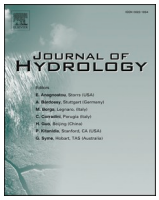
Journal of Hydrology 634 (2024) 131051

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol



Research papers

A novel land surface temperature reconstruction method and its application for downscaling surface soil moisture with machine learning

Onur Güngör Şahin [a], Orhan Gündüz [b,*]

[a] Department of International Water Resources, Izmir Institute of Technology, Izmir Turkiye
[b] Department of Environmental Engineering, Izmir Institute of Technology, Izmir Turkiye

ARTICLE INFO

Keywords:
Land Surface Temperature
Reconstruction
SMAP
Surface Soil Moisture
Downscaling
Random Forest

ABSTRACT

Downscaling of soil moisture data is important for high resolution hydrological modeling. Most downscaling studies in the literature have used spatially discontinuous land surface temperature (LST) maps as the main auxiliary parameter, which limits the creation of continuous soil moisture maps. The number of studies on soil moisture downscaling with machine learning that use gapless LST maps is limited. With this motivation, a hybrid reconstruction method has been proposed in this study to practically obtain continuous LST maps, which are then used to produce high resolution surface soil moisture (SSM) datasets. The proposed method is shown to have high mean performance with R^2 and RMSE values of 0.94 and 1.84°K, respectively, for the period between 2019 and 2022. The developed reconstructed LST maps were then used to downscale original 9 km spatial resolution soil moisture datasets of SMAP L3 and SMAP L4 with Random Forest (RF) machine learning algorithm. The RF model were run with four different rainfall datasets, and the MSWEP rainfall dataset was found to produce the best results. The use of antecedent rainfall values as input variables in machine learning models has been shown to improve the performance of the models R^2 0.76 to 0.93. The accuracy of the downscaled data was later evaluated for Western Anatolia Basins (WAB) in Türkiye with 31 in-situ stations. The downscaled SMAP L4 had good average statistical indicators R (0.815 ± 0.1), RMSE (0.09 ± 0.047 cm^3/cm^3), and ubRMSE (0.058 ± 0.025 cm^3/cm^3). Downscaled SMAP L3 was also validated with in-situ observations with satisfactory R (0.79 ± 0.074), RMSE (0.09 ± 0.043 cm^3/cm^3), and ubRMSE (0.06 ± 0.026 cm^3/cm^3) statistics. Furthermore, the performance of the downscaled SMAP L3 was also cross validated with SMAP + Sentinel 1 (L2) dataset between 2019 and 2022. The mean statistics of R (0.761 ± 0.11) and Root Mean Squared Difference (RMSD) (0.05 ± 0.014 cm^3/cm^3) between downscaled SMAP L3 and L2 data revealed that the new reconstruction method of LST used in the RF model for downscaling of soil moisture performed well to obtain high resolution soil moisture datasets. The proposed technique also overcame the difficulties associated with coastal regions where data was masked for quality considerations, by not only enhancing overall spatial resolution but also filling these data gaps and giving a complete SSM coverage.

1. Introduction

Soil moisture is a key hydrologic variable controling infiltration, surface runoff, and evapotranspiration in the vadose zone. It is important for weather forecasting, drought and flood monitoring, agricultural production planning, and hydrological modeling. For these reasons, it is defined as a fundamental climate variable by the Global Climate Observing System (Dorigo et al., 2011; Huang et al., 2022b). The use of microwave remote sensing methods that are based on the principle of measuring the dielectric properties of soil has been proven to be successful in SSM measurements in bare lands or vegetated areas with the help of various low-frequency radio bands (i.e., X, C, and L) (Mohanty et al., 2017; Li et al., 2021). Passive microwave sensors capture the land surface's brightness temperature, enable the monitoring of spatial and temporal variations in land surface properties. Active microwave sensors transmit microwave signals and measure the backscatter fraction, resulting in soil moisture data with finer spatial resolution compared to passive sensors (Peng et al., 2021; Tavakol et al., 2021). The L-band (1–2 GHz) passive microwave is the promising approach for global SSM measurement with frequent revisit times and

* Corresponding author.
E-mail addresses: onursahin@iyte.edu.tr (O. Güngör Şahin), orhangunduz@iyte.edu.tr (O. Gündüz).

https://doi.org/10.1016/j.jhydrol.2024.131051
Received 26 November 2023; Received in revised form 28 January 2024; Accepted 25 February 2024
Available online 11 March 2024

minimal impacts of atmosphere, surface roughness and vegetation on received signals (Wigneron et al., 2017).

The coarse resolution of the passive microwave sensing or modelled soil moisture products (9–50 km) limits to conduct hydro-meteorological studies at basin and regional scales (Entekhabi et al., 2014; Zhao et al., 2018). Therefore, alternative techniques such as machine learning (Fuentes et al., 2022, Abowarda et al., 2021), HydroBlocks (Vergopolan et al., 2020) or Thermal Inertia Theorem (Fang et al., 2022) are used to produce high (1 km) or hyper (30 m) spatial resolution surface soil moisture maps. Vergopolan et al. (2021) used hyper resolution land surface model and Random Forest (RF) model to predict maize yield prediction. They concluded soil moisture is the most influential and reliable indicator for prediction of crop yield. Fang et al. (2021) used 1 km downscaled data for calculation of two different drought indices for drought monitoring over Australia. Dari et al. (2021) used high resolution soil datasets for detecting and mapping irrigated activities in Mediterranean region.

One of the universally used methods to obtain high resolution downscaled soil moisture datasets are machine learning techniques. These algorithms are capable to learn complex and nonlinear relationships between soil moisture and other hydro-meteorological parameters. The majority of downscaling studies used LST as the main auxiliary variable. These studies can be divided into two main categories based on how LST data is used in the downscaling process. In the first category, LST datasets are directly used in the machine learning models. Here, spatial discontinuities in LST maps can occur due to a variety of factors, one of which is the cloud density that negatively affects thermal infrared measurements (Wu et al., 2021). Alternatively, in the second category, LST maps are reconstructed to avoid any discontinuities in the analysis domain. The gaps in remotely sensed LST data can be filled by a variety of techniques, which can be broadly be classified into four categories: a) daily merging method that utilizes data from different satellite observations (Crosson et al., 2012), b) correlation method where empirical relationship between reference data with other auxiliary variables are derived (F. Xu et al., 2022) c) spatiotemporal gap filling method, within which the temporal, spatial, or spatiotemporal information of the dataset are used to fill the gaps (Gerber et al., 2018; Li et al., 2018), and d) a hybrid usage of first three category (Ma et al., 2022; Zhang et al., 2023). Globally reconstructed LST maps by different techniques are presented in many studies (Shiff et al., 2021; Yu et al., 2022; T. Zhang et al., 2022), where these databases cover certain time periods, and they may not meet user requirements.

Each LST reconstruction methods have tradeoffs. For example, the daily merging scheme is efficient but it cannot fully fill LST gaps. Temporal interpolation method is fast and easy to implement but can sometimes be inaccurate (Wu et al., 2021). Spatial interpolation technique can be easily used to fill in gaps, but it may lead to accuracy problems. Spatiotemporal interpolation, on the other hand, can be used to increase accuracy, but is more complex and computationally costly than other methods (Mo et al., 2021).

During the downscaling process, it is possible to train models with spatially and temporally discontinuous maps, but then the resulting soil moisture product also becomes discontinuous, which will likely to limit its use. Therefore, the main challenge is to produce gapless soil moisture maps using machine learning models. M. Xu et al., (2022) used wide & deep learning method to downscale SMAP for Continental United States and they concluded that downscaled dataset is discontinuous and future studies should be focused to construct continuous LST auxiliary variables.

Based on these premises, reconstruction of LST for use as the primary auxiliary variable in the downscaling methodology is essential for generating high-resolution soil moisture data. Accordingly, this study aims to develop an accurate and practically applicable reconstruction method for LST. While Moderate Resolution Imaging Spectro-radiometer (MODIS) MOD11 Aqua and Terra LST datasets are commonly used as LST data sources in the literature (Wan et al., 2021), the Visible Infrared Imaging Radiometer Suite (VIIRS) VNP21 LST dataset offers superior temporal and spatial continuity compared to MODIS products. These two datasets can be merged using the daily merging technique. Consequently, the MODIS (MOD21, MYD21) and VIIIRS (VNP21) datasets were merged in this study using the daily merging method. As the daily merging scheme cannot fill all the gaps in the data, model-based LST datasets can be used as additional tools for filling gaps in IR-based LST data. Accordingly, model-based LST data was used in this study to fill the remaining gaps in the LST dataset. Finally, gapless LST data generated by the proposed method was employed as an auxiliary parameter to downscale the SMAP L3 and L4 datasets to a spatial resolution of 1 km.

According to the scope mentioned above, the novelties of this study can be summarized as follows:

- A new hybrid algorithm is proposed in this study to provide seamless Night (ascending) LST maps. First, a daily merging method is applied using MODIS and VIIRS to fill the all the gaps in the data. Then, a model based dataset (CLSM) is used to fill remain gaps in the reference LST dataset.
- Reconstructed LST maps are then used with other auxiliary variables in the Random Forest method to downscale SMAP L3 and SMAP L4 surface soil moisture data and to create high spatial resolution (1 km) gapless surface soil moisture datasets.
- Different gridded precipitation products can be used for training RF models. In this study, the performance of four precipitation datasets has been compared in generating soil moisture maps by comparing RF test results. We also performed tests to analyze the performance of machine learning method when different combinations of previous $(P_{t-1}, P_{t-2} \cdots P_{t-n})$ precipitation events are used as auxiliary variables in the RF model.
- Then, a validation study is conducted to test the accuracy of the final outputs. A sparse network analysis with only one-point scale measurement within each corresponding pixel was utilized to determine the accuracy of the original, downscaled SMAP L4 and downscaled SMAP L3 datasets in Western Anatolia (Türkiye) using 31 in-situ stations.
- SMAP + Sentinel 1 (L2) dataset is capable of producing high-resolution (1 km or 3 km) soil moisture measurements with good accuracy, but low temporal resolution. Hence, our final downscaled L3 products were temporally and spatially validated with SMAP + Sentinel 1 (L2) dataset in addition to in-situ validation.

2. Study area

Western Anatolia Basin (WAB) consist of the four hydrological river basins Küçük Menderes (7060 km^2), Büyük Menderes (26133 km^2), Gediz (17034 km^2) and Kuzey Ege (9974 km^2) located in the Aegean Region between 37.00° − 40.00° N latitudes and 26.00°-30.30° E longitudes. WAB covers an area of approximately 60,000 km^2 (Fig. 1). The climate conditions in these basins are predominantly of the Mediterranean type with dry and hot summers and cold and rainy winters. The annual average rainfall in the region varies between 450 and 1200 mm (Khorrami et al., 2023). According to the ESRI land use data of 2022, the ratio of agricultural land use to total basin area in the Büyük Menderes, Gediz, Kuzey Ege and Küçük Menderes basins is 27.9 %, 33.0 %, 22.1 % and 22.0 %, respectively (Karra et al., 2021). Agricultural activities are quite high in all four basins, especially in the Gediz basin, where soil moisture is a determining factor in agricultural production.

3. Data used

MOD21 (Aqua), MYD21 (Terra), and VNP21 (VIIRS) LST datasets were utilized to reconstruct and generate daily gapless LST maps. Precipitation, LST, soil properties, NDVI, EVI, elevation, and slope variables were later used as auxiliary variables for training machine learning

Fig. 1. Geographical location of Western Anatolia Basins (WAB): Küçük Menderes, Büyük Menderes, Gediz, and Kuzey Ege basins and soil moisture measurement network in the study area.

model and producing high resolution surface soil moisture maps. Four precipitation datasets (CHIRPS, MSWEP, IMERG, SM2RAIN) were further analyzed to find out which product train the machine learning model better. In-situ measurements (MEVBIS) collected from the State Meteorological Organization of Türkiye were used for the validation of downscaled SSM maps. A summary of the datasets used in this study are given in Table 1.

3.1. Land surface temperature

The Visible Infrared Imaging Radiometer Suite (VIIRS) Land Surface Temperature and Emissivity (LST&E) algorithm and data products (VNP21) have been developed by NASA in collaboration with the MODIS. The overall goal of the NASA VIIRS is to make the algorithms and products compatible with the C6 MODIS Terra and Aqua to ensure data product continuity (Hulley and Hook, 2018). VIIRS collects global satellite observations across the visible and infrared wavelengths. It has 22 channels ranging from 0.41 μm to 12.01 μm, with 5 high-resolution, and 16 moderate-resolution bands. The VNP21 dataset is available for download from https://viirsland.gsfc.nasa.gov/Products/NASA/LSTESDR.html.

Terra and Aqua satellites are the two main parts of the MODIS missions, which were launched in 1999 and 2002, respectively. Two satellites collect earth observation data from multi-spectral bands consisting of Thermal Infrared (TIR) and Passive Microwave (Hulley & Ghent, 2019). The MOD21 is available for download from https://modis.gsfc.nasa.gov/data/dataprod/mod21.php.

3.2. SMAP surface soil moisture

The NASA Soil Moisture Active Passive (SMAP) mission launched in 2015 uses an L-band radar and an L-band radiometer for integrated measurements as a single observation system (Entekhabi et al., 2014). The accuracy requirement is determined as unbiased root mean square error $< 0.04\ m^3/m^3$. Comparative analyses of SMAP surface soil moisture datasets are prepared at three levels (i.e., L2, L3, and L4) and it is confirmed that these datasets produce data compatible with Core Validation Sites (CVS) and sparse networks (Colliander et al., 2017; Colliander et al., 2022). Different level SMAP products are available for download from https://nsidc.org/data/smap/data.

3.3. Soil properties

SoilGrids250m is a dataset of soil properties, including sand, silt, clay content, pH, etc., at six different depths. More than 200,000 soil profile data from around the world are used to train the machine learning model that was validated with 10 different cross-validations (Hengl et al., 2017). The soil properties used in this study were obtained from the SoilGrids250m dataset. Soil properties were downloaded from https://soilgrids.org/.

3.4. Precipitation

Precipitation is the main hydro-meteorological parameter that soil moisture is dependent on. Four different gridded precipitation datasets were used in the downscaling model as auxiliary variables.

Table 1
All satellite-based, model-based, and ground-based datasets used in this study.

