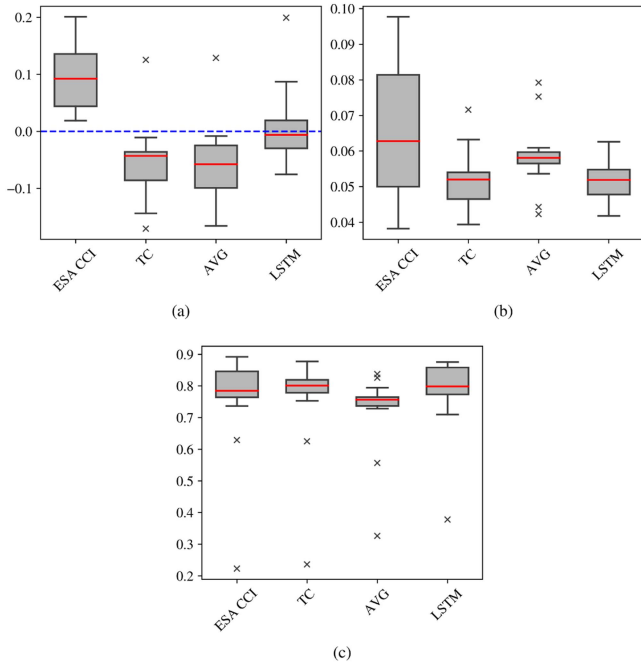


Fig. 11. Box plots of statistical metrics for comparison between merged soil moisture products at 13 CRNS stations. (a) Bias. (b) ubRMSD. (c) R.

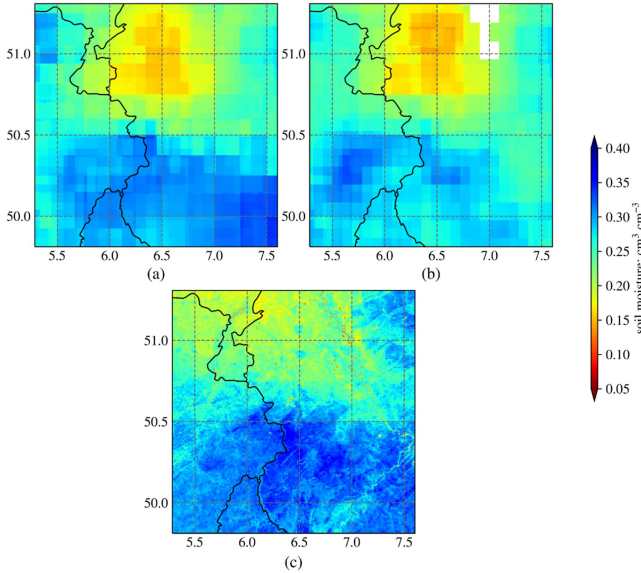


Fig. 12. Comparison of spatial soil moisture distributions derived from merged datasets. Shown are average soil moisture distributions for the period from January 10, 2018 to December 31, 2018. (a) Arithmetic average method. (b) TC based method. (c) LSTM scenario.

passive in terms of all evaluation metrics. Recall from previous sections, the median bias, median ubRMSD, and median R are $0.093 \text{ cm}^3/\text{cm}^3$, $0.063 \text{ cm}^3/\text{cm}^3$, and 0.75, respectively.

Fig. 12 gives the spatial patterns of the three merged soil moisture products. It reveals that the merged products incorporated the characteristics from the original products. In theory, the merged soil moisture products should reduce the random

retrieval error associated with remote sensing instruments, algorithms, and other factors. The spatial pattern of each merged soil moisture product shows similar spatial distributions, exhibiting elevated soil moisture levels in the southern mountainous regions and lower soil moisture levels in the northern regions. Some discrepancies are observed between the TC merging approach and arithmetic average method in southeast and northeast areas. Notably, the weight assigned to the SMAP L3E product exceeds 0.6 in these areas (see Fig. 9). The LSTM merged soil moisture exhibits significant spatial heterogeneity, as it integrates fine scale resolution information derived from both static and dynamic parameters.

To have a comprehensive understanding of the performance of different data fusion methods, the soil moisture time series are visually presented in Fig. 13. It is found that the TC and arithmetic mean merged products exhibit generally comparable variability particularly in scenarios characterized by minimal discrepancies between the original products. However, the TC method demonstrates superior performance under low soil moisture conditions compared to the simple averaging method. Nevertheless, the TC based method should be preferred over the simple arithmetic mean method, as it could yield optimal weighing factors and result in a better merged product. The LSTM merged product agrees well with the temporal dynamics of in situ measurements, indicating its capability to capture the relationship between remote sensing observation and CRNS soil moisture measurement, considering the influence of penetration depth.

