



OPEN

DATA DESCRIPTOR

Seamless finer-resolution soil moisture from the synergistic merging of the FengYun-3 satellite series

Daniel F. T. Hagan^{1,2,9}, Seokhyeon Kim^{3,9}, Guojie Wang²✉, Xiaowen Ma^{2,9}, Yifan Hu², YiY. Liu⁴, Alexander Barth⁵, Haonan Liu², Waheed Ullah⁶, Isaac K. Nooni⁷ & Samuel A. Bhatti⁸

In the last years, more satellites with microwave imagers have been launched, making more observations available to obtain soil moisture estimates globally. China's endeavour has resulted in the launch of the FengYun (FY) passive microwave observations (FY-3B, C, D) capable of filling in global soil moisture data gaps. In this study, we develop a merged soil moisture dataset at a spatial resolution of 0.15° from the FY series which spans 2011 to present time (2020 in this study) by a merging technique that minimizes mean square error (MSE) using the signal-to-noise ratio of the input parent products. Here, we combine the ascending and descending observations from the three satellite observations to obtain sub-daily estimates. Finally, we averaged the merged sub-daily FY soil moisture into daily estimates and gap-fill it using a deep learning interpolation approach to reconstruct the missing days while preserving the characteristics of the Merged FY data. The results of this study aim to provide datasets that meet challenges in using global satellite soil moisture observations.

Background & Summary

In the past four decades, passive microwaves (L-, C-, and X-bands) have become crucial for retrieving global near-surface soil moisture due to their ability to penetrate clouds and vegetation^{1–4}. Major institutions like the European Space Agency (ESA) have thus focused on providing observations from passive microwave imagers onboard satellites as an alternative to *in situ* soil moisture measurements. Additionally, it has been recognized as an essential climate variable by GCOS(2010) because of its role in regulating terrestrial water, energy, and carbon cycles⁵. This recognition has led to special satellite missions like the Soil Moisture and Ocean Salinity (SMOS)^{6,7} and the Soil Moisture Active Passive (SMAP)⁸, dedicated to providing high-quality soil moisture observations globally from L-band (1.4) GHz brightness temperature observations. Recently, China's FengYun (FY-) 3B, C, and D passive microwave observations (PMWs) have extended the legacy of existing observations from 2011 to the present at sub-daily local times of 1:30, 10:15, 13:30, and 22:15⁹. However, soil moisture estimates from PMWs so far have come at coarse resolutions with spatial and temporal gaps, limiting their near real-time applications. Additional limitations include finite satellite lifespans leading to the cessation of observations and data gaps due to discontinuous revisit times at each location. This study aims to develop a soil moisture dataset usable in various fields by addressing the shortcomings of the FY3 series through two approaches. The first approach is to improve performance by merging existing finer-resolution FY3 soil moisture datasets, leveraging their individual strengths. The second approach is to fill the data gaps in the merged product to increase the available FY3

¹Hydro-Climate Extremes Lab (H-CEL), Ghent University, Department of Environment, Ghent, 9000Gent, Belgium.

²Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing, 210044, China.

³Kyung Hee University, 1732 Deogyong-daero, Department of Civil Engineering, Giheung-gu, Yongin-si, 17104, Republic of Korea. ⁴School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia. ⁵GeoHydrodynamics and Environment Research (GHER), University of Liège, Liège, Belgium.

⁶Faculty of Defense and Security, Rabdan Academy, Abu Dhabi, United Arab Emirates. ⁷School of Atmospheric Science & Remote Sensing, Wuxi University, Wuxi, 214105, China. ⁸Bacha Khan University Charsadda, Department of Geology & Geophysics, Peshawar, Pakistan. ⁹These authors contributed equally: Daniel F. T. Hagan, Seokhyeon Kim, Xiaowen Ma. ✉e-mail: gwang@nuist.edu.cn

Frequency(GHz)	Polarization	Bandwidths(MHz)	Sample size footprint(km × km)	Orbit type	Incidence angle(°)	Swath width(km)
10.65	V/H	180	51 × 85	Polar	55.4	1400
18.70	V/H	200	30 × 50	Polar	55.4	1400
23.80	V/H	400	27 × 45	Polar	55.4	1400
36.50	V/H	900	18 × 30	Polar	55.4	1400
89.00	V/H	4600	9 × 15	Polar	55.4	1400

Table 1. Sensor characteristics of the FY3 observations used in this study.

soil moisture estimates by interpolating the gaps of unobserved times using a deep learning approach, although interpolations are mainly as good as the amount and quality of available observations from which to learn.