Data name	Dataset	Spatial Resolution	Temporal Resolution	Reference
Surface Soil Moisture	SMAP L4	9 km	3 hourly	(Reichle et al., 2017)
Surface Soil Moisture	SMAP L3	9 km	1–2 days	(Entekhabi et al., 2016)
Surface Soil Moisture	SMAP L2	1–3 km	2–6 days, Europe 6–12 days, Global	(Das et al., 2018)
Land Surface Temperature (LST)	CLSM	9 km	3 hourly	(Reichle et al., 2017)
Land Surface Temperature (LST) (MOD21, MYD21), VNP21	MODIS, VIIRS	1 km	Daily	(Hulley et al., 2016) (Hulley, 2018)
Normalized Vegetation Index (NDVI), Enhanced Vegetation Index (EVI)	VIIRS	1 km	Monthly	(Didan, 2018)
Precipitation	CHIRPS	0.05°	Daily	(Funk et al., 2015)
Precipitation	MSWEP	0.1°	Daily	(Beck et al., 2019b)
Precipitation	SM2RAIN	0.1°	Daily	(Brocca et al., 2019)
Precipitation	IMERG	0.1°	Daily	(Huffman et al., 2019)
Soil Properties	SoilGrids	250 *m*	Static	(Hengl et al., 2017)
Elevation and Slope (DEM, S)	SRTM	90 *m*	Static	(Farr et al., 2007)
Soil Moisture at 20 cm (in situ)	MEVBIS	Point	Hourly	–
Soil Temperature at 20 cm (in situ)	MEVBIS	Point	Hourly	–
Precipitation (in situ)	MEVBIS	Point	Daily	–

3.4.1. Multi-source weighted-ensemble precipitation (MSWEP)

Multi-Source Weighted-Ensemble Precipitation (MSWEP) is a global precipitation dataset that combines gauge, satellite, and reanalysis data for hydrological modeling (Beck et al., 2017). MSWEP data are available from 1979 to ~ 3 h from real-time at http://www.gloh2o.org.

3.4.2. Climate hazards group infrared precipitation with station data (CHIRPS)

Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) is a quasi – global precipitation dataset that are widely used in the downscaling studies. In regions with limited access to ground-based precipitation measurements, CHIRPS serves as a valuable alternative data source The CHIRPS dataset is available for download from http://chg.geog.ucsb.edu/data/chirps/.

3.4.3. Soil moisture to rain (SM2RAIN)

Soil Moisture to Rain (SM2RAIN) dataset is a relatively new gridded precipitation product (Brocca et al., 2019). that inversely use soil moisture balance equation for prediction of precipitation based on a "bottom-up" approach. The SM2RAIN datasets are accessible at http://hydrology.irpi.cnr.it/download-area/sm2rain-data-sets/.

3.4.4. Integrated multi-satellite retrievals for the global precipitation measurement mission (IMERG)

Amjad et al. (2020) compared the validity of several precipitation datasets for Türkiye and they concluded that Integrated Multi-Satellite Retrievals for the Global Precipitation Measurement Mission (IMERG) has good accuracy. The latest release of the IMERG data is accessible at https://disc.gsfc.nasa.gov/datasets/GPM_3IMERGHH_06/summary.

3.5. Vegetation parameters

VIIRS mission produces three different vegetation indexes with 8-day, 16 days, and monthly intervals. Two vegetation indexes used are the Normalized Difference Vegetation Index (NDVI), one of the longest remote sensing based time series data records, and the Enhanced Vegetation Index (EVI), which is a more sensitive and accurate measure of vegetation health because it accounts for variations in canopy structure and atmospheric interference (Didan, 2018). NDVI and EVI datasets are available at https://lpdaac.usgs.gov/products/vnp13a3v001/.

3.6. Digital elevation model (DEM) and slope

The elevation data is crucial for variation of the soil moisture, and it was downloaded from the NASA Shuttle Radar Topography Mission (SRTM) (https://srtm.csi.cgiar.org/). DEM data was resampled to 1 km resolution and slope map were produced using Geographical Information System (GIS) analysis. The planar method within ArcGIS software is used to determine slope based on the maximum elevation difference between a cell and its surrounding eight neighbors.

3.7. In-situ soil moisture

Ground-based soil moisture data are measured using the Campbell Scientific CS616 (Campbell Scientific, 2020) reflectometer that determines the volumetric soil moisture at 20 cm. The quality of the sensor data is strongly affected by parameters such as electrical conductivity and soil temperature. For this reason, the raw measurements must be corrected with the soil temperature values at the same depth as described in Campbell Scientific (2020) and Bulut et al. (2019). Hourly precipitation, soil temperature and soil moisture at 20 cm obtained from the State Meteorological Service (MGM) of Türkiye through the online platform MEVBIS for a period of two years (2021–2022) for 39 ground stations in the study area.

The visual inspection is typically a good option for quality control of moisture data (Dorigo et al., 2013). SSM has a predictable response to precipitation. Following a recorded precipitation event, corresponding increase in soil moisture measurements is expected. Conversely, during dry periods, a gradual and continuous decline in soil moisture should be observed (Bulut et al., 2019). Therefore, the compatibility of the average daily moisture data with the rainfall data measured from the same stations is visually examined. An example of visual analysis of precipitation and soil moisture data is given in the Fig. 2. Upon visual inspection, 8 of these stations were removed from further analysis due to inconsistent and irrational response patterns.

4. Methodology

In this study, we propose to utilize various LST datasets to reconstruct and generate daily gapless LST maps, which can then be used together with additional auxiliary variables to downscale SMAP L3 and SMAP L4 data. All auxiliary variables are resampled to resolution of SMAP (9 km) and RF model is trained for both SMAP L3 and SMAP L4. Afterwards, high spatial resolution SSM maps are produced, which are resampled to 1 km resolution. In-situ measurements are then used to validate the downscaled SMAP L4 and SMAP L3 datasets. Additionally, SMAP L2 data is also used to further validate the downscaled SMAP L3 dataset. The proposed LST reconstruction method and the soil moisture

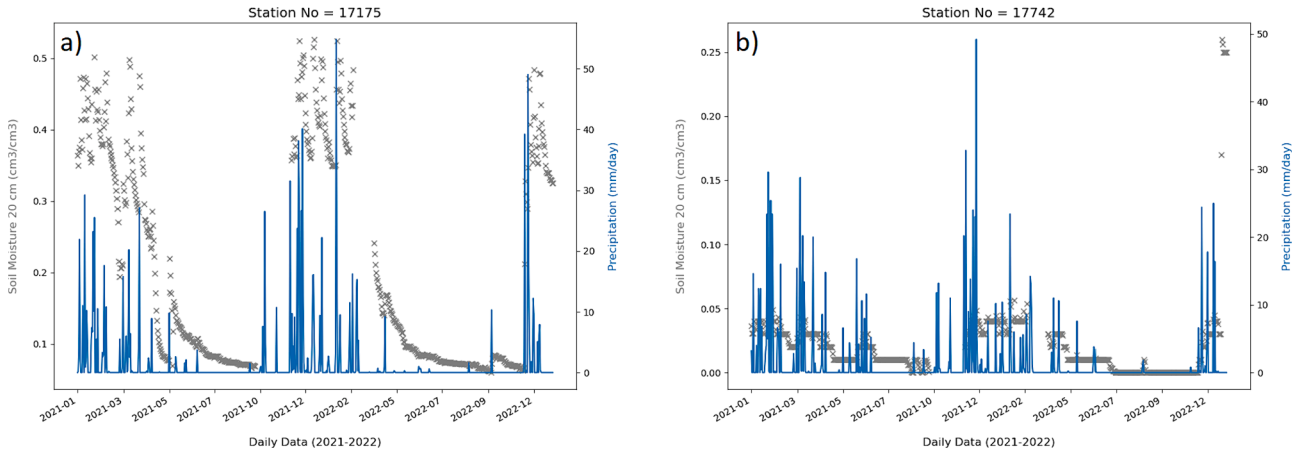


Fig. 2. Sensor measurements of soil moisture corrected with soil temperature data at 20 cm versus precipitation shown for a) reliable station, b) unreliable station.

downscaling procedure applied in this study are schematically shown in Fig. 3.

4.1. Reconstruction of land surface temperature

The proposed reconstruction method consists of three stages. The first and initial step is data processing to assess the quality of the LST products as defined in the section 4.1.1. The second step is the daily merging step that merges VNP21 (T1), MYD21 (T2), and MOD21 (T3) to Night LST (T4) datasets (Section 4.1.2). In the last step, remaining gaps are filled with model-based LST datasets (Section 4.1.3). Fig. 4 illustrates the steps on how VNP Night product can be filled by using the proposed hybrid method.

4.1.1. Data preprocessing

The initial step of the analysis is the quality assessments of MOD21, MYD21 and VNP products. This includes a data control step to label LST accuracy as; >2°K (Poor performance), 1–1.5°K (Good performance), 1.5 − 2°K (Marginal performance), and < 1°K (Excellent performance). Later, poor performance pixels are removed from the dataset. Furthermore, physically unreasonable values are also eliminated from the dataset such that the minimum and maximum LST values beyond −40 °C and 65 °C would set the usable pixel values, respectively. The remaining pixels are further processed in the next steps.

4.1.2. Daily merging scheme

Several research conducted the daily merge scheme by using four daily LST observations from Aqua and Terra (Crosson et al., 2012; Li et al., 2018; T. Zhang et al., 2022). In this study, a new algorithm is implemented to merge three daily night observations from the VIIRS, Aqua, and Terra to produce a daily merged (T4) Mid – Night LST product as shown in Fig. 4. In this new method, VNP Night (T1) product is selected as the reference data. Then, valid pixel percentage of yearly time series data for each pixel is calculated for all observations. If T1 is available for the pixel and valid percentage of T1 is larger than 10 %, then that pixel (T4) will be filled directly using the value of T1. Otherwise, other products T2 and T3 will be selected respectively as independent variables with T1 dataset being the dependent variable to build linear regressions sequentially (T1-T2, T1-T3). The slope and bias of the linear function between the T1-T2 or T1-T3 are sequentially calculated

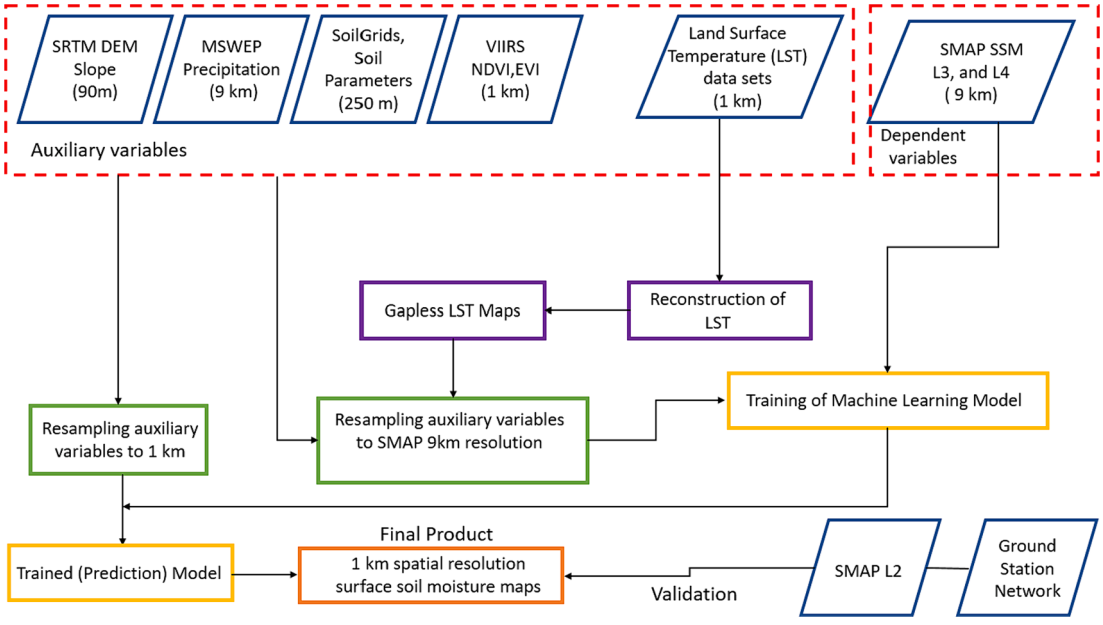


Fig. 3. Flowchart of LST reconstruction and soil moisture downscaling procedure. LST maps are reconstructed to fill in gaps and later used as the primary auxiliary variable of the soil moisture downscaling algorithm. The random forest model is first trained and validated. Later, high resolution soil moisture maps are created for WAB region using the trained model. These maps are validated using sparse soil moisture sensors and SMAP + Sentinel L2 high resolution data.

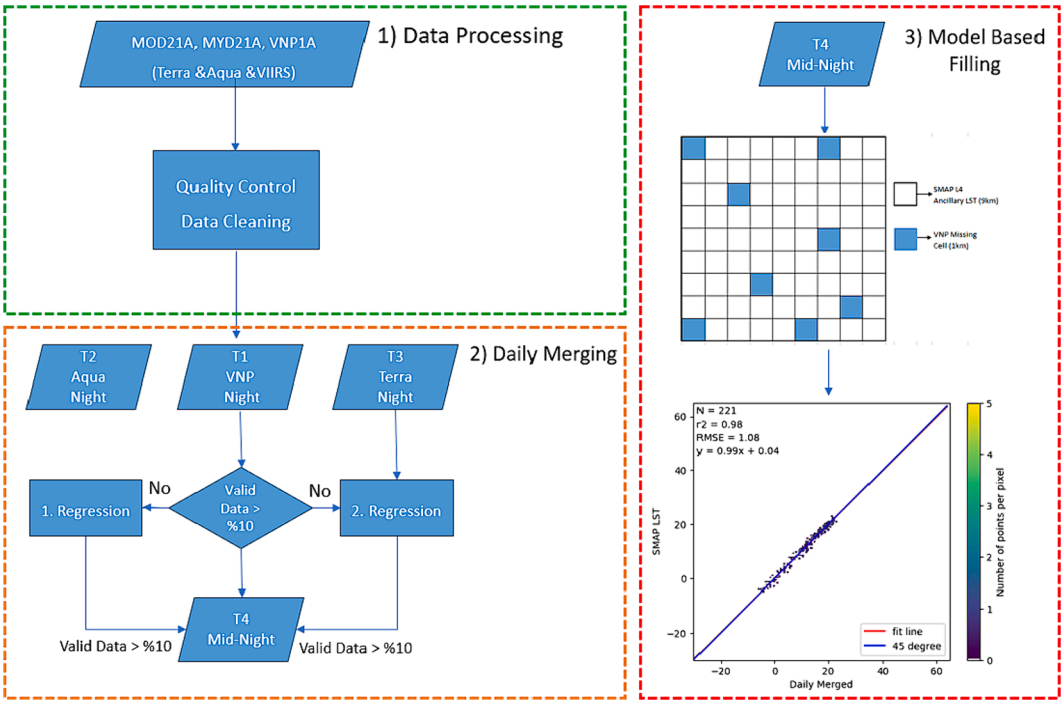


Fig. 4. Reconstruction of VNP Night LST using daily merged + model-based hybrid method. The proposed algorithm includes three steps: 1) Data processing, quality control and data clean up, 2) Daily merging of the satellite data (VNP21, MYD21, and MOD21) and 3) Model-based gap filling. Seasonal linear regression fitted between daily merged (T4) and SMAP L4 data are used as final step for gap-filling.

using the array programming library (Harris et al., 2020) in the Python programming language. In the final step of the procedure, the missing pixel in the T4 will be filled using the T2 or T3 respectively via a fitted linear regression model.