V. DISCUSSIONS

As shown in Figs. 12 and 13, the LSTM-based merged product exhibits more variability in spatial patterns and a closer alignment with CRNS measurements compared to the TC-based merging dataset. The TC merged dataset exhibits more sensitivity to the extreme events (e.g, wetting and drying conditions) that are subject to the similar performance in the parental remote sensing products. This might be attributed to the limited information constrained in the remotely sensed soil moisture penetration depth, which lacks information about the deeper soil. In contrast, the LSTM is capable of exploiting this temporal information from CRNS during the training process. In addition, its window size mechanism allows the model to learn long-term dependencies and effectively capture the soil moisture memory.

While complete independence of in situ measurements would be optimal for validation purposes, the acquisition of such data may be constrained by the fact that only 13 in situ stations are located in the research area. Some validation sites, in particular those with high bulk density or dense vegetation cover, exhibited unique performance characteristics. Therefore, we proposed that the inclusion of training data across a range of conditions could enhance the robustness of the LSTM model. Alternatively, we implemented an evaluation strategy, which involves training the model on data spanning from 2015 to 2017 and evaluating on an entirely independent year (2018), which was characterized by significant dryness. This allows us to assess the generalizability of the model across different conditions.

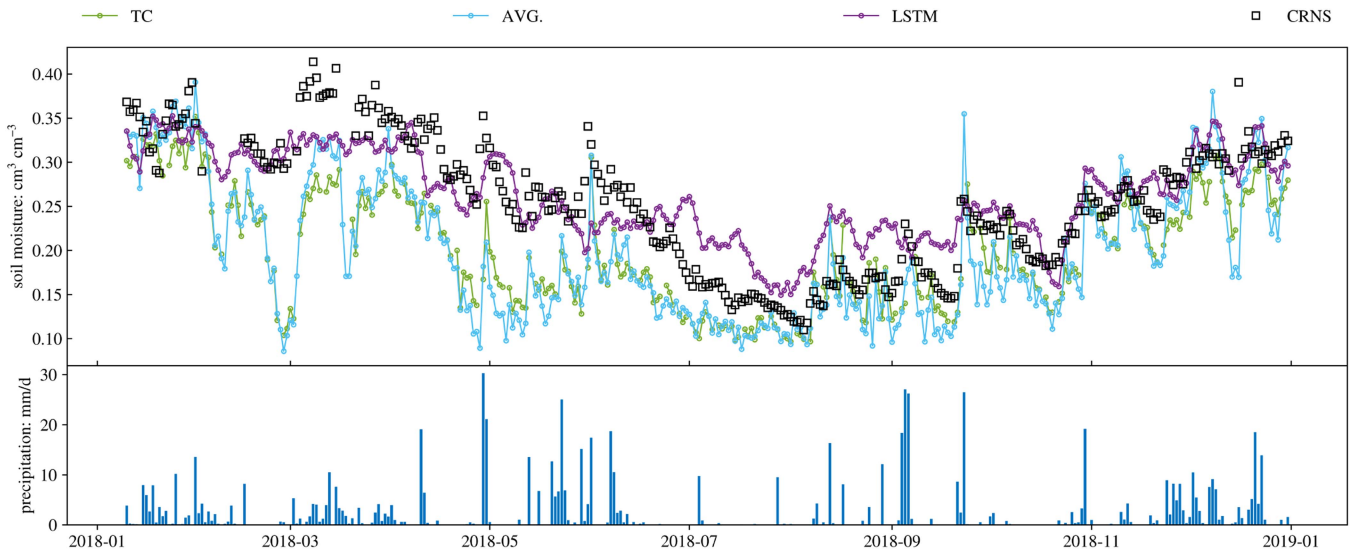


Fig. 13. Time series of soil moisture in situ observations (black square) and merged soil moisture from TC (green dots), merged soil moisture by arithmetic mean (cyan dots), and merged soil moisture by LSTM (purple dots) over 13 stations for 10 January 2018–31 December 2018. The bottom graph represents the daily time series of precipitation forcing from COSMO-REA6.

Our findings demonstrate that LSTM can be an effective approach to merge soil moisture observations from active and passive satellite microwave sensors, however, there are some limitations in the approach. The LSTM model targets in situ data with parameter tuning. In contrast, the TC merging scheme relies solely on error calculations. Thus, the comparison with TC-based merging datasets may not be entirely fair. Despite this, we present our findings to demonstrate the applicability of the LSTM model and encourage further research in this direction. The LSTM model application can be hindered in the areas where the in situ soil moisture measurements are sparse, such as Africa. A spatial mismatch exists between the LSTM grid and in situ sampling, which increases the uncertainty in the training process. In addition, the LSTM merging scheme generates a higher resolution merged dataset by leveraging information from high resolution static and dynamic feature datasets, but its effectiveness can be limited in complex landscapes. In addition, there is also a vertical mismatch between the remote sensing products and the CRNS measurement, which could affect the comparability of the merged product with the CRNS soil moisture datasets. Although the CRNS shows advantages for the validation of the near-surface soil moisture products [60], [85], [86], the higher penetration depth of the CRNS should be considered [60], [87].