The FY3 PMWs have recently demonstrated reliable potential for creating satellite-based long-term soil moisture data². Consequently, the CCI-SM began including the FY3 PMWs within their framework. Furthermore, Wang *et al.*¹⁰ showed in a preliminary study that combining the FY-3B and FY-3C observations could provide reliable global soil moisture estimates. This study extends the works of Hagan *et al.*² and Wang *et al.*¹⁰ by merging six sub-daily soil moisture retrievals from the three FY3 PMWs (two overpasses from each), from 2011 to 2020, to create global daily soil moisture estimates using the signal-to-noise-ratio optimization (SNR-opt) approach¹¹ for merging multiple datasets. Unlike the approach used by Hagan *et al.*², the SNR-opt does not require reference data. This implies that the merged output would be entirely intrinsic to the characteristics of the parent estimates. Here, we rely on the land parameter retrieval model (LPRM)¹², which is also the soil moisture retrieval model for the CCI-SM PMWs, to obtain the soil moisture retrievals. While previous FY3 datasets developed at a spatial resolution of 0.25°, we obtain the retrievals here at a finer spatial resolution of 0.15° following Parinussa *et al.*¹³. However, even with merging schemes like SNR-opt and other common approaches, merged datasets are still limited by gaps from the 2–4 day revisit times of polar-orbiting satellites. Whereas model-based merging schemes, outputs from statistical¹⁴ and machine learning^{15,16} gap-filling approaches are generally intrinsic to the satellite observations themselves. One such approach is the recently proposed DINCAE (Data INterpolating Convolutional Auto-Encoder), which uses a neural network with the structure of a convolutional auto-encoder to reconstruct a satellite dataset with gaps in it¹⁷. Here, we apply the DINCAE to the daily FY3 global soil moisture to fill in the gaps to generate gap-filled consistent soil moisture estimates from July 2011 to December 2020, which will be extended as more observations become available. Additionally, DINCAE provides reliable pixel-wise time-varying error estimates of the reconstructed field, which is generally difficult to obtain with traditional reconstruction methods. This gap-filled soil moisture dataset can complement existing ones like the CCI-SM by addressing limitations caused by missing observations in current records for research and application. Furthermore, we validate the merged, gap-filled datasets using global networks of *in situ* and SMAP soil moisture observations and reanalysis products to assess the relative qualities and uncertainties in the merged FY3 observations. Evaluations over different climate conditions showed that the Merged FY leveraged the strengths of the parent products to produce superior skills in terms of consistently high correlations and minimized errors.

Methods

Data Sources. *Satellite datasets.* This study uses the brightness temperatures of the microwave radiation imager (MWRI) onboard the FY-3B, FY-3C and FY-3D polar-orbiting satellites for the years 2011 to 2020, when observations from all the satellites are available⁹. FY-3B, and FY-3D were launched in November 2010 and 2017 respectively, with a common equator overpass local time of 01:30 for its descending and 13:30 for its ascending overpasses. On the other hand, FY3C, launched in September 2013, comes with a local equatorial overpass time of 10:15 for its descending and 22:15 for its ascending. Nonetheless, because of the high similarities between the FY-3B&D and FY-3C regarding the instrument specification and errors, we only need to focus on their observation time differences in the merging scheme here¹⁰. Details of the specifications of the three sensors are presented in Table 1. In this study, 0.15° soil moisture anomalies are retrieved with LPRM for FY-3B and FY-3D^{12,13,18}, while the retrievals from FY-3C are based on the retrieval setup presented by Wang *et al.*¹⁰ due to the differences in thermal equilibrium at the different overpass times noted in the aforementioned studies. To adequately understand the relative qualities in the data developed in this study, we selected an existing high-resolution satellite soil moisture data, whose spatial resolution would be close to 15 km, as an independent reference for validation. SMAP, which was launched by NASA as a special mission to provide global land surface soil moisture observations (about the top 5 cm)⁸, serves as a good candidate in this regard. The observations come at a revisit time of about 2 to 3 days at a spatial resolution of 36 km. The ascending and descending overpass observations come at 6:00 and 18:00 local time which have been suggested to be optimal times for soil moisture monitoring because of increased thermal equilibrium⁸. In this study, we use the Version 4 (SPL3SMP_E) SMAPL3 soil moisture observations that come at a spatial resolution of 9 km. These are resampled to fit the resolution of the FY3 datasets (i.e., 0.15°). Details of the SMAP data used here can be found at https://nsidc.org/data/spl3smp_e/versions/4, last access: 21 November 2022.