4.1.3. *Model based filling*

The GEOS-5 Catchment land surface model is enhanced with radiative transfer model for energy transport process. Surface soil moisture capabilities of SMAP are extended to the root zone (0–100) using the Catchment land surface model capabilities and GEOS-5 ensemble-based land data assimilation system. SMAP L4 ancillary variables have continuous 3 h temporal and 9 km spatial resolution (Reichle et al., 2017). Due to the capabilities of the radiative transfer model, SMAP L4 LST ancillary variable could be a viable alternative to filling the gaps in IR-based LST data. Therefore, we hybridized daily satellite products with model-based SMAP L4 LST product with following steps: i) first we obtained daily merged T4 products with data preprocessing and daily merging scheme, ii) if a pixel was missing in the T4, the missing pixel is matched with the closest SMAP LST pixel, iii) later, a seasonal linear regression model is fitted (slope and bias of the linear function) between T4 and model based LST pixels, iv) finally missing pixels are filled with the available corrected SMAP LST for obtaining final gapless daily LST maps.

4.1.4. *Validation of reconstructed LST*

The typical procedure of validating the reconstructed LST maps is done by generating artificial gaps in the base LST product to evaluate performance of seamless LST maps that are created using different reconstruction algorithms (T. Zhang et al., 2022). Accordingly, we randomly selected 15 days from each year between the period 2019–2022 and randomly excluded 50 % of LST data from each selected day. Excluded data points were then used to validate the performance of the proposed reconstruction algorithm using statistical indicators R and RMSE shown in Section 4.5.

4.2. *Selection of auxiliary parameters*

The selection of auxiliary variables of the RF model is the key aspect of the analysis. The model typically uses a combination of static and dynamic variables. Static variables are time independent parameters such as soil texture and morphologic characteristics while dynamic variables are time dependent parameters such as temperature and precipitation. For the use of static parameters as auxiliary variables, Karthikeyan and Mishra (2021) concluded that SSM exhibits a strong positive correlation with the percentage of clay and silt in the soil texture, likely due to their higher moisture holding capacity. Conversely, a negative correlation is observed with sand percentage, as it has lower porosity. Consequently, soil texture distribution emerges as a critical factor in SSM prediction. Also, other soil properties, such as soil texture and bulk density, have a major influence on the spatial distribution of soil moisture. As a result, these properties are often used as static auxiliary datasets in soil moisture downscaling studies (M. Xu et al., 2022). Liu et al. (2020) discussed that over areas with significant elevation changes, a potential coupling might exist between DEM and SSM suggesting the potential usefulness of DEM in downscaling soil moisture data. Furthermore, slope of the topography dictates the infiltration potential and subsurface flow character of the soil, which can be an additional predictor to determine SSM. Accordingly, researchers have extensively employed static parameters that are influential on soil moisture as auxiliary variables as shown in Table 2.

The incorporation of time dependent auxiliary variables into the analysis provides the prediction capability of the RF model in a time dependent manner. Since soil moisture is highly dependent on temporally variable conditions such as surface temperature, precipitation input, evapotranspiration etc., a suitable subset of these dynamic parameters need to be included in the prediction model. Since the main objective of this study is the production of spatially gapless SSM maps, LST and precipitation are the most preferred parameters often used in the literature (Table 2). LST maps are reconstructed defined as in the section 4.1. Although surface reflectance (SR) and albedo (ALB) could have been used as valuable auxiliary parameters, they also need to be

Table 2
Literature review summary of downscaling studies for selection of auxiliary parameters and machine learning model.

Study Area	Method*	Auxiliary Variables **	Reference
İber Island (Spain)	RF	LST, LAI, NDVI, EVI, NDWI, DEM, ALB	(Zhao et al., 2018)
Haihe Basin (China)	RF	LST, SR, NDVI, P, BT	(Abowarda et al., 2021)
Hebei Region (China)	RF	LST, SR, P, NDVI	(Long et al., 2019)
Southwest France	RF, CART, GBDT, XGB	LST, NDVI, EVI, ALB, DEM, ET	(Liu et al., 2020)
Australia	LSTM	LST, SR, NDVI, P, SP	(Fuentes et al., 2022)
Mongolia	RF	LST, NDVI, EVI, DEM	(Hu et al., 2020)
Southernwest USA	DBN	LST, SR, P, DEM,SP	(Huang et al., 2022b)
Continental America	XGB	LST, NDVI, EVI, DEM, P, SP	(Karthikeyan and Mishra, 2021)
Continental America	WDL	LST, SR, P, SP	(F. Xu et al., 2022)
Pearl River Basin (China)	LSTM	LST, ALB, P, EVI, DEM	(Y. Zhang et al., 2022)

*RF – Random Forest, CART- Classification and regression tree, GBDT- Gradient Boosted Decision Tree, XGB – Extreme Gradient Boosted Regression Tree, LSTM – Long Short Term Memory, WDL- Wide Deep Learning, DBN – Deep Belief Learning, ** LST – Land Surface Temperature, LAI – Leaf Area Index, NDVI- Normalized Difference Vegetation Index, EVI – Enhanced Vegetation Index, NDWI- Normalized Difference Water Index, DEM- Digital Elevation Map, ALB- Albedo, SR- Surface Reflectance, P- Precipitation, SP- Soil Properties, BT- Brightness Temperature, ET-Evapotranspiration

reconstructed before use. Furthermore, vegetation related parameters such as NDVI and EVI are also commonly used in the literature as dynamic parameters reflecting hydro-meteorological conditions of the area (Table 2).

4.3. Machine learning model

A general functional representation of downscaling procedure implemented in this study is shown in Equation (1) based on a thorough analysis of the literature (Table 2). Accordingly, the SSM is considered a function of night (ascending) land surface temperature (LST_N), normalized difference vegetation index ($NDVI$), enhanced vegetation index (EVI), precipitation (P), clay, silt, sand contents of the soil (Cl, Si, Sa, respectively), the soil organic carbon content (soc), pH of water ($phh2o$), bulk density of soil ($bdod$) volumetric water content of soil at 10 kPa ($wv0010$), digital topographical elevation (DEM), and topographical slope (S):

$$SSM = f(LST_N, NDVI, EVI, P, Cl, Si, Sa, soc, phh2o, bdod, wv0010, DEM, S) \quad (1)$$

This concept is independent from the type of machine learning model used in the application. The parameters inside the function are independent auxiliary variables that are accepted to be significant in downscaling the soil moisture. The dependent variable used in the model is soil moisture value from SMAP.

In this study, Random Forest algorithm is used as the machine learning technique to downscale the soil moisture data. RF is the most widely used decision tree technique in the literature, which provides good accuracy and successful results (Zhao et al., 2021; Abowarda et al., 2021; Long et al., 2019). RF is first used by Breiman (2001) as a multiple decision tree model that produces optimal estimates by averaging the output of a large number of regression trees. The RF method is implemented in three stages to produce optimal estimates of the dependent variable. In the first stage, all data is divided into subsets and each subset is used to create a single decision tree model. Then, an output is obtained from the created subset model. In the final stage, the output estimates are generated by averaging the results of the models created from different decision trees. In this study, the number of trees was set to 250 and 80 % of the data set was used for training purposes. The remaining 20 % was reserved for testing the model. Cross validation procedure was conducted by randomly selecting training and test data five times.

4.4. Analysis of precipitation

Precipitation is the main driver of soil moisture. Previous research used only the precipitation value corresponding to day the soil moisture is forecasted. However, it is known that the soil moisture conditions are not only related to the precipitation event on that particular day, but also on antecedent precipitation events from previous days (P_{t-1}, $P_{t-2}\cdots P_{t-n}$). Therefore, we performed tests to analyze the performance of machine learning method when precipitation from previous days (P_{t-n}) are included as additional auxiliary variables in addition to the particular day's precipitation value (P_t).

Currently, numerous precipitation products are available for use in the RF model. We compared the suitability of four precipitation datasets (CHIRPS, MSWEP, IMERG, SM2RAIN) by comparing machine learning test results. The product that gives the highest performance for the study area is used in further steps of the downscaling analysis. The performance of the datasets was checked by using the R and RMSE statistics presented in Section 4.5.

4.5. Downscaling and validation

The model-based SMAP L4 and SMAP L3 soil moisture datasets were downscaled using auxiliary variables and the RF machine learning technique (Equation (1) by using the flowchart given in Fig. 3. There are two daily observations for L3 data and the mean of these observations are used in this study. For the WAB region, the performance of the original SMAP and downscaled SMAP datasets were compared and validated using ground observations recorded at 31 in-situ stations. In addition, the performance of downscaled SMAP L3 data was also compared with SMAP L2 product that is prepared by joint use of SMAP and Sentinel 1 datasets, yielding a high-resolution (1 km) soil moisture product with good accuracy, despite the fact that its temporal resolution is coarser (i.e., 6 day over Europe and 12 days over the world) (Das et al., 2019). Colliander et al. (2022) validated 3 km resolution SMAP L2 dataset using core validation sites (CVS). One of the advantages of the using high resolution spatial data for validation is that it provides a means for directly validating the downscaled data on a spatial basis. On the other hand, ground station comparison is only possible for small number of pixels distributed over a large area that directly match with ground stations.

4.6. Statistical evaluation

Correlation Coefficient (R), Root Mean Square Error (RMSE), and Unbiased Root Mean Square Error (ubRMSE) metrics given in Equations (2)–(4) are used to test the performance of the developed methodology. In the equations, r and o represent the remotely sensed and ground-based datasets, respectively and n corresponds to the number of data. As the R value approaches 1, it is concluded that the remote sensing data produces results that are consistent with ground observation stations. Similarly, the RMSE and ubRMSE values indicate that the remotely sensed data and ground-based data are more consistent when the corresponding metric approaches to 0.

$$R = \frac{\sum_{i=1}^{n}(r_i - \overline{r})(o_i - \overline{o})}{\sqrt{\sum_{i=1}^{n}(r_i - \overline{r})^2}\sqrt{\sum_{i=1}^{n}(o_i - \overline{o})^2}} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i - o_i)^2} \quad (3)$$

$$ubRMSE = \sqrt{RMSE^2 - \left[\frac{1}{n}\sum_{i=1}^{n}(r_i - o_i)\right]^2} \quad (4)$$

5. Results

5.1. Reconstruction of LST

Once the mean valid pixel percentages are evaluated for the study area, day products were found to have similar percentages, but VIIRS night product had nearly 1.33 and 1.77 times more valid pixel than Terra and Aqua products, respectively in 2022 (Table 3). Thus, VNP Night product was reconstructed by the proposed method given in Section 4.1 (Fig. 4).

Statistical performance indicators for the evaluation of proposed reconstruction algorithm are given in Table 4. Accordingly, the final Daily Merged + SMAP product mean performances were 0.94 (R^2) and 1.84°K (RMSE), respectively. The performance of model-based filling varied between 0.86 and 0.94 (R^2) and 2.05–2.64°K (RMSE). Scatter plots of the Daily Merged, SMAP and Final Daily Merged + SMAP are given in Fig. 5 for 2022. Mean Valid Pixel percentage of VNP (%68.37) increased to 79.37 % and 98.62 % after daily merging scheme and the model-based filling steps, respectively (Table 5). The daily merging algorithm can only fill 10–15 % of the gaps and thus leave missing pixels. The remaining gaps should be filled with a model-based or another reconstruction method. The model-based reconstruction method used in this study showed relatively good performance and, gapless LST maps can be produced with this two-step reconstruction technique with R^2 and RMSE ranges of 0.93–0.95 and 1.77–2.02°K, respectively.

5.2. Machine learning model test results with precipitation datasets

The effectiveness of using precipitation values from previous days were tested on the prediction capabilities of machine learning models on the soil moisture prediction. Different previous precipitation values P_{t-1}, $P_{t-2}\cdots P_{t-n}$ are used as additional auxiliary variables for both SMAP L3 and SMAP L4 products. The parameter "n" controls how many previous days from today the rainfall values are used as auxiliary variables. For example, when n = 6, the model will use the rainfall values from the past 6 days to predict soil moisture and when n = 0, it will only use the corresponding day's precipitation (no antecedent precipitation) value as an auxiliary variable. For the SMAP L3, the performance of machine learning model for several previous precipitation values on soil moisture are given in Fig. 6. R^2 values of 0.760, 0.876, 0.904, and 0.931 were obtained for n = 0, 3, 6, and 13, respectively. There was no improvement in statistical indicators in test results produced for n > 13. It can be said that the rainfall values of the previous 2 weeks have an effect on surface soil moisture for WAB region. Therefore, it can be concluded that, for the SMAP L3, antecedent rainfall has a significant impact on soil moisture and better predictions are obtained by incorporating the precipitation values of higher number of previous days for MSWEP dataset. For SMAP L4, on the other hand, only the LST Night variable has major influence for downscaling and it can be said that antecedent precipitation values have no major effect to improve test results.

Furthermore, the performances of the other three precipitation products have been evaluated for n = 0 and n = 13. The scatter plots of the results are given in Fig. 7 for n = 13. The R^2 and RMSE values were calculated as 0.815 and 0.0322 cm^3/cm^3 for CHIRPS, 0.786 and 0.0347 cm^3/cm^3 for IMERG, 0.907 and 0.0218 cm^3/cm^3 for SM2RAIN, respectively. The results revealed that antecedent rainfall has relatively smaller impact on soil moisture for this three precipitation product when compared to MSWEP. In all different scenarios, MSWEP outperforms other precipitation datasets when the test results are compared.

5.3. Validation of downscaled SMAP L3 and SMAP L4 with in-situ measurements and SMAP + Sentinel L2

For original SMAP L4, the R statistic varied from 0.664 to 0.977, with an average of 0.88 ± 0.074. RMSE and ubRMSE also varied from 0.017 to 0.26 cm^3/cm^3 and from 0.013 to 0.13 cm^3/cm^3, respectively, with average values being 0.088 ± 0.057 cm^3/cm^3 and 0.052 ± 0.029 cm^3/cm^3, respectively. For downscaled SMAP L4, mean statistical indicators were calculated to be quite similar with the original data with R, RMSE and ubRMSE values of (0.82 ± 0.1), (0.09 ± 0.047 cm^3/cm^3) and (0.058 ± 0.025 cm^3/cm^3), respectively. There is slight decrease in the R statistic for the downscaled SMAP L4 but RMSE and ubRMSE have similar performance when compared with the original data. Therefore, it can be concluded that the original SMAP L4 dataset is downscaled to high (1 km) spatial resolution with approximately the same statistical performance.