Previous studies [27] have demonstrated that the TC merging scheme does not always produce a superior product to the simple arithmetic mean method. Peng et al. [21] have confirmed this finding and demonstrated that TC-based approaches provide better results when the parent products exhibit large differences. The results obtained in this study align with those studies and show that the TC merged product exhibits superior performance in terms of smaller bias and lower ubRMSE and higher Pearson r when compared to the arithmetic mean soil moisture product. The TC merging scheme could be further improved to optimally retain both spatial and temporal information by taking account

the spatial and temporal nonstationary errors [88]. This would facilitate a more accurate comparison with other ML merging methodologies.

Future research can enhance the predicting capacity of the model and evaluate its performance in areas that are not included in the training, which could be done by increasing the training data or using transfer learning techniques. Moreover, related physical constraints can be introduced into LSTM model to guide the training process. Studies [89], [90] have shown that the performance of physics-informed ML can be significantly improved in scenarios traditionally challenging for traditional ML methods, such as sparse datasets and extreme values.

VI. CONCLUSION

The acquisition of high-quality soil moisture data at high spatial and temporal resolution is of importance for hydrological and meteorological applications. In this study, we select a region in western Europe exhibiting very heterogeneous land use and topographical variations. In the research area, multiple sources of soil moisture products are accessible, from either satellite retrievals (e.g., SMAP L3E, AMSR2, and ASCAT) or field measurements by CRNS. We train a LSTM model with retrievals and a limited number of in situ measurements. Subsequently, we employ the pretrained LSTM model to make predictions in research areas where in situ measurements are not attainable. Our investigation includes the assessment of various parameter configurations within the LSTM model. A TC based merging scheme and arithmetic mean method are also employed to merge the different satellite soil moisture products and generate a merged soil moisture product.

Results for the different scenarios of the LSTM model illustrate the importance of incorporating both static and dynamic data in the soil moisture merging scheme at finer spatial resolution. Static data like soil texture information influences the

spatial soil moisture variation not only by the high resolution information, but also by its role in the soil drying process over time. Accurate static information at high spatial resolution exerts an importance influence on enhancing the correlation between the merged soil moisture dataset and in situ measurements.

The LSTM and TC merged soil moisture datasets exhibit the ability to effectively integrate soil moisture information derived from diverse remote sensing products. They demonstrate comparable performance in terms of ubRMSE and R when compared to in situ measurements, better than a soil moisture dataset based on the simple arithmetic mean of the different remote sensing products. Notably, the LSTM approach demonstrates reduced bias and superior handling of spatial and temporal variations than the TC merged dataset, owing to its intricate weighting calculations by incorporating information from both static and dynamic parameters at finer resolution. This demonstrates the potential of LSTM as a robust tool for soil moisture prediction and characterization and scaling.

ACKNOWLEDGMENT

This article used publicly available data. The satellite soil moisture datasets (SMAP, AMSR2, ASCAT, and ESA CCI) can be downloaded from their official websites. Forcing and soil texture data can be obtained from the COSMO-REA6 and SoilGrids websites, respectively. We thank the Terrestrial Environmental Observatories (TERENO) community for providing observation data and the JUWELS supercomputer in Forschungszentrum Jülich for providing storage and computational time.