Reanalysis dataset. We rely on the soil moisture and precipitation datasets from the fifth generation of the global model products of the European Centre for Medium Range Weather Forecast (ERA5) for preprocessing and validating the satellite datasets. The ERA5 was developed as an improvement and successor of the ERA-Interim. Here, we use soil moisture from the offline land model product of the ERA5-Land, which comes

at a spatial resolution of 0.125° , and the atmospheric reanalysis product, which has a spatial resolution of 0.25° . Both come at an hourly temporal resolution and are resampled to a spatial resolution of 0.15° by nearest neighbor interpolation. Global and regional evaluation studies have demonstrated the reliability of these products in capturing precipitation¹⁹ and soil moisture dynamics^{20–23}. More details of these products can be found at <https://confluence.ecmwf.int/display/CKB/ERA5>, and access to data can be found at <https://cds.climate.copernicus.eu/cdsapp#!/home>, last access: 22 October 2022.

In situ soil moisture. Point-scale *in situ* datasets provide ground truth measurements of physical variables such as soil moisture. However, due to the point-scale spatial representation, they are limited beyond small-scale applications and can be used as independent references for validating satellite and model-based estimates. The international Soil Moisture Network (ISMN) provides global soil moisture measurement networks. Here, we rely on hourly ISMN datasets (depth < 10 cm), to evaluate the surface soil moisture anomalies deseasonalized by removing a 31-day moving window and the entire multiyear dataset - from the FY3 satellite observations. First, to ensure the quality of the *in situ* datasets and reduce systematic differences, we applied quality controls suggested by²⁴. Next, stations in high-density vegetation regions, indicated with the multiyear mean of the normalized difference vegetation index (NDVI > 0.85) were masked out. Finally, stations with less than 100 paired observations with the FY3 satellites are also masked out to ensure statistical significance in the comparisons, while all stations with significant ($p = 0.05$) negative correlations to two or more of the FY3 observations were removed. This resulted in a total of 507 global stations from the ISMN database.

Ancillary Datasets. Existing studies have shown that satellite and even model-based soil moisture quality varies with vegetation density^{22,25}. In this study, we use NDVI data obtained from the surface reflectance data of the sensor of the advanced very high-resolution radiometer (AVHRR), which comes at a spatial resolution of 0.05° . It can be obtained from <https://climatedataguide.ucar.edu/climate-data/ndvi-normalized-difference-vegetation-index-noaa-avhrr>. First, negative values are masked out, and the multi-year mean of monthly NDVI of the entire study period is computed. Next, we resampled it to a 0.15° spatial resolution to match the satellite datasets and scale the values between 0 and 1. NDVI values between 0.1 and 0.8 are used in this study²⁵. Soil moisture also varies as a function of climate zones. In this study, different climate zones are indicated using the Köppen-Geiger climate classification²⁶ for the present-day period (1980–2016). This data is developed from an ensemble of four topographically-corrected, high-resolution climatic maps, which provide independence from the satellite soil moisture estimates. Details of this data can be found in Beck *et al.*²⁶. We also rely on the on Climate Hazards group Infrared Precipitation with Stations (CHIRPS, v2.0) precipitation product for the Rvalue analysis²⁷. It is a quasi-global rainfall product, spanning between latitudes 50°S and 50°N , with a temporal coverage from 1981 to near-present. It incorporates both satellite and *in situ* (ground) rainfall estimates to create gridded precipitation datasets at a spatial resolution of 0.05° which have been used for several applications, including trend and drought analysis. In this study, we resample the CHIRPS data to match the spatial resolution of the FY3 datasets before proceeding to apply the Rvalue analysis. Validation studies have demonstrated the reliability of the CHIRPS product globally¹⁹ and regionally²⁸.