The original and downscaled SMAP L4 maps for the fifteenth days of January, April, July, and October for the year 2021 are shown in Fig. 8 to compare the downscaling performance of SMAP L4 by visually incorporating the effect of seasonality (dry – wet). It can be seen that the original and downscaled SMAP L4 datasets have similar spatial patterns for dry – wet seasons. Moreover, spatial mean values of the original and the downscaled SMAP L4 have been calculated for each day between 2019 and 2022. Maximum and mean differences between these values were calculated to be 0.096 and 0.017, respectively (Fig. 9). Thus, it can be revealed that the spatial difference between the original and downscaled data is minor for WAB region.

For the downscaled SMAP L3, the R, RMSE and ubRMSE statistics were calculated to be (0.79 ± 0.074), (0.09 ± 0.043 cm^3/cm^3), and (0.06 ± 0.026 cm^3/cm^3), respectively. The results of the statistical comparisons of the original and downscaled SMAP products with ground stations are given in Table 6. Downscaled SMAP L3 data has slightly poor performance with in-situ stations when compared with Original and Downscaled SMAP L4 datasets. One reason for this difference could be the moisture sensing depth between SMAP L3 (0–5 cm) and in-situ measurements (20 cm). Nevertheless, the observed discrepancies between the datasets were found to be minor and a comparative analysis of their statistical parameters revealed fairly compatible results.

Moreover, the spatial performance of downscaled SMAP L3 was compared with SMAP + Sentinel (L2) because of the sensing depth differences between SMAP L3 and in-situ datasets. According to the

Table 3
Mean values of passing times and valid pixel percentages for WAB in 2022.

Parameter	Terra (MOD21D)	AQUA (MYD21D)	VIIRS (VNP1D)	Terra (MOD21A)	AQUA (MYD21A)	VIIRS (VNP1A)
	Day	Day	Day	Night	Night	Night
Passing Times (Local Time)	10:30	13:36	12:22	21:22	2:13	2:42
Valid Pixel Percentage (%)	51.0	48.6	53.4	46.5	35.0	***61.9***

Table 4
Statistical performance of proposed LST reconstruction algorithm for the period 2019 – 2022.

Excluded pixels %50	2019		2020		2021		2022	
	R^2	RMSE (°K)	R^2	RMSE (°K)	R^2	RMSE (°K)	R^2	RMSE (°K)
Daily Merged	0.94	1.61	0.95	1.65	0.94	1.9	0.94	1.74
SMAP	0.86	2.59	0.92	2.14	0.9	2.64	0.94	2.05
Daily Merged + SMAP	0.93	1.77	0.95	1.78	0.93	2.02	0.94	1.79

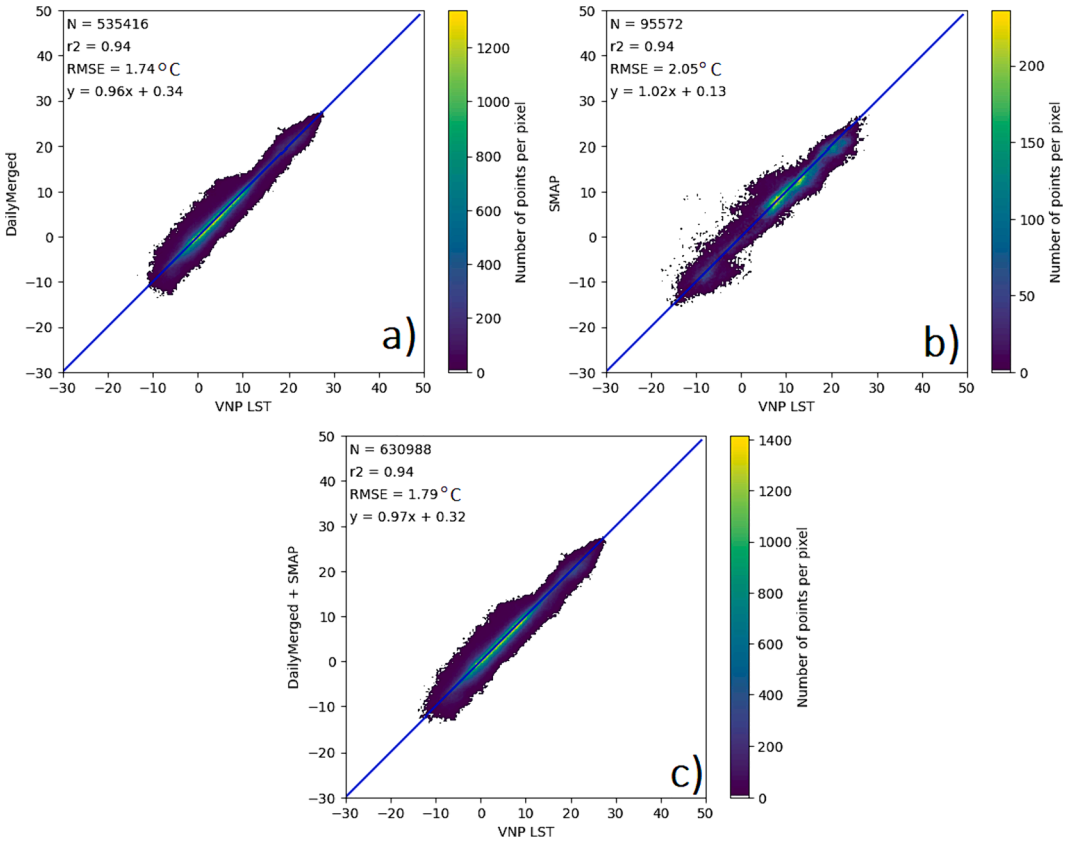


Fig. 5. Validation of reconstructed LST dataset. Performances of a) Daily-Merged Scheme, b) Model based reconstruction, and c) Final product for year 2022.

Table 5
Valid Pixel Percentages after reconstruction steps.

Year	Valid Pixel Percentage		
	VPN Night	Daily Merged	Daily Merged + SMAP
2022	66.57 %	79.06 %	95.00 %
2021	68.50 %	79.30 %	99.79 %
2020	72.66 %	82.62 %	99.83 %
2019	65.73 %	76.48 %	99.87 %
Mean	***68.37 %***	***79.37 %***	***98.62 %***

results, mean statistics of R and Root Mean Squared Difference (RMSD) between downscaled SMAP L3 and L2 data were calculated as (0.761 ± 0.11) and (0.05 ± 0.014 cm^3/cm^3), respectively. It can be concluded that the downscaled SMAP L3 shows similar temporal and spatial trends with the high resolution L2 dataset where SMAP and Sentinel are used together. Generally, decreased R and RMSE values indicated a decline in model performance, particularly at watershed boundaries characterized by complex topography and high altitudes (Fig. 10). Fig. 11 presents the original and downscaled SMAP L3 maps for August 1, 2021. It also illustrates the temporal resolution of the original SMAP dataset. The findings indicate that the SMAP satellite passes over the WAB region 135 times (mean of pixels) out of 365 days. Notably, there is a lack of data for 46 % and 51 % of the total surface areas of Kuzey Ege and Küçük Menderes basins, respectively.

Finally, time series obtained from the downscaled SMAP L3 and L4 datasets are visually compared with in-situ soil moisture and precipitation measurements. Time series for 15 out of the 31 stations are given in Fig. 12. Accordingly, downscaled SMAP L3 were found to be more sensitive to precipitation than the downscaled SMAP L4 but they have a similar temporal pattern with in-situ soil moisture.

6. Discussion

6.1. Performance of proposed reconstruction method

Majority of the downscaling studies in the literature use LST as an auxiliary parameter in downscaling the soil moisture datasets using trained machine learning models. While some studies (Zhao et al., 2018; Karthikeyan and Mishra, 2021; M. Xu et al., 2022; Shangguan et al., 2023) use original LST datasets with the gaps they have, some others (Long et al., 2019; Zhao et al., 2021; Huang et al., 2022b) utilize additional gap filling methodologies to prepare a complete LST dataset in the entire spatial domain. While gap filling methodology is quite new, research has proven its benefits in downscaling soil moisture data (Abowarda et al., 2021; Huang et al., 2022a). Majority of the studies that

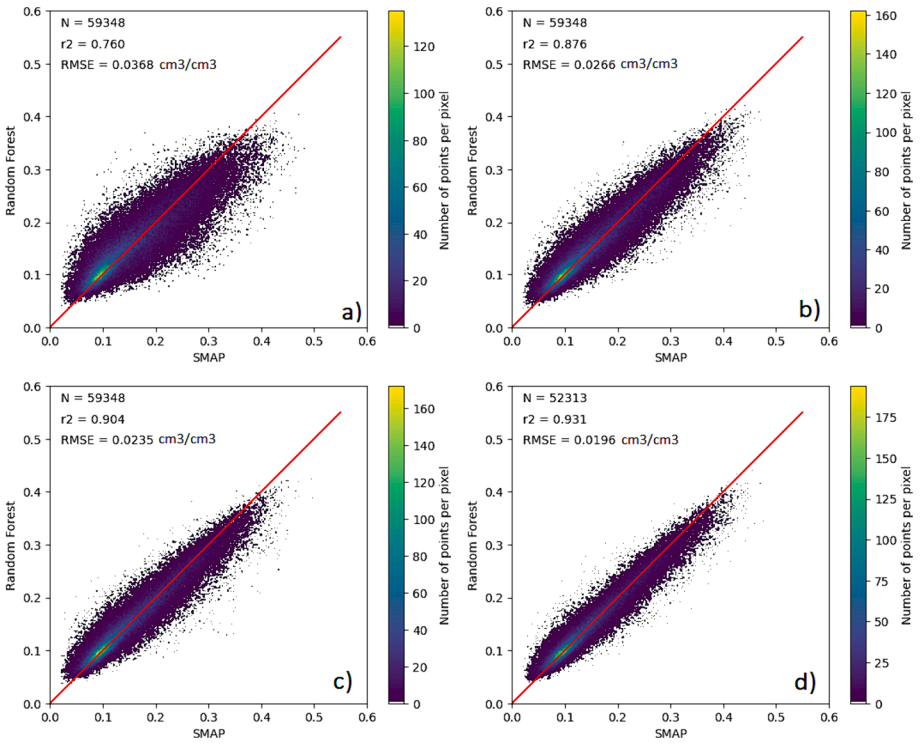


Fig. 6. Performance of using previous MSWEP precipitations as auxiliary variables a) n = 0, b) n = 3, c) n = 6, d) n = 13. The parameter n controls how many days ago the rainfall values are used as dependent variables. For example, when n = 6, the model will use the rainfall values from the past 6 days to predict soil moisture and when n = 0, the model will only use the corresponding day's precipitation value (no antecedent precipitation) as the auxiliary variable.

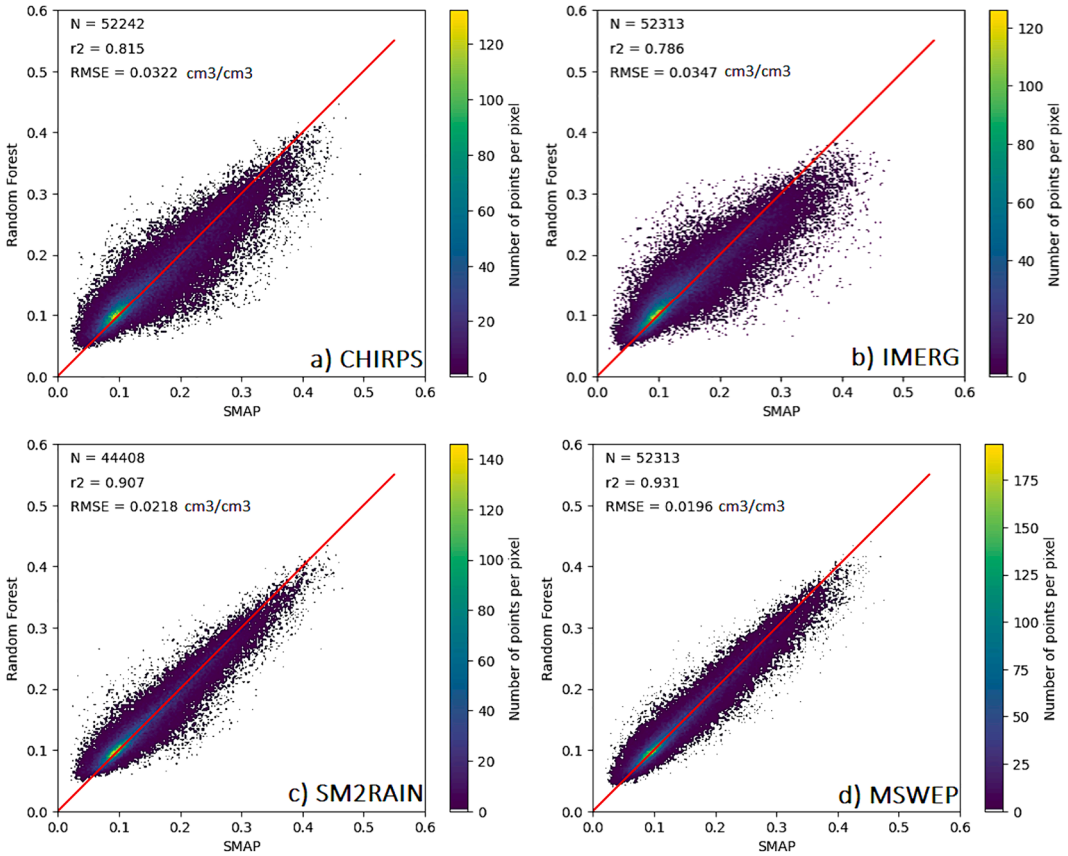


Fig. 7. The performance of different precipitation products on the machine learning algorithm (for n = 13) a) CHIRPS, b) IMERG, c) SM2RAIN, and d) MSWEP.