REFERENCES

- [1] A. Henderson-Sellers, "Soil moisture: A critical focus for global change studies," *Glob. Planet. Change*, vol. 13, no. 1–4, pp. 3–9, 1996.
- [2] S. I. Seneviratne et al., "Investigating soil moisture–climate interactions in a changing climate: A review," *Earth-Sci. Rev.*, vol. 99, no. 3/4, pp. 125–161, 2010.
- [3] P. Dobriyal, A. Qureshi, R. Badola, and S. A. Hussain, "A review of the methods available for estimating soil moisture and its implications for water resource management," *J. Hydrol.*, vol. 458, pp. 110–117, 2012.
- [4] E. T. Engman, "Applications of microwave remote sensing of soil moisture for water resources and agriculture," *Remote Sens. Environ.*, vol. 35, no. 2/3, pp. 213–226, 1991.
- [5] B. P. Mohanty, M. H. Cosh, V. Lakshmi, and C. Montzka, "Soil moisture remote sensing: State-of-the-science," *Vadose Zone J.*, vol. 16, no. 1, pp. 1–9, 2017.
- [6] T. Zhao et al., "Retrievals of soil moisture and vegetation optical depth using a multi-channel collaborative algorithm," *Remote Sens. Environ.*, vol. 257, 2021, Art. no. 112321.
- [7] E. G. Njoku, T. J. Jackson, V. Lakshmi, T. K. Chan, and S. V. Nghiem, "Soil moisture retrieval from AMSR-E," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 2, pp. 215–229, Feb. 2003.
- [8] K. Imaoka, M. Kachi, M. Kasahara, N. Ito, K. Nakagawa, and T. Oki, "Instrument performance and calibration of AMSR-E and AMSR2," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 38, no. 8, pp. 13–18, 2010.
- [9] Z. Bartalis et al., "Initial soil moisture retrievals from the METOP-A advanced scatterometer (ASCAT)," *Geophys. Res. Lett.*, vol. 34, no. 20, 2007, Art. no. L20401, doi: [10.1029/2007GL031088](https://doi.org/10.1029/2007GL031088).
- [10] L. Brocca et al., "A review of the applications of ASCAT soil moisture products," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 2285–2306, May 2017.
- [11] J.-P. Wigneron et al., "Modelling the passive microwave signature from land surfaces: A review of recent results and application to the L-band SMOS & SMAP soil moisture retrieval algorithms," *Remote Sens. Environ.*, vol. 192, pp. 238–262, 2017.
- [12] M. Baur, T. Jagdhuber, M. Link, M. Piles, R. Akbar, and D. Entekhabi, "Multi-frequency estimation of canopy penetration depths from SMAP/AMSR2 radiometer and IceSAT lidar data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 365–368.
- [13] H. Ma et al., "An assessment of L-band surface soil moisture products from SMOS and SMAP in the tropical areas," *Remote Sens. Environ.*, vol. 284, 2023, Art. no. 113344.
- [14] Y. H. Kerr, P. Waldteufel, J.-P. Wigneron, J. Martinuzzi, J. Font, and M. Berger, "Soil moisture retrieval from space: The soil moisture and ocean salinity (SMOS) mission," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 8, pp. 1729–1735, Aug. 2001.
- [15] D. Entekhabi et al., "The soil moisture active passive (SMAP) mission," *Proc. IEEE*, vol. 98, no. 5, pp. 704–716, May 2010.
- [16] M. C. Dobson and F. T. Ulaby, "Active microwave soil moisture research," *IEEE Trans. Geosci. Remote Sens.*, vol. GE-24, no. 1, pp. 23–36, Jan. 1986.
- [17] B. W. Barrett, E. Dwyer, and P. Whelan, "Soil moisture retrieval from active spaceborne microwave observations: An evaluation of current techniques," *Remote Sens.*, vol. 1, no. 3, pp. 210–242, 2009.
- [18] L. Karthikeyan, M. Pan, N. Wanders, D. N. Kumar, and E. F. Wood, "Four decades of microwave satellite soil moisture observations: Part I. A review of retrieval algorithms," *Adv. Water Resour.*, vol. 109, pp. 106–120, 2017.
- [19] K. Zhang et al., "Using multi-satellite microwave remote sensing observations for retrieval of daily surface soil moisture across China," *Water Sci. Eng.*, vol. 12, no. 2, pp. 85–97, 2019.
- [20] N. Zhang, S. M. Quiring, and T. W. Ford, "Blending noah, SMOS, and in situ soil moisture using multiple weighting and sampling schemes," *J. Hydrometeorol.*, vol. 22, no. 7, pp. 1835–1854, 2021.
- [21] J. Peng et al., "Estimation and evaluation of high-resolution soil moisture from merged model and earth observation data in the great britain," *Remote Sens. Environ.*, vol. 264, 2021, Art. no. 112610.
- [22] S. Kim, R. M. Parinussa, Y. Y. Liu, F. M. Johnson, and A. Sharma, "A framework for combining multiple soil moisture retrievals based on maximizing temporal correlation," *Geophys. Res. Lett.*, vol. 42, no. 16, pp. 6662–6670, 2015.
- [23] S. Kim, R. M. Parinussa, Y. Y. Liu, F. M. Johnson, and A. Sharma, "Merging alternate remotely-sensed soil moisture retrievals using a non-static model combination approach," *Remote Sens.*, vol. 8, no. 6, 2016, Art. no. 518.
- [24] K. Scipal, W. Dorigo, and R. deJeu, "Triple collocation—A new tool to determine the error structure of global soil moisture products," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2010, pp. 4426–4429.
- [25] C.-H. Su, D. Ryu, W. T. Crow, and A. W. Western, "Beyond triple collocation: Applications to soil moisture monitoring," *J. Geophys. Res., Atmos.*, vol. 119, no. 11, pp. 6419–6439, 2014.
- [26] A. Gruber, C.-H. Su, S. Zwieback, W. Crow, W. Dorigo, and W. Wagner, "Recent advances in (soil moisture) triple collocation analysis," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 45, pp. 200–211, 2016.
- [27] M. Yilmaz, W. Crow, M. Anderson, and C. Hain, "An objective methodology for merging satellite- and model-based soil moisture products," *Water Resour. Res.*, vol. 48, no. 11, 2012, Art. no. W11502, doi: [10.1029/2011WR011682](https://doi.org/10.1029/2011WR011682).
- [28] W. Anderson et al., "Towards an integrated soil moisture drought monitor for east Africa," *Hydrol. Earth Syst. Sci.*, vol. 16, no. 8, pp. 2893–2913, 2012.
- [29] A. Gruber, W. A. Dorigo, W. Crow, and W. Wagner, "Triple collocation-based merging of satellite soil moisture retrievals," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 6780–6792, Dec. 2017.
- [30] H. Kim et al., "Global-scale assessment and combination of SMAP with ASCAT (active) and AMSR2 (passive) soil moisture products," *Remote Sens. Environ.*, vol. 204, pp. 260–275, 2018.
- [31] S. Kim, H. T. Pham, Y. Y. Liu, L. Marshall, and A. Sharma, "Improving the combination of satellite soil moisture data sets by considering error cross correlation: A comparison between triple collocation (TC) and extended double instrumental variable (EIVD) alternatives," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7285–7295, Sep. 2021.
- [32] J. Kang, R. Jin, and X. Li, "An advanced framework for merging remotely sensed soil moisture products at the regional scale supported by error structure analysis: A case study on the Tibetan plateau," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3614–3624, 2021.
- [33] Q. Xie, L. Jia, M. Menenti, and G. Hu, "Global soil moisture data fusion by triple collocation analysis from 2011 to 2018," *Sci. Data*, vol. 9, no. 1, 2022, Art. no. 687.
- [34] L. Li, Q. Zhu, Y. Liu, X. Lai, and K. Liao, "Fusing satellite-based surface soil moisture products over a typical region with complex land surface characteristics," *J. Hydrol.*, vol. 612, 2022, Art. no. 128158.

- [35] X. Min, Y. Shangguan, D. Li, and Z. Shi, "Improving the fusion of global soil moisture datasets from SMAP, SMOS, ASCAT, and MERRA2 by considering the non-zero error covariance," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 113, 2022, Art. no. 103016.
- [36] S. Salcedo-Sanz et al., "Machine learning information fusion in earth observation: A comprehensive review of methods, applications and data sources," *Inf. Fusion*, vol. 63, pp. 256–272, 2020.
- [37] E. Babaeian, S. Paheding, N. Siddique, V. K. Devabhaktuni, and M. Tuller, "Estimation of root zone soil moisture from ground and remotely sensed soil information with multisensor data fusion and automated machine learning," *Remote Sens. Environ.*, vol. 260, 2021, Art. no. 112434.
- [38] L. Karthikeyan and A. K. Mishra, "Multi-layer high-resolution soil moisture estimation using machine learning over the United States," *Remote Sens. Environ.*, vol. 266, 2021, Art. no. 112706.
- [39] 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, 2022, Art. no. 155066.
- [40] V. Batchu, G. Nearing, and V. Gulshan, "A deep learning data fusion model using Sentinel-1/2, SoilGrids, SMAP-USDA, and GLDAS for soil moisture retrieval," *J. Hydrometeorol.*, vol. 24, no. 10, pp. 1789–1823, 2023.
- [41] F. Herrmann, L. Keller, R. Kunkel, H. Vereecken, and F. Wendland, "Determination of spatially differentiated water balance components including groundwater recharge on the federal state level—A case study using the MGROWA model in north Rhine-Westphalia (Germany)," *J. Hydrol., Regional Stud.*, vol. 4, pp. 294–312, 2015.
- [42] H. Zhao, C. Montzka, R. Baatz, H. Vereecken, and H.-J. H. Franssen, "The importance of subsurface processes in land surface modeling over a temperate region: An analysis with SMAP, cosmic ray neutron sensing and triple collocation analysis," *Remote Sens.*, vol. 13, no. 16, 2021, Art. no. 3068.
- [43] P. E. O'Neill et al., "SMAP enhanced L3 radiometer global and polar grid daily 9 km EASE-Grid soil moisture, version 5," NASA National Snow and Ice Data Center Distributed Active Archive Center. 2021, Accessed on: Jul. 1, 2023. [Online]. Available: https://nsidc.org/data/SPL3SMP_E/versions/5
- [44] M. Owe and R. de Jeu, "AMSR2/GCOM-W1 surface soil moisture (LPRM) L3 1 day 25 km x 25 km descending V001, Edited by goddard earth sciences data and information services center (GES DISC) (Bill Teng)," Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), 2014, Accessed on: Jul. 1, 2023. [Online]. Available: https://disc.gsfc.nasa.gov/datasets/LPRM_AMSR2_D_SOILM3_001/summary
- [45] H. Saf, "Metop ASCAT surface soil moisture climate data record v7 12.5 km sampling (H119)," EUMETSAT SAF on Support to Operational Hydrology and Water Management, 2021, Accessed on: Jul. 1, 2023, doi: [10.15770/EUM_SAF_H_0009](https://doi.org/10.15770/EUM_SAF_H_0009).
- [46] W. Dorigo et al., "ESA CCI soil moisture for improved earth system understanding: State-of-the-art and future directions," *Remote Sens. Environ.*, vol. 203, pp. 185–215, 2017.
- [47] A. Gruber, T. Scanlon, R. van der Schalie, W. Wagner, and W. Dorigo, "Evolution of the ESA CCI soil moisture climate data records and their underlying merging methodology," *Earth Syst. Sci. Data*, vol. 11, no. 2, pp. 717–739, 2019.
- [48] W. Preimesberger, T. Scanlon, C.-H. Su, A. Gruber, and W. Dorigo, "Homogenization of structural breaks in the global ESA CCI soil moisture multisatellite climate data record," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 2845–2862, Apr. 2021.
- [49] S. K. Chan et al., "Development and assessment of the SMAP enhanced passive soil moisture product," *Remote Sens. Environ.*, vol. 204, pp. 931–941, 2018.
- [50] J. Du, J. S. Kimball, L. A. Jones, Y. Kim, J. Glassy, and J. D. Watts, "A global satellite environmental data record derived from AMSR-E and AMSR2 microwave earth observations," *Earth Syst. Sci. Data*, vol. 9, no. 2, pp. 791–808, 2017.
- [51] R. M. Parinussa, T. R. Holmes, N. Wanders, W. A. Dorigo, and R. A. de Jeu, "A preliminary study toward consistent soil moisture from AMSR2," *J. Hydrometeorol.*, vol. 16, no. 2, pp. 932–947, 2015.
- [52] W. Wagner et al., "The ASCAT soil moisture product: A review of its," *Meteorologische Zeitschrift*, vol. 22, no. 1, pp. 1–29, 2013.