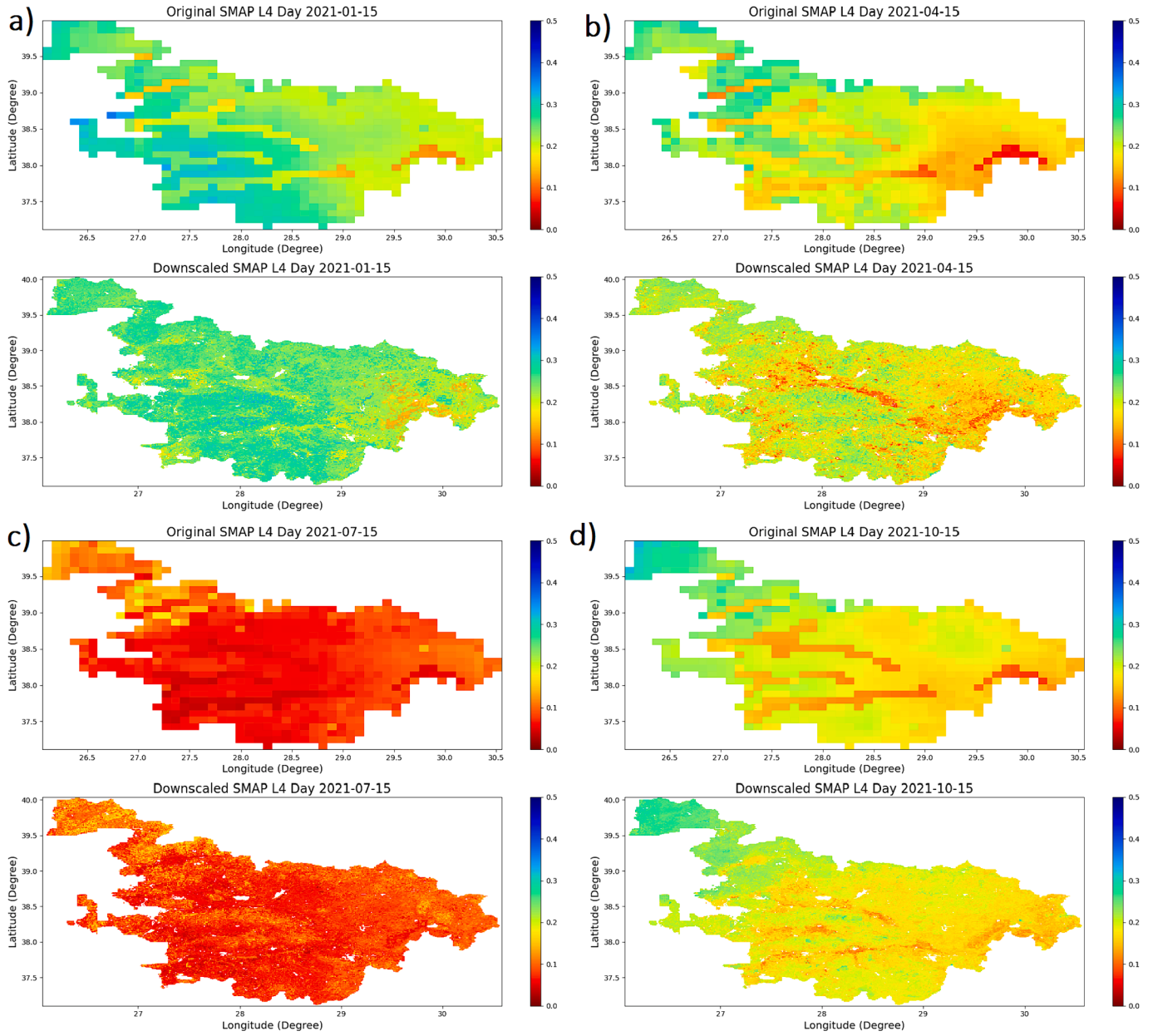


Fig. 8. Visual comparison of original SMAP L4 and downscaled SMAP L4 data for fifteenth day of a) January, b) April, c) July, and d) October for the year 2021. Moreover, spatial mean values of soil moisture are also presented for both original and downscaled data. (Units are cm^3/cm^3 for all maps.)

downscale soil moisture data utilize MODIS LST datasets, regardless of whether they used a reconstruction algorithm or not. When MODIS and VIIRS datasets are used together, discontinuities in LST maps can significantly be reduced (Table 3).

In this study, seamless LST Night maps are generated using a two-step approach. The first step involves merging MODIS (MOD21, MYD21) and VIIRS (VNP21) LST datasets using a daily merging procedure. However, this procedure cannot completely fill all data gaps. Therefore, a second step is implemented to fill the remaining gaps using seasonal linear regression with model-based data (CLSM). The advantage of this method is its practicality, as it only requires linear regression for both daily merging and model-based merging steps.

Huang et al. (2022b) proposed a reconstruction method that uses a multivariate regression model between LST and auxiliary parameters that show similar characteristics such as NDVI and DEM. In the validation process, they artificially removed 5 % of the quality data on clear sky days and compared the performance of their proposed method through the statistical indicators R^2 (0.828) and RMSE (3.79°K). On the other hand, Zhao et al. (2021) reconstructed LST by the annual temperature cycle (ATC) method with good statistical agreement (R^2: 0.96 and RMSE: 1.78°K), but they did not discuss the validation procedure in detail such as whether artificial gaps or in-situ LST datasets were used in their study or what was the number of pixels used to evaluate performance of ATC method. Long et al. (2019), on the other hand, obtained seamless LST maps by combining data from the MODIS and the China Meteorological Administration Land Data Assimilation System (CLDAS), achieving an R^2 of 0.865 and an RMSE of 1.86°K. Comparably, the new reconstruction method proposed in this study and the Final Daily Merged + SMAP product had mean statistical performances of 0.94 (R^2) and 1.84°K (RMSE). This outcome demonstrates the capability and comparably higher performance of the proposed LST reconstruction methodology.

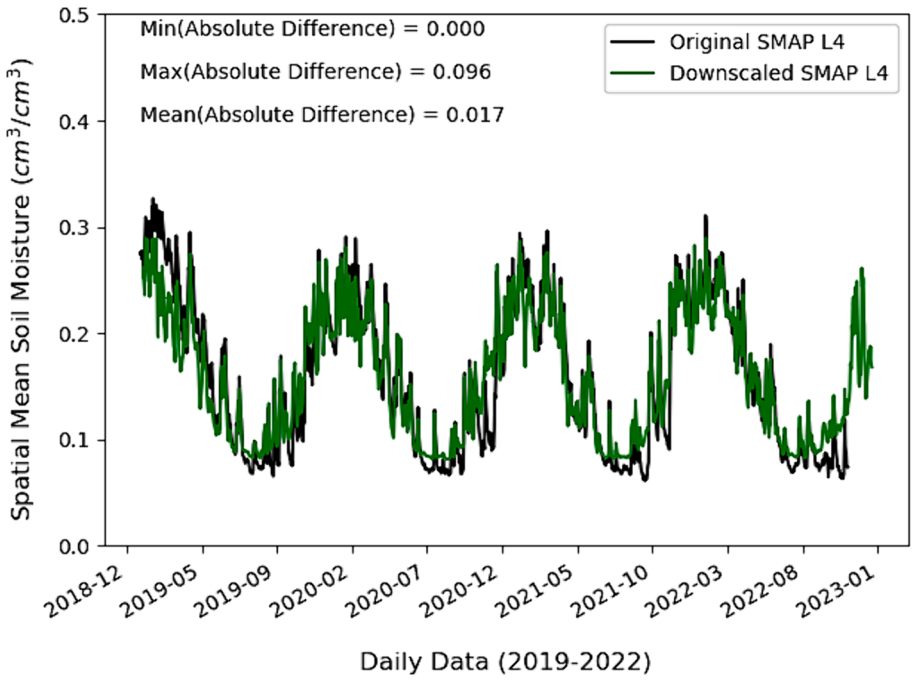


Fig. 9. Mean spatial difference between original and downscaled datasets. For each day between 2019 and 2022, mean soil moisture values have been calculated for original SMAP L4 and downscaled SMAP L4. Statistics of absolute mean spatial differences range between 0.00 and 0.096 with a mean of 0.017 cm^3/cm^3. It can be interpreted that the spatial difference between the original and downscaled data is minor for the WAB region.

Table 6
Comparison of Original SMAP L4, Dowscaled SMAP L4, and Downscaled SMAP L3 with insitu observations. For each station, closest grid cell is selected and compared with the remotely sensed data. SMAP L3 dataset is not spatially continous since this dataset is masked near coastal regions within an average range of approximately 36 km for quality standarts. This is the reason why original SMAP L3 dataset is not directly compared with insitu stations.

Station Number	Original SMAP L4			Downscaled SMAP L4			Downscaled SMAP L3		
	R	RMSE (cm^3/cm^3)	ubRMSE (cm^3/cm^3)	R	RMSE (cm^3/cm^3)	ubRMSE (cm^3/cm^3)	R	RMSE (cm^3/cm^3)	ubRMSE (cm^3/cm^3)
17184	0.907	0.222	0.078	0.909	0.202	0.072	0.883	0.187	0.071
17746	0.910	0.033	0.033	0.889	0.041	0.036	0.836	0.052	0.044
17750	0.892	0.074	0.071	0.836	0.084	0.080	0.780	0.089	0.084
17789	0.893	0.126	0.096	0.889	0.106	0.083	0.822	0.090	0.084
17233	0.913	0.049	0.048	0.835	0.084	0.079	0.846	0.081	0.080
17234	0.950	0.041	0.035	0.886	0.056	0.039	0.846	0.059	0.044
17862	0.727	0.103	0.065	0.578	0.089	0.070	0.712	0.064	0.061
18909	0.823	0.032	0.030	0.711	0.041	0.041	0.675	0.043	0.043
17175	0.847	0.151	0.132	0.865	0.131	0.120	0.788	0.147	0.140
17722	0.805	0.114	0.058	0.874	0.061	0.041	0.859	0.044	0.044
17771	0.885	0.261	0.099	0.808	0.212	0.109	0.772	0.209	0.111
17792	0.912	0.054	0.033	0.793	0.075	0.042	0.703	0.093	0.054
17850	0.940	0.079	0.042	0.861	0.118	0.039	0.833	0.104	0.043
17860	0.977	0.017	0.017	0.925	0.030	0.030	0.782	0.056	0.049
17886	0.868	0.109	0.055	0.853	0.119	0.058	0.826	0.104	0.065
17186	0.935	0.033	0.026	0.859	0.058	0.036	0.885	0.065	0.035
17220	0.886	0.109	0.029	0.822	0.075	0.037	0.801	0.066	0.039
17232	0.959	0.128	0.033	0.904	0.097	0.048	0.861	0.085	0.054
17822	0.959	0.048	0.036	0.839	0.050	0.049	0.817	0.053	0.051
17145	0.862	0.075	0.040	0.859	0.075	0.040	0.823	0.074	0.045
17180	0.719	0.091	0.091	0.602	0.125	0.105	0.594	0.103	0.103
17787	0.924	0.161	0.099	0.885	0.182	0.094	0.863	0.170	0.104
17188	0.664	0.153	0.089	0.613	0.149	0.093	0.628	0.155	0.092
17221	0.926	0.094	0.039	0.850	0.118	0.069	0.696	0.120	0.084
17820	0.946	0.071	0.029	0.876	0.081	0.048	0.840	0.082	0.050
17854	0.936	0.037	0.028	0.906	0.052	0.035	0.854	0.113	0.040
18442	0.787	0.063	0.055	0.604	0.067	0.057	0.682	0.066	0.059
17237	0.893	0.054	0.027	0.856	0.040	0.031	0.808	0.058	0.032
17824	0.899	0.035	0.029	0.823	0.037	0.036	0.819	0.045	0.038
17825	0.865	0.035	0.029	0.662	0.073	0.044	0.754	0.046	0.039
17827	0.869	0.063	0.036	0.796	0.057	0.042	0.794	0.070	0.042
Mean	***0.880***	***0.088***	***0.052***	***0.815***	***0.090***	***0.058***	***0.790***	***0.090***	***0.062***

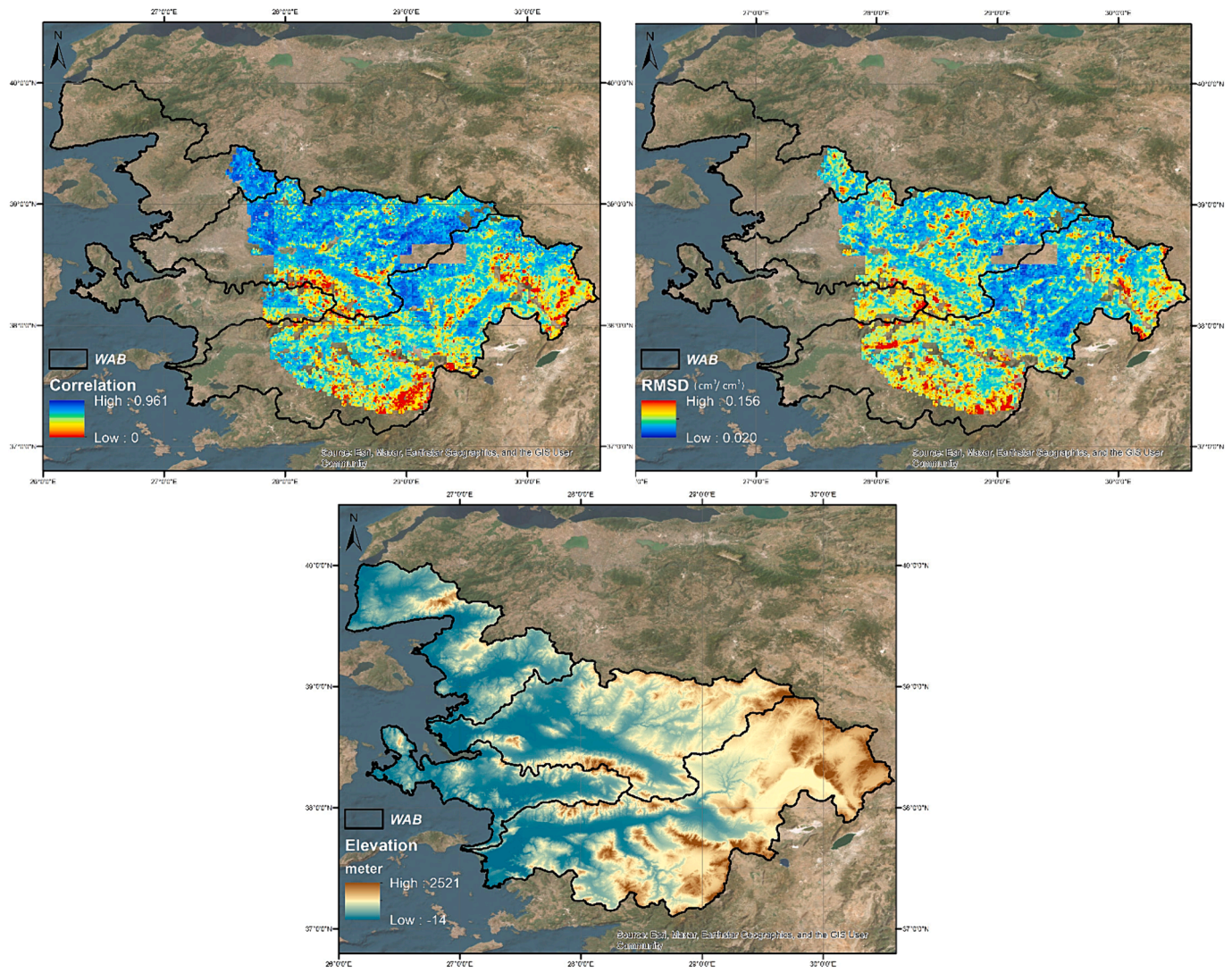


Fig. 10. Spatial distribution of a) Correlation Coefficient and b) Root Mean Squared Difference (RMSD) between SMAP L2 and Downscaled SMAP L3 for WAB region (Since SMAP L2 is also a satellite dataset RMSE is denoted here as RMSD to represent the root mean squared difference between two satellite datasets) c) DEM of the WAB region. SMAP L2 data is masked near the coastal region to maintain quality standards due to the native resolution of SMAP L3 radar and this is the reason for the spatial gaps on the correlation and RMSD maps. Variations on the R and RMSE values are seen particularly at watershed boundaries that are characterized by complex topography and high altitudes.

6.2. *Impact of rainfall datasets in training machine learning models*

Beck et al. (2019a) compared the performance of 26 different precipitation datasets using gauge-radar system. They concluded that MSWEP performed best performance among the other gauge corrected datasets. Casson et al. (2018) also showed that MSWEP is a good predictor for estimation of the snow water equivalent (SWE). Beck et al. (2017) analyzed the performance of thirteen uncorrected and nine gauge-corrected gridded precipitation datasets using in-situ observations by calibrating an hydrological model with streamflow records. They concluded that MSWEP dataset had comparably better performance in calibrating the hydrological model and outperformed other gridded precipitation datasets.

In this study, four machine learning models were trained using different precipitation datasets. The performance of the models improved in the order of IMERG, CHIRPS, S2MRAIN and MSWEP, as shown in Fig. 6. Thus, MSWEP was found to be the most accurate precipitation dataset to yield the best soil moisture results when compared with in-situ data. Although CHIRPS has better spatial resolution, more accurate machine learning models can be trained using the MSWEP dataset. In addition, machine learning test performances were significantly improved when antecedent precipitation records from previous days were used as additional auxiliary variables for MSWEP. The results revealed no improvements for other gridded precipitation datasets when previous day's values were used as auxiliary variables.

6.3. *Feature importance of auxiliary variables in SMAP L3 and SMAP L4*

Known for its simplicity and efficiency, Scikit-learn is a Python-based machine-learning library. It is accessible to both experts and beginners, and offers reusable tools for diverse applications. It is carefully built using the powerful array of numerical and scientific Python software, including NumPy and SciPy (Pedregosa et al., 2012; Buitinck et al., 2013). Feature importance values for auxiliary variables in SMAP L3 and L4 were calculated using the Scikit-learn module, and are shown in Fig. 13. For SMAP L3 datasets, precipitation variables P_{t-1}, P_{t-2}, P_{t-4}, and LST Night were the most important variables according to feature importance analysis for SMAP L3 (Fig. 13). On the other hand, LST night was the predominant factor in influencing the SMAP L4 data.

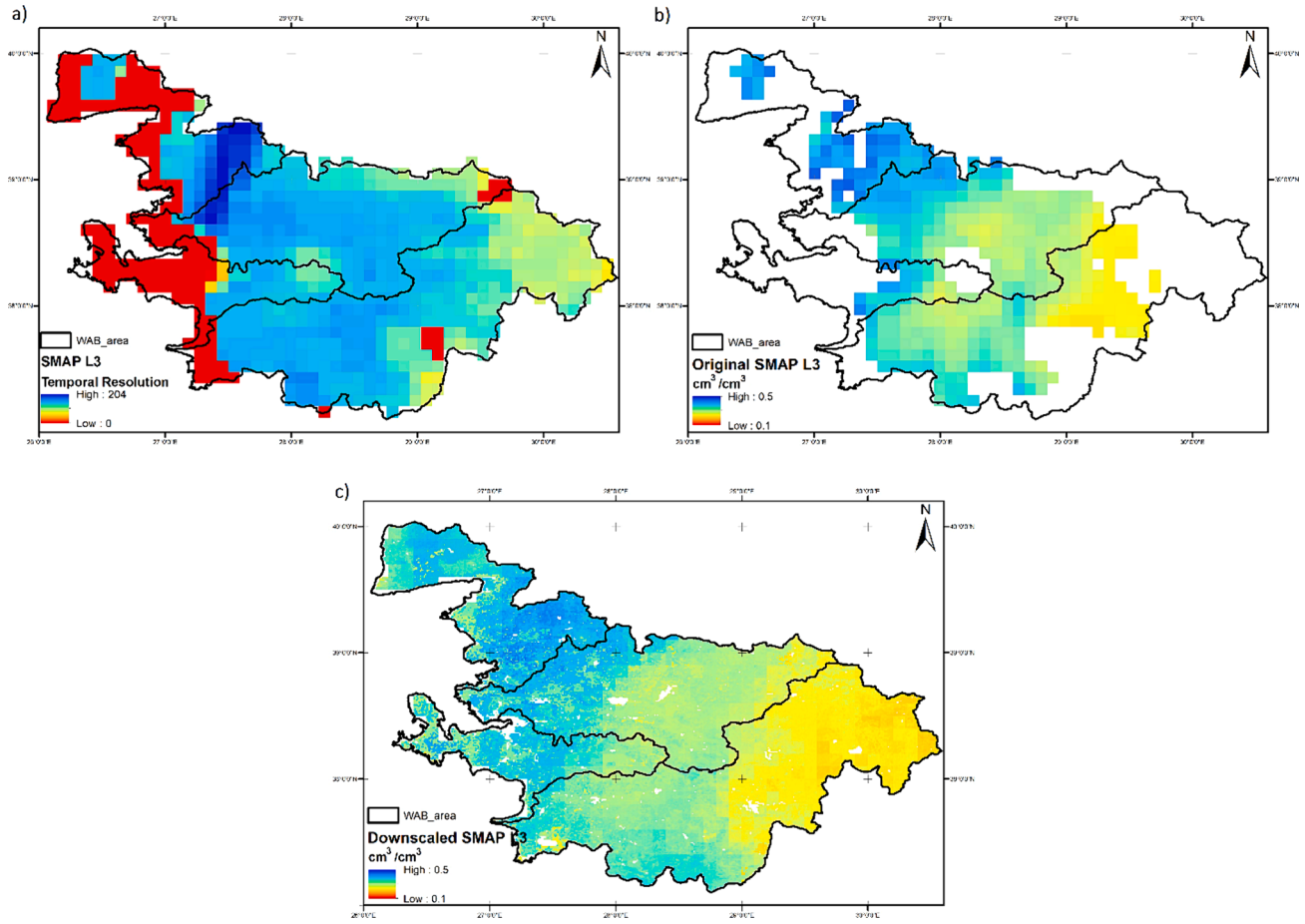


Fig. 11. a) Temporal resolution of Smap L3 data for WAB region and b) downscaled and c) original Smap L3 for first day of august 2021. the temporal resolution of Smap L3 for WAB region is defined as how many times smap passed from WAB region out of 365 days for 2021 (The mean is 135 out of 365). Nearly the half of the total surface area is empty for Kuzey Ege and Küçük Menderes Basin.

6.4. Performance of the trained random forest model

According to the comparison results, the RF model trained in this study was found to be quite successful in predicting soil moisture compared to other models given in Table 7. In addition, no reconstruction procedure was applied in other methods, which meant that these models were mostly trained on soil moisture values during dry periods (cloud free days). Conversely, the model trained in this study was also able to successfully predict soil moisture during rainy days. According to machine learning results, more accurate models can be trained using the SMAP L3 dataset than SMAP L4 (Table 7). Model performance analysis revealed a substantial reduction in the RMSE, with the error dropping by 50 %. Additionally, R^2 values demonstrated a significant improvement. It can be said that developed model outperforms the other RF models in the literature. Only Wide & Deep learning method produced slightly better R^2 results (Table 7).

6.5. Validation of downscaled data

The accuracy of the original SMAP L4 dataset first tested in the WAB region (Türkiye) for comparison with downscaled SMAP L4. The results revealed R values that varied from 0.664 (lowest) to 0.977 (highest), with an average of 0.88 ± 0.073 cm^3/cm^3. Similarly, RMSE and ubRMSE also varied as shown in Table 6 with average values of 0.088 ± 0.056 cm^3/cm^3 and 0.052 ± 0.028 cm^3/cm^3, respectively for the original SMAP L4 product. Colliander et al. (2022) evaluated the performance of SMAP L4 using sparse networks (with a total of nine networks and more than 750 stations).They concluded that for croplands and grasslands statistics were 0.74 and 0.7 for R and 0.059 cm^3/cm^3 and 0.053 cm^3/cm^3 for ubRMSE, respectively.

For the downscaling of the SMAP L4, the performance of the trained RF model in this study had a R^2 value of 0.90 and a RMSE value of 0.025 cm^3/cm^3. These values demonstrated a superior downscaling capability of the reconstructed-LST-integrated RF model for the WAB region. When compared to statistical evaluation analyses with in-situ observation stations, the downscaled SMAP L4 dataset showed similar performance to the original dataset with a much higher spatial resolution. In terms of average values, the RMSE and ubRMSE values were almost the same. However, the average R value in the downscaled SMAP L4 dataset decreased slightly (Table 6). To visually inspect the accuracy of downscaled maps, the original and downscaled maps for dry and wet seasons were reviewed. It was concluded that the high-resolution 1 km maps resembled the original 9 km maps satisfactorily. The spatial variation of soil moisture was clearly observable for both datasets (Fig. 8). In addition, the average of the original and downscaled data was calculated for each day between 2019 and 2022. When these values were compared, the time series showed minor differences (Fig. 9). The time series of downscaled data and in-situ observations were investigated for a 2-year period between 2021 and 2022. Accordingly, it can be concluded that the time series were generally in adequate agreement (Fig. 12).

For the downscaling of the SMAP L3, the performance of the trained RF model in this study had a R^2 value of 0.931 and a RMSE value of 0.0019 cm^3/cm^3. These results also demonstrated that the RF model formulated in this study was satisfactory to downscale the soil moisture in the WAB region. Mean statistical results were calculated as (0.79 ± 0.074), (0.09 ± 0.043 cm^3/cm^3), and (0.06 ± 0.026 cm^3/cm^3) for R,

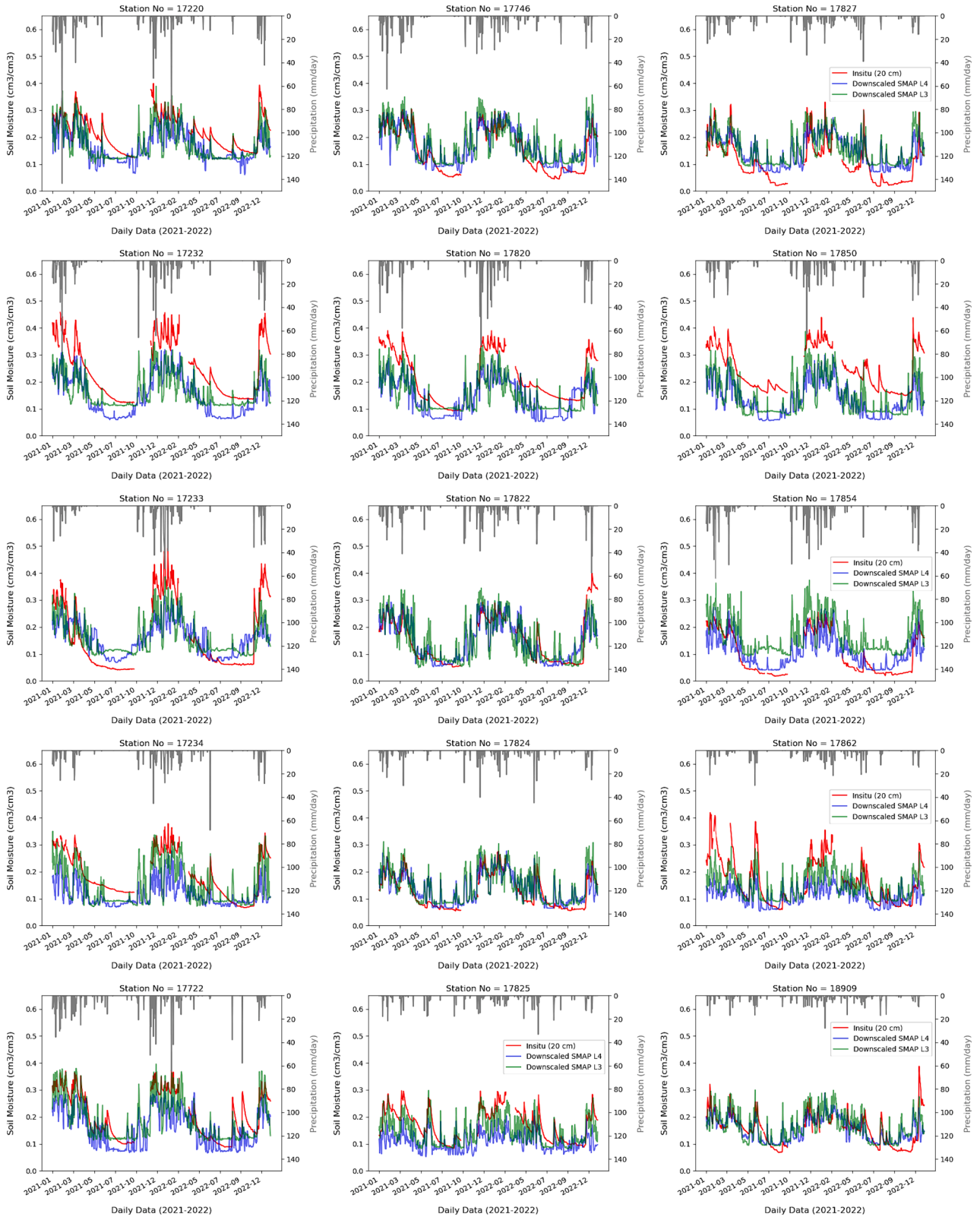


Fig. 12. Timeseries of downscaled SMAP datasets, insitu observations of soil moisture and precipitation for visual inspection of the proposed downscaling methodology for two years period (2021–2022). 15 out of the 31 in situ stations are presented.