- [53] M. Owe, R. de Jeu, and T. Holmes, "Multisensor historical climatology of satellite-derived global land surface moisture," *J. Geophys. Res.: Earth Surf.*, vol. 113, no. F1, 2008, Art. no. F01002, doi: [10.1029/2007JF000769](https://doi.org/10.1029/2007JF000769).
- [54] A. Chakravorty, B. R. Chahar, O. P. Sharma, and C. Dhanya, "A regional scale performance evaluation of SMOS and ESA-CCI soil moisture products over India with simulated soil moisture from Merra-Land," *Remote Sens. Environ.*, vol. 186, pp. 514–527, 2016.
- [55] X. Xu and S. K. Frey, "Validation of SMOS, SMAP, and ESA CCI soil moisture over a humid region," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10784–10793, 2021.
- [56] H. Ma, J. Zeng, N. Chen, X. Zhang, M. H. Cosh, and W. Wang, "Satellite surface soil moisture from SMAP, SMOS, AMSR2 and ESA CCI: A comprehensive assessment using global ground-based observations," *Remote Sens. Environ.*, vol. 231, 2019, Art. no. 111215.
- [57] M. S. Burgin et al., "A comparative study of the SMAP passive soil moisture product with existing satellite-based soil moisture products," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2959–2971, May 2017.
- [58] H. Cui et al., "Evaluation and analysis of AMSR-2, SMOS, and SMAP soil moisture products in the Genhe area of China," *J. Geophys. Res., Atmos.*, vol. 122, no. 16, pp. 8650–8666, 2017.
- [59] Z. Wang, T. Che, T. Zhao, L. Dai, X. Li, and J.-P. Wigneron, "Evaluation of SMAP, SMOS, and AMSR2 soil moisture products based on distributed ground observation network in cold and arid regions of China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8955–8970, 2021.
- [60] C. Montzka et al., "Validation of spaceborne and modelled surface soil moisture products with cosmic-ray neutron probes," *Remote Sens.*, vol. 9, no. 2, 2017, Art. no. 103.
- [61] M. Zreda et al., "Cosmos: The cosmic-ray soil moisture observing system," *Hydrol. Earth Syst. Sci.*, vol. 16, no. 11, pp. 4079–4099, 2012.
- [62] S. Zacharias et al., "A network of terrestrial environmental observatories in Germany," *Vadose Zone J.*, vol. 10, no. 3, pp. 955–973, 2011.
- [63] M. Köhli, M. Schrön, M. Zreda, U. Schmidt, P. Dietrich, and S. Zacharias, "Footprint characteristics revised for field-scale soil moisture monitoring with cosmic-ray neutrons," *Water Resour. Res.*, vol. 51, no. 7, pp. 5772–5790, 2015.
- [64] M. Schrön et al., "Spatial sensitivity of cosmic-ray neutron sensors applied to improve calibration and validation," *Hydrol. Earth Syst. Sci.*, vol. 21, pp. 5009–5030, 2017.
- [65] H. Bogaen, J. Huisman, R. Baatz, H.-J. Hendricks Franssen, and H. Vereecken, "Accuracy of the cosmic-ray soil water content probe in humid forest ecosystems: The worst case scenario," *Water Resour. Res.*, vol. 49, no. 9, pp. 5778–5791, 2013.
- [66] R. Baatz et al., "Calibration of a catchment scale cosmic-ray probe network: A comparison of three parameterization methods," *J. Hydrol.*, vol. 516, pp. 231–244, 2014.
- [67] A. Colliander et al., "Validation of SMAP surface soil moisture products with core validation sites," *Remote Sens. Environ.*, vol. 191, pp. 215–231, 2017.
- [68] C. Montzka, H. R. Bogaen, M. Herbst, M. H. Cosh, T. Jagdhuber, and H. Vereecken, "Estimating the number of reference sites necessary for the validation of global soil moisture products," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 9, pp. 1530–1534, Sep. 2021.
- [69] T. Zhao et al., "Soil moisture experiment in the Luan river supporting new satellite mission opportunities," *Remote Sens. Environ.*, vol. 240, 2020, Art. no. 111680.
- [70] H. Bogaen 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, pp. 1–22, 2018.
- [71] H. R. Bogaen et al., "Cosmos-Europe: A European network of cosmic-ray neutron soil moisture sensors," *Earth Syst. Sci. Data*, vol. 14, no. 3, pp. 1125–1151, 2022.
- [72] A. Stoffelen, "Toward the true near-surface wind speed: Error modeling and calibration using triple collocation," *J. Geophys. Res.: Oceans*, vol. 103, no. C4, pp. 7755–7766, 1998.
- [73] M. T. Yilmaz and W. T. Crow, "Evaluation of assumptions in soil moisture triple collocation analysis," *J. Hydrometeorol.*, vol. 15, no. 3, pp. 1293–1302, 2014.
- [74] S. Siarni-Namini, N. Tavakoli, and A. S. Namin, "A comparison of ARIMA and LSTM in forecasting time series," in *Proc. 17th IEEE Int. Conf. Mach. Learn. Appl.*, 2018, pp. 1394–1401.
- [75] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [76] 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, 2021.