RMSE and ubRMSE, respectively, which were similar to other compared datasets (Table 6). Because most of the stations measured soil moisture at 20 cm from ground surface, it was not possible to fully assess the performance of the downscaled SMAP L3 with in-situ observations for WAB region. Therefore, the performance of the downscaled SMAP L3 were also validated with SMAP + Sentinel (L2) high spatial, low temporal resolution dataset in 2019–2022 period. The mean statistics were calculated to be (0.761 ± 0.11) for R and (0.05 ± 0.014 cm^3/cm^3) for

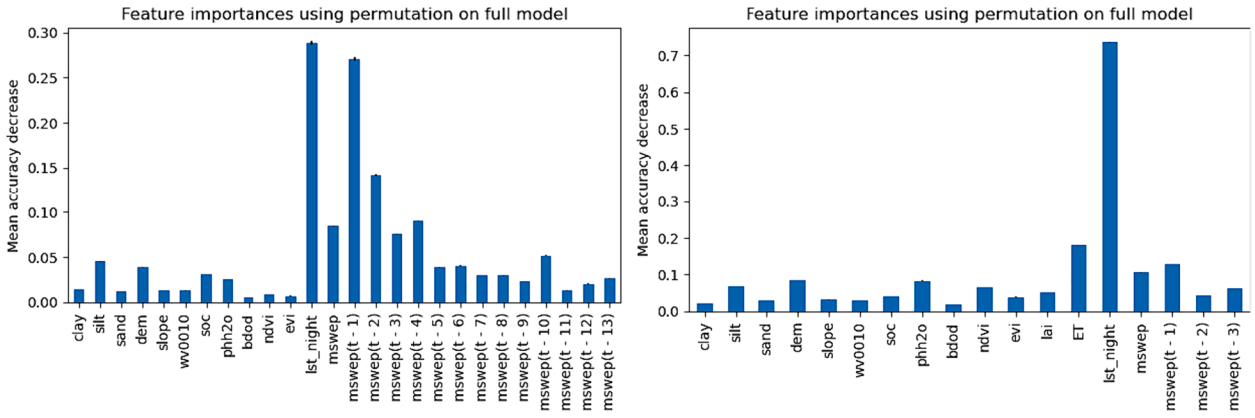


Fig. 13. Feature importance of a) SMAP L3 and b) SMAP L4 soil moisture data. For SMAP L4, LST_Night parameter was the single most important variable for downscaling. For SMAP L3, lst_night and the precipitation variables were the most important parameters. Permutation based feature importance values were calculated using Scikit-learn module in Python (Buitinck et al., 2013).

Table 7
Comparison of test performance values of the trained machine learning model in this study with other research from the literature (Units of the RMSE values are (cm^3/cm^3).

Method	Soil Moisture Dataset	Performance	LST Gap filling	Reference
Long short-termmemory network (LSTM)	SMAP L3 36 km	RMSE = 0.043 R^2 = 0.706	No	(Y. Zhang et al., 2022)
Random Forest		RMSE = 0.04 R^2 = 0.753		
Random Forest	Essential Climate Variable (ECV)	RMSE = 0.036 R^2 = 0.704	No	(Liu et al., 2020)
Wide & Deep learning method	SMAP L3 36 km	RMSE = 0.026 R^2 = 0.95	No	(M. Xu et al., 2022)
Random Forest	SMAP L3 9 km	RMSE = 0.019 R^2 = 0.931	Yes	Current Study
	SMAP L4 9 km	RMSE = 0.025 R^2 = 0.90		

RMSD between downscaled SMAP L3 and L2 data (Fig. 10).

Native resolution of the SMAP radiometer is approximately 36 km. Hence, regions close to big water masses have been masked for the quality considerations. This is the main reason for the lack of data for 46 % and 51 % of the total surface areas of the Kuzey Ege and Küçük Menderes basins, respectively, which are coastal basins where data is missing. The mean of the temporal resolution of the original SMAP L3 is 135 days out of 365 due to the native resolution of SMAP satellite, which has data gaps in the coastal basins such as Kuzey Ege and Küçük Menderes. In other words, the procedure presented in this study offers a promising solution for soil moisture downscaling also in coastal regions, by not only enhancing the resolution from 9 km to 1 km but also enabling spatial gap filling rather than interpolation (Fig. 11). It can be seen from Fig. 12 that a strong agreement between the downscaled SMAP L3 and SMAP L4 data with soil moisture and precipitation measurements derived from in-situ stations. The observed trends in the datasets exhibit remarkable consistency.

While some previous studies have relied on in-situ observations at 5 cm for SSM validation, this study lacked in-situ data from 5 cm depth and thus utilized measurements at a deeper depth of 20 cm within the WAB region for validation purposes. Recently, Feldman et al. (2023) discussed that L-band measurements can often probe beyond the traditional 5 cm depth limit, and reveal information about deeper soil moisture content when combining the analysis of microwave emission depth and the statistical properties of vertically correlated soil moisture data. Previous studies have utilized L-band satellite soil moisture retrievals to estimate effective depth. While these depths tend to be greater than 5 cm in wetter conditions, they rarely exceed 30 cm (Akbar et al., 2018; Short Gianotti et al., 2019). Furthermore, the validation statistics of this study also yielded compatible results with the other sparse network comparison analysis (Beck et al., 2021; Colliander et al., 2022). Hence, it can be considered that the depth inconsistency of the comparison procedure does not significantly hinder the validation process as long as the comparison depth difference does not go beyond 30 cm. Nonetheless, potential issues due to depth discrepancies of remotely sensed data with in-situ observations should be taken into account.

7. Conclusions

Accurate downscaling of soil moisture data is a significant step for hydrological modeling and agricultural planning. Most of the downscaling methods in the literature have used spatially discontinuous LST datasets as the main auxiliary parameter due to cloud contamination. While this situation does not cause a problem for the training of machine learning models, it limits the creation of continuous soil moisture maps, which later result in additional handicaps. Based on this motivation, a hybrid reconstruction method has been proposed in this study to obtain continuous LST maps, which were later integrated into a robust machine learning algorithm for creating high resolution gapless soil moisture products. Key findings of the study can be listed as follows:

- The effectiveness of the reconstruction method was verified in the western watersheds of Türkiye, where valid pixel percentage values were significantly increased with the joint use of the MODIS (MOD21, MYD21) VIIRS (VNP21) datasets.
- The proposed method has been tested with artificially generated gaps created in the original maps. In addition, the performance of the proposed algorithm was also compared with the methods used in other downscaling studies. The results revealed that the proposed method is extremely practical and easily applicable, and has a promising accuracy. It only uses linear regression to produce the final product and requires less computational burden.
- The test performances of four different rainfall datasets were compared for their potential uses in downscaling soil moisture, and the MSWEP product was found to give the best results.
- The use of antecedent rainfall values as input variables in machine learning models was shown to improve the test performance of the models. However, a major limitation of using global rainfall products

was their low spatial resolution (0.1 degrees) and the need to resample to improved spatial resolution if they will be used as an auxiliary variable for predicting SSM using trained machine learning models.

- The developed LST reconstruction technique was successfully used to downscale 9 km spatial resolution datasets of SMAP L3 and SMAP L4 and gave promising results to produce 1 km datasets. SMAP L3 was much more sensitive to precipitation data unlike SMAP L4 when the feature importance of RF model was evaluated.
- The inherent resolution of SMAP radar data is approximately 36 km. Thus, a significant portion of the data for basins proximal to coastal sites must be masked due to quality considerations associated with the original SMAP L3 dataset. The proposed technique provided an alternative approach for creating downscaled SSM in these regions by not only enhancing spatial resolution but also filling these data gaps rather than interpolation.

Sources of funding

This research was supported by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) with the project numbered 123Y039. The first author is financially sponsored by the Higher Education Council of Türkiye through 100/2000 PhD Scholarship Program.

CRediT authorship contribution statement

Onur Güngör Şahin: Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Orhan Gündüz:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data and codes that support the findings of this study are available upon a rational request from the corresponding author.

Acknowledgements

The authors would like to offer sincere gratitude to the Turkish State Meteorological Service for providing our research with the in-situ observations soil moisture records, without which this research could not be implemented. The authors also acknowledge the support of the Scientific and Technological Research Council of Türkiye (TÜBİTAK) through the project numbered 123Y039 as well as the financial support provided to the first author by the Higher Education Council of Türkiye through 100/2000 PhD Scholarship Program.

References

Abowarda, A.S., Bai, L., Zhang, C., Long, D., Li, X., Huang, Q., Sun, Z., 2021. Generating surface soil moisture at 30 m spatial resolution using both data fusion and machine learning toward better water resources management at the field scale. Remote Sens. Environ. 255, 112301 https://doi.org/10.1016/j.rse.2021.112301.

Akbar, R., Short Gianotti, D., McColl, K.A., Haghighi, E., Salvucci, G.D., Entekhabi, D., 2018. Hydrological storage length scales represented by remote sensing estimates of soil moisture and precipitation. Water Resour. Res. 54, 1476–1492. https://doi.org/10.1002/2017WR021508.

Amjad, M., Yilmaz, M.T., Yucel, I., Yilmaz, K.K., 2020. Performance evaluation of satellite- and model-based precipitation products over varying climate and complex topography. J. Hydrol. 584, 124707 https://doi.org/10.1016/j.jhydrol.2020.124707.

Beck, H.E., Vergopolan, N., Pan, M., Levizzani, V., Van Dijk, A.I.J.M., Weedon, G.P., Brocca, L., Pappenberger, F., Huffman, G.J., Wood, E.F., 2017. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. Hydrol. Earth Syst. Sci. 21, 6201–6217. https://doi.org/10.5194/HESS-21-6201-2017.

Beck, H.E., Pan, M., Roy, T., Weedon, G.P., Pappenberger, F., Van Dijk, A.I.J.M., Huffman, G.J., Adler, R.F., Wood, E.F., 2019a. Daily evaluation of 26 precipitation datasets using stage-IV gauge-radar data for the CONUS. Hydrol. Earth Syst. Sci. 23, 207–224. https://doi.org/10.5194/HESS-23-207-2019.

Beck, H.E., Wood, E.F., Pan, M., Fisher, C.K., Miralles, D.G., Van Dijk, A.I.J.M., McVicar, T.R., Adler, R.F., 2019b. MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment. Bull. Am. Meteorol. Soc. 100, 473–500. https://doi.org/10.1175/BAMS-D-17-0138.1.

Beck, H.E., Pan, M., Miralles, D.G., Reichle, R.H., Dorigo, W.A., Hahn, S., Sheffield, J., Karthikeyan, L., Balsamo, G., Parinussa, R.M., van Dijk, A.I.J.M., Du, J., Kimball, J. S., Vergopolan, N., Wood, E.F., 2021. Evaluation of 18 satellite- and model-based soil moisture products using in-situ measurements from 826 sensors. Hydrol. Earth Syst. Sci. 25, 17–40. https://doi.org/10.5194/hess-25-17-2021.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Brocca, L., Filippucci, P., Hahn, S., Ciabatta, L., Massari, C., Camici, S., Schüller, L., Bojkov, B., Wagner, W., 2019. SM2RAIN-ASCAT (2007–2018): global daily satellite rainfall data from ASCAT soil moisture observations. Earth Syst. Sci. Data 11, 1583–1601. https://doi.org/10.5194/ESSD-11-1583-2019.

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Müller, A.C., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., Vanderplas, J., Joly, A., Holt, B., Varoquaux, G., 2013. API design for machine learning software: experiences from the scikit-learn project.

Bulut, B., Yilmaz, T.M., Afshar, M.H., Şorman, Ü.A., Yücel, I., Cosh, M.H., Şimşek, O., 2019. Evaluation of remotely-sensed and model-based soil moisture products according to different soil type, vegetation cover and climate regime using station-based observations over Turkey. Remote Sens. 11 https://doi.org/10.3390/rs11161875.

Campbell Scientific, 2020. CS616 and CS625 water content reflectometers, Instruction manual.

Casson, D.R., Werner, M., Weerts, A., Solomatine, D., 2018. Global re-analysis datasets to improve hydrological assessment and snow water equivalent estimation in a sub-Arctic watershed. Hydrol. Earth Syst. Sci. 22, 4685–4697. https://doi.org/10.5194/HESS-22-4685-2018.

Colliander, A., Jackson, T.J., Bindlish, R., Chan, S., Das, N., Kim, S.B., Cosh, M.H., Dunbar, R.S., Dang, L., Pashaian, L., Asanuma, J., Aida, K., Berg, A., Rowlandson, T., Bosch, D., Caldwell, T., Caylor, K., Goodrich, D., 2017. Validation of SMAP surface soil moisture products with core validation sites. Remote Sens. Environ. 191, 215–231. https://doi.org/10.1016/j.rse.2017.01.021.

Colliander, A., Reichle, R.H., Crow, W.T., Cosh, M.H., Chen, F., Chan, S., Das, N.N., Bindlish, R., Chaubell, J., Kim, S., Liu, Q., Oaneill, P.E., Dunbar, R.S., Dang, L.B., Kimball, J.S., Jackson, T., Al-Jassar, H., Asanuma, J., Bhattacharya, B., Berg, A.A., Bosch, D.D., Bourgeau-Chavez, L., Caldwell, T., Calvet, J.C., Collins, C.H., Jensen, K. H., Livingston, S., Lopez-Baeza, E., Martínez-Fernández, J., McNairn, H., Moghaddam, M., Montzka, C., Notarnicola, C., Pellarin, T., Greimeister-Pfeil, I., Pulliainen, J., Hernández, J.G.R., Seyfried, M., Starks, P.J., Su, B., Van Der Velde, R., Zeng, Y., Thibeault, M., Vreugdenhil, M., Walker, J.P., Zribi, M., Entekhabi, D., Yueh, S.H., 2022. Validation of soil moisture data products from the NASA SMAP mission. IEEE J Sel. Top. Appl. Earth Obs. Remote Sens. 15, 364–392. https://doi.org/10.1109/JSTARS.2021.3124743.

Crosson, W.L., Al-Hamdan, M.Z., Hemmings, S.N.J., Wade, G.M., 2012. A daily merged MODIS aqua-Terra land surface temperature data set for the conterminous United States. Remote Sens. Environ. 119, 315–324. https://doi.org/10.1016/J.RSE.2011.12.019.

Dari, J., Quintana-Seguí, P., Escorihuela, M.J., Stefan, V., Brocca, L., Morbidelli, R., 2021. Detecting and mapping irrigated areas in a Mediterranean environment by using remote sensing soil moisture and a land surface model. J. Hydrol. 596, 126129 https://doi.org/10.1016/J.JHYDROL.2021.126129.

Das, N.N., Entekhabi, D., Dunbar, R.S., Chaubell, M.J., Colliander, A., Yueh, S., Jagdhuber, T., Chen, F., Crow, W., O'Neill, P.E., Walker, J.P., Berg, A., Bosch, D.D., Caldwell, T., Cosh, M.H., Collins, C.H., Lopez-Baeza, E., Thibeault, M., 2019. The SMAP and copernicus sentinel 1A/B microwave active-passive high resolution surface soil moisture product. Remote Sens. Environ. 233, 111380 https://doi.org/10.1016/J.RSE.2019.111380.

Das, N., Entekhabi, D., Dunbar, R.S., Kim, S., Yueh, S., Colliander, A., O'Neill, P.E., Jackson, T., T.Jagdhuber, Chen, F., Crow, W.T., Walker, J., Berg, A., Bosch, D., Caldwell, T., Cosh, M., 2018. SMAP/Sentinel-1 L2 Radiometer/Radar 30-Second Scene 3 km EASE-Grid Soil Moisture, Version 2 [Indicate subset used]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/KE1CSVXMI95Y. [Date Accessed].

Didan, K., 2018. VIIRS/NPP Vegetation Indices 16-Day L3 Global 1km SIN Grid V001. distributed by NASA EOSDIS Land Processes Distributed Active Archive Center. https://doi.org/10.5067/VIIRS/VNP13A2.001. Accessed 2023-09-12.