- [77] N. H. Batjes, E. Ribeiro, A. Van Oostrum, J. Leenaars, T. Hengl, and J. Mendes de Jesus, "Wosis: Providing standardised soil profile data for the world," *Earth Syst. Sci. Data*, vol. 9, no. 1, pp. 1–14, 2017.
- [78] C. Bollmeyer et al., "Towards a high-resolution regional reanalysis for the European cordex domain," *Quart. J. Roy. Meteorol. Soc.*, vol. 141, no. 686, pp. 1–15, 2015.
- [79] R. B. Myneni, Y. Knyazikhin, and T. Park, "MCD15A2H MODIS/Terra aqua leaf area Index/FPAR 8-day L4 global 500 m SIN grid V006 [data set]," NASA EOSDIS Land Processes Distributed Active Archive Center, 2015, Accessed on: Jul. 1, 2023, doi: [10.5067/MODIS/MCD15A2H.006](https://doi.org/10.5067/MODIS/MCD15A2H.006).
- [80] L. Zhu, H. Wang, C. Tong, W. Liu, and B. Du, "Evaluation of ESA active, passive and combined soil moisture products using upscaled ground measurements," *Sensors*, vol. 19, no. 12, 2019, Art. no. 2718.
- [81] Y. Yang et al., "Evaluation of multi-source soil moisture datasets over central and eastern agricultural area of China using in situ monitoring network," *Remote Sens.*, vol. 13, no. 6, 2021, Art. no. 1175.
- [82] W. T. Crow et al., "Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products," *Rev. Geophys.*, vol. 50, no. 2, 2012, Art. no. RG2002, doi: [10.1029/2011RG000372](https://doi.org/10.1029/2011RG000372).
- [83] E. Ayres, A. Colliander, M. H. Cosh, J. A. Roberti, S. Simkin, and M. A. Genazzio, "Validation of SMAP soil moisture at terrestrial national ecological observatory network (neon) sites show potential for soil moisture retrieval in forested areas," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10903–10918, 2021.
- [84] K. M. Andreadis, D. F. Meason, B. Höck, P. Lad, and N. Das, "Evaluation of multiscale SMAP soil moisture products in forested environments," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 2505805.
- [85] O. D. Howells, G. P. Petropoulos, P. K. Srivastava, D. Triantakoustantis, and I. Sandric, "Exploring the potential of SCAT-SAR SWI for soil moisture retrievals at selected COSMOS- UK sites," *Int. J. Remote Sens.*, vol. 42, no. 23, pp. 9155–9169, 2021.
- [86] B. Fersch, T. Jagdhuber, M. Schrön, I. Völksch, and M. Jäger, "Synergies for soil moisture retrieval across scales from airborne polarimetric SAR, cosmic ray neutron roving, and an in situ sensor network," *Water Resour. Res.*, vol. 54, no. 11, pp. 9364–9383, 2018.
- [87] A. T. Bayat et al., "Mapping near-surface soil moisture in a mediterranean agroforestry ecosystem using cosmic-ray neutron probe and sentinel-1 data," in *Proc. IEEE Int. Workshop Metrol. Agriculture Forestry*, 2020, pp. 201–206.
- [88] J. Zhou, W. T. Crow, Z. Wu, J. Dong, H. He, and H. Feng, "A triple collocation-based 2D soil moisture merging methodology considering spatial and temporal non-stationary errors," *Remote Sens. Environ.*, vol. 263, 2021, Art. no. 112509.
- [89] A. Singh, K. Gaurav, G. K. Sonkar, and C.-C. Lee, "Strategies to measure soil moisture using traditional methods, automated sensors, remote sensing, and machine learning techniques: Review, bibliometric analysis, applications, research findings, and future directions," *IEEE Access*, vol. 11, pp. 13605–13635, 2023.
- [90] Y. Wang et al., "A deep learning approach based on physical constraints for predicting soil moisture in unsaturated zones," *Water Resour. Res.*, vol. 59, no. 11, 2023, Art. no. e2023WR035194.