Dorigo, W.A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., Van Oevelen, P., Robock, A., Jackson, T., 2011. The international soil moisture network: a data hosting facility for global in-situ soil moisture measurements. Hydrol. Earth Syst. Sci. 15, 1675–1698. https://doi.org/10.5194/HESS-15-1675-2011.

Dorigo, W.A., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiová, A., Sanchis-Dufau, A.D., Zamojski, D., Cordes, C., Wagner, W., Drusch, M., 2013. Global automated quality control of in-situ soil moisture data from the international soil moisture network. Vadose Zo. J. 12, 1–21. https://doi.org/10.2136/VZJ2012.0097.

Entekhabi, D., Yueh, Si., O'Neil, P.E., Kellogg, K.H., Allen, A., Bindlish, R., Administration, N.A. and S., 2014. SMAP Handbook. Mapp. Soil Moisture Free. from Sp. 192.

Entekhabi, D., Das, N., Njoku, E.G., Johnson, J.T., Shi, J., 2016. SMAP L3 radar/radiometer global daily 9 km EASEGrid soil moisture, Version 3 [Indicate subset used]. . Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/7KKNQ5UURM2W. Date Accessed 09-12-2023.

Fang, B., Kansara, P., Dandridge, C., Lakshmi, V., 2021. Drought monitoring using high spatial resolution soil moisture data over Australia in 2015–2019. J. Hydrol. 594, 125960 https://doi.org/10.1016/J.JHYDROL.2021.125960.

Fang, B., Lakshmi, V., Cosh, M., Liu, P.W., Bindlish, R., Jackson, T.J., 2022. A global 1-km downscaled SMAP soil moisture product based on thermal inertia theory. Vadose Zo. J. 21, e20182.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D.E., 2007. The shuttle radar topography mission. Rev. Geophys. 45, 2004. https://doi.org/10.1029/2005RG000183.

Feldman, A.F., Short Gianotti, D.J., Dong, J., Akbar, R., Crow, W.T., McColl, K.A., Konings, A.G., Nippert, J.B., Tumber-Dávila, S.J., Holbrook, N.M., Rockwell, F.E., Scott, R.L., Reichle, R.H., Chatterjee, A., Joiner, J., Poulter, B., Entekhabi, D., 2023. Remotely sensed soil moisture can capture dynamics relevant to plant water uptake. Water Resour. Res. 59, e2022WR033814. https://doi.org/10.1029/2022WR033814.

Fuentes, I., Padarian, J., Vervoort, R.W., 2022. Towards near real-time national-scale soil water content monitoring using data fusion as a downscaling alternative. J. Hydrol. 609 https://doi.org/10.1016/j.jhydrol.2022.127705.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Sci. Data 2015 21 2, 1–21. https://doi.org/10.1038/sdata.2015.66.

Gerber, F., De Jong, R., Schaepman, M.E., Schaepman-Strub, G., Furrer, R., 2018. Predicting missing values in spatio-temporal remote sensing data. IEEE Trans. Geosci. Remote Sens. 56, 2841–2853. https://doi.org/10.1109/TGRS.2017.2785240.

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. Nat. 2020 5857825 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2.

Hengl, T., De Jesus, J.M., Heuvelink, G.B.M., Gonzalez, M.R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: global gridded soil information based on machine learning. PLoS One 12, e0169748.

Hu, F., Wei, Z., Zhang, W., Dorjee, D., Meng, L., 2020. A spatial downscaling method for SMAP soil moisture through visible and shortwave-infrared remote sensing data. J. Hydrol. 590, 125360 https://doi.org/10.1016/j.jhydrol.2020.125360.

Huang, S., Zhang, X., Chen, N., Ma, H., Fu, P., Dong, J., Gu, X., Nam, W.H., Xu, L., Rab, G., Niyogi, D., 2022a. A novel fusion method for generating surface soil moisture data with high accuracy, high spatial resolution, and high spatio-temporal continuity. Water Resour. Res. 58, e2021WR030827. https://doi.org/10.1029/2021WR030827.

Huang, S., Zhang, X., Chen, N., Ma, H., Zeng, J., Fu, P., Nam, W.H., Niyogi, D., 2022b. Generating high-accuracy and cloud-free surface soil moisture at 1 km resolution by point-surface data fusion over the southwestern U.S. Agric. for. Meteorol. 321, 108985 https://doi.org/10.1016/j.agrformet.2022.108985.

Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Tan, J., 2019. Integrated multi-satellitE retrievals for GPM (IMERG) Technical Documentation.

Hulley, G.C., Ghent, D., 2019. Taking the temperature of the Earth : steps towards integrated understanding of variability and change. ISBN: 978-0-12-814458-9 https://doi.org/10.1016/C2017-0-01600-2.

Hulley, G., Freepartner, R., Malakar, N., Sudipta, S., 2016. Moderate resolution imaging spectroradiometer (MODIS) land surface temperature and emissivity product (MxD21) user guide collection-6. National Aeronautics and Space Administration.

Hulley, G., 2018. VIIRS/NPP Land surface temperature and emissivity daily L3 global 1km SIN grid day V001. distributed by NASA EOSDIS land processes distributed active archive center. https://doi.org/10.5067/VIIRS/VNP21A1D.001. Accessed 2023-09-12.

Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J.C., Mathis, M., Brumby, S.P., 2021. Global land use/land cover with sentinel 2 and deep learning. Int. Geosci. Remote Sens. Symp. 2021-July, 4704–4707. https://doi.org/10.1109/IGARSS47720.2021.9553499.

Karthikeyan, L., Mishra, A.K., 2021. Multi-layer high-resolution soil moisture estimation using machine learning over the United States. Remote Sens. Environ. 266, 112706 https://doi.org/10.1016/j.rse.2021.112706.

Khorrami, B., Ali, S., Gündüz, O., 2023. Investigating the local-scale fluctuations of groundwater storage by using downscaled GRACE/GRACE-FO JPL mascon product based on machine learning (ML) algorithm. Water Resour. Manag. 37, 3439–3456. https://doi.org/10.1007/S11269-023-03509-W/TABLES/2.

Li, Z.L., Leng, P., Zhou, C., Chen, K.S., Zhou, F.C., Shang, G.F., 2021. Soil moisture retrieval from remote sensing measurements: current knowledge and directions for the future. Earth-Science Rev. 218, 103673 https://doi.org/10.1016/J.EARSCIREV.2021.103673.

Li, X., Zhou, Y., Asrar, G.R., Zhu, Z., 2018. Creating a seamless 1 km resolution daily land surface temperature dataset for urban and surrounding areas in the conterminous United States. Remote Sens. Environ. 206, 84–97. https://doi.org/10.1016/j.rse.2017.12.010.

Liu, Y., Jing, W., Wang, Q., Xia, X., 2020. Generating high-resolution daily soil moisture by using spatial downscaling techniques: a comparison of six machine learning algorithms. Adv. Water Resour. 141 https://doi.org/10.1016/j.advwatres.2020.103601.

Long, D., Bai, L., Yan, L., Zhang, C., Yang, W., Lei, H., Quan, J., Meng, X., Shi, C., 2019. Generation of spatially complete and daily continuous surface soil moisture of high spatial resolution. Remote Sens. Environ. 233, 111364 https://doi.org/10.1016/j.rse.2019.111364.

Ma, J., Shen, H., Wu, P., Wu, J., Gao, M., Meng, C., 2022. Generating gapless land surface temperature with a high spatio-temporal resolution by fusing multi-source satellite-observed and model-simulated data. Remote Sens. Environ. 278, 113083 https://doi.org/10.1016/J.RSE.2022.113083.

Mo, Y., Xu, Y., Chen, H., Zhu, S., 2021. A review of reconstructing remotely sensed land surface temperature under cloudy conditions. Remote Sens. 2021, Vol. 13, Page 2838 13, 2838. https://doi.org/10.3390/RS13142838.

Mohanty, B.P., Cosh, M.H., Lakshmi, V., Montzka, C., 2017. Soil moisture remote sensing: state-of-the-science. Vadose Zo. J. 16, 1–9. https://doi.org/10.2136/vzj2016.10.0105.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2012. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.

Peng, J., Albergel, C., Balenzano, A., Brocca, L., Cartus, O., Cosh, M.H., Crow, W.T., Dabrowska-Zielinska, K., Dadson, S., Davidson, M.W.J., de Rosnay, P., Dorigo, W., Gruber, A., Hagemann, S., Hirschi, M., Kerr, Y.H., Lovergine, F., Mahecha, M.D., Marzahn, P., Mattia, F., Musial, J.P., Preuschmann, S., Reichle, R.H., Satalino, G., Silgram, M., van Bodegom, P.M., Verhoest, N.E.C., Wagner, W., Walker, J.P., Wegmüller, U., Loew, A., 2021. A roadmap for high-resolution satellite soil moisture applications – confronting product characteristics with user requirements. Remote Sens. Environ. 252, 112162 https://doi.org/10.1016/j.rse.2020.112162.

Reichle, R., De Lannoy, G., Koster, R.D., Crow, W.T., Kimball, J.S., 2017. SMAP L4 9 km EASE-Grid Surface and Root Zone Soil Moisture Land Model Constants, Version 3: 3-hourly Analysis Update, 3-hourly Geophysical Data, and Land Model Constants. https://doi.org/10.5067/20ULJH6EZKFJ.

Shangguan, Y., Min, X., Shi, Z., 2023. Inter-comparison and integration of different soil moisture downscaling methods over the Qinghai-Tibet plateau. J. Hydrol. 617, 129014 https://doi.org/10.1016/J.JHYDROL.2022.129014.

Shiff, S., Helman, D., Lensky, I.M., 2021. Worldwide continuous gap-filled MODIS land surface temperature dataset. Nat. (Scientific Data) 8, 1–10. https://doi.org/10.1038/s41597-021-00861-7.

Short Gianotti, D.J., Salvucci, G.D., Akbar, R., McColl, K.A., Cuenca, R., Entekhabi, D., 2019. Landscape water storage and subsurface correlation from satellite surface soil moisture and precipitation observations. Water Resour. Res. 55, 9111–9132. https://doi.org/10.1029/2019WR025332.

Tavakol, A., McDonough, K.R., Rahmani, V., Hutchinson, S.L., Hutchinson, J.M.S., 2021. The soil moisture data bank: the ground-based, model-based, and satellite-based soil moisture data. Remote Sens. Appl. Soc. Environ. 24 https://doi.org/10.1016/j.rsase.2021.100649.

Vergopolan, N., Chaney, N.W., Beck, H.E., Pan, M., Sheffield, J., Chan, S., Wood, E.F., 2020. Combining hyper-resolution land surface modeling with SMAP brightness temperatures to obtain 30-m soil moisture estimates. Remote Sens. Environ. 242, 111740 https://doi.org/10.1016/J.RSE.2020.111740.

Vergopolan, N., Xiong, S., Estes, L., Wanders, N., Chaney, N.W., Wood, E.F., Konar, M., Caylor, K., Beck, H.E., Gatti, N., Evans, T., Sheffield, J., 2021. Field-scale soil moisture bridges the spatial-scale gap between drought monitoring and agricultural yields. Hydrol. Earth Syst. Sci. 25, 1827–1847. https://doi.org/10.5194/HESS-25-1827-2021.

Wan, Z., Hook, S., Hulley, G., 2021. MODIS/terra land surface temperature/emissivity 5-min L2 swath 1km V061. distributed by NASA EOSDIS land processes distributed active archive center. https://doi.org/10.5067/MODIS/MOD11_L2.061.Accessed 2023-10-29.

Wigneron, J.P., Jackson, T.J., O'Neill, P., De Lannoy, G., de Rosnay, P., Walker, J.P., Ferrazzoli, P., Mironov, V., Bircher, S., Grant, J.P., Kurum, M., Schwank, M., Munoz-Sabater, J., Das, N., Royer, A., Al-Yaari, A., Al Bitar, A., Fernandez-Moran, R., Lawrence, H., Mialon, A., Parrens, M., Richaume, P., Delwart, S., Kerr, Y., 2017. 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. 192, 238–262. https://doi.org/10.1016/j.rse.2017.01.024.

Wu, P., Yin, Z., Zeng, C., Duan, S.B., Gottsche, F.M., Ma, X., Li, X., Yang, H., Shen, H., 2021. Spatially continuous and high-resolution land surface temperature product generation: a review of reconstruction and spatiotemporal fusion techniques. IEEE Geosci. Remote Sens. Mag. 9, 112–137. https://doi.org/10.1109/MGRS.2021.3050782.

Xu, F., Fan, J., Yang, C., Liu, J., Zhang, X., 2022. Reconstructing all-weather daytime land surface temperature based on energy balance considering the cloud radiative effect. Atmos. Res. 279, 106397 https://doi.org/10.1016/J.ATMOSRES.2022.106397.

Xu, M., Yao, N., Yang, H., Xu, J., Hu, A., Goncalves, G., de Goncalves, L., Liu, G., 2022. Downscaling SMAP soil moisture using a wide & deep learning method over the continental United States. J. Hydrol. 609, 127784 https://doi.org/10.1016/j.jhydrol.2022.127784.

Yu, P., Zhao, T., Shi, J., Ran, Y., Jia, L., Ji, D., Xue, H., 2022. Global spatiotemporally continuous MODIS land surface temperature dataset. Nat. (scientific Data) 9, 1–15. https://doi.org/10.1038/s41597-022-01214-8.

Zhang, Y., Chen, Y., Chen, L., Xu, S., Sun, H., 2022. A machine learning-based approach for generating high-resolution soil moisture from SMAP products. Geocarto Int. 1–22. https://doi.org/10.1080/10106049.2022.2105406.

Zhang, Y., Yang, Y., Pan, X., Ding, Y., Hu, J., Dai, Y., 2023. Multiinformation fusion network for mapping gapless all-sky land surface temperature using thermal infrared and reanalysis data. IEEE Trans. Geosci. Remote Sens. 61 https://doi.org/10.1109/TGRS.2023.3269622.

Zhang, T., Zhou, Y., Zhu, Z., Li, X., Asrar, G.R., 2022. A global seamless 1km resolution daily land surface temperature dataset (2003–2020). Earth Syst. Sci. Data 14, 651–664. https://doi.org/10.5194/ESSD-14-651-2022.

Zhao, W., Sánchez, N., Lu, H., Li, A., 2018. A spatial downscaling approach for the SMAP passive surface soil moisture product using random forest regression. J. Hydrol. 563, 1009–1024. https://doi.org/10.1016/j.jhydrol.2018.06.081.

Zhao, W., Wen, F., Wang, Q., Sanchez, N., Piles, M., 2021. Seamless downscaling of the ESA CCI soil moisture data at the daily scale with MODIS land products. J. Hydrol. 603, 126930 https://doi.org/10.1016/j.jhydrol.2021.126930.