Carsten Montzka (Senior Member, IEEE) received the Ph.D. degree in geography from the University of Bonn, Bonn, Germany, in 2007.

He is currently with the Institute of Bio- and Geosciences: Agrosphere (IBG-3) of the Forschungszentrum Jülich, Jülich, Germany, where he is focusing on hydrological remote sensing.



Harry Vereecken received the M.Sc. degree in agricultural engineering and the Ph.D. degree in agricultural sciences from Katholieke Universiteit Leuven, Leuven, Belgium, in 1982 and 1988, respectively.

His Ph.D. thesis was on the development of pedo-transfer functions to estimate soil hydraulic properties. From 1988 to 1990, he was as a Research Assistant, involved in modeling nitrogen and water fluxes in soils and groundwater. From 1990 to 1992, he was a Researcher with the Institute of Petroleum and Organic Geochemistry, Forschungszentrum Jülich, Jülich, Germany, where he was the Head of the Division "Behavior of Pollutants in Geological Systems" from 1992 to 2000. He was appointed as the Director of the Institute of Bio- and Geosciences, Agrosphere (IBG-3), Forschungszentrum Jülich, in 2000. His research interests include modeling of flow and transport processes in soils and hydrogeophysics.



Harrie-Jan Hendricks Franssen received the M.Sc. degree in soil science from Wageningen Agricultural University, Wageningen, The Netherlands, in 1994, and atmospheric and climate sciences from ETH Zurich, Zurich, Switzerland, in 2008, and the Ph.D. degree in hydrogeology from the Technical University of Valencia, Valencia, Spain, in 2001, on groundwater inverse modeling.

After his Ph.D. he worked as a postdoctoral and a Lecturer with the ETH Zurich in Switzerland, with a focus on stream-aquifer interaction and data assimilation. From 2009 onwards, he is a Full Professor with the RWTH Aachen, Germany, and leads a research group at Forschungszentrum Jülich, Germany. His current research interests include land surface modelling and hydrological modeling, and model-data fusion, with the aim to provide better short- and long-term hydrological and land surface model predictions and projections, and to develop continental-scale land surface and subsurface reanalysis.



Haojin Zhao received the M.Sc. degree in environmental sciences from Peking University, Beijing, China, in 2018. She is currently working toward the Ph.D. degree in hydrology with Forschungszentrum Jülich, Jülich, Germany.

Her research interests include remote sensing, data assimilation, and land surface modeling, with aim to improve hydrological predictions and enhance our understanding of land-atmosphere interactions.