The Land Parameter Retrieval Model. The Land Parameter Retrieval Model (LPRM) is a commonly used approach to retrieve soil moisture from passive microwave observation and is the main retrieval algorithm for the long-term CCI PMW soil moisture retrievals^{5,29}. It simultaneously solves for soil moisture and vegetation optical depth (VOD) from low-frequency microwave observations using land surface temperatures retrieved offline from vertically polarized Ka-band³⁰ or Ku-bands³¹ as input where available. The LPRM solves for its retrievals uniquely at each location and time-step while minimizing the use of external ancillary datasets, thereby making the soil moisture retrievals intrinsic to the satellite observations^{32,33}. Since its development, the LPRM has been continuously developed to improve its retrieval outputs, including using FY3B to improve its scattering albedo and roughness parameters³⁴. In this study, we rely on the LPRM version presented by van de Schalie *et al.*²⁷ to retrieve soil moisture anomalies from the FY3 observations. So far, the LPRM has been used to retrieve soil moisture from FY3B³⁵ and FY3C¹⁰ and extensively validated globally³⁶ and regionally²⁵.

The SNR-opt merging approach. Over the years, many merging approaches have been proposed to combine satellite observations^{37–40}. The approach used here, the signal-to-noise ratio optimization (SNR-opt), was proposed by Kim *et al.*¹¹ to merge multiple satellite soil moisture observations, and can also be extended to other climate variables. The SNR-opt seeks to minimize mean square error (MSE) merging weights using the signal-to-noise ratio of the parent products. Unlike other commonly used merging approaches like the triple collocation (TC) which merges three inputs at a time⁴⁰, or the linear combination approach, which combines two inputs⁴¹, the SNR-opt can be applied to an arbitrary number of inputs at a time. Furthermore, the SNR-opt may or may not be used with reference data, making it easily applicable to a wide range of uses.

A brief introduction of the SNR-opt formalism is given as follows: Given N number of parent datasets $(x_1, x_2, \dots, x_N) = \mathbf{x}$, a set of real numbers as $\mathbf{x} = y\mathbf{1} + \mathbf{e}$ with additive noise $\mathbf{e}$, the unknown quantity (e.g. soil moisture) y , a real number, which is the weighted average can be predicted as $\hat{y} = \mathbf{x}^T \mathbf{u}$. $\mathbf{u}$ is a vector that minimizes the MSE of $\hat{y}$ which is determined by solving the problem

$$\min f(\mathbf{u}) = E(\mathbf{x}^T \mathbf{u} - y)^2$$

using the solution of the optimum weights

$$\mathbf{u}^* = E(\mathbf{x}\mathbf{x}^T)^{(-1)}E(y\mathbf{x})$$

where $E(\mathbf{x}\mathbf{x}^T)$ is always positive definite. We note that $E(\mathbf{x}\mathbf{x}^T) = E(\mathbf{e}\mathbf{e}^T) + E(y^2)\mathbf{1}\mathbf{1}^T$ and $E(y\mathbf{x}) = E(y^2)\mathbf{1}$. Thus, the above solution for optimum weights can be rewritten as

$$\mathbf{u}^* = (\mathbf{N} + \mathbf{1}\mathbf{1}^T)^{(-1)}\mathbf{1}$$

where $\mathbf{N} = E(\mathbf{e}\mathbf{e}^T)/E(y^2)$ ¹¹. noted that this could be interpreted as the coefficient of a MISO (multiple-input single-output) Wiener filter which assumes noise spectra and flat signal. Furthermore, the derivation of $\mathbf{u}^*$ shows that the optimum weights depend on a noise-to-signal ratios matrix expressed as $E(\mathbf{e}\mathbf{e}^T)/E(y^2)$. More generally, for the case where the scaling factor of $\mathbf{x}$ is not $\mathbf{1}$ but $\mathbf{a}$, the solution can be presented as

$$\mathbf{u}^* = (\mathbf{N} + \mathbf{a}\mathbf{a}^T)^{(-1)}\mathbf{a}$$

Since the ground truth is not fully available for spatiotemporally concurrent satellite data, $\mathbf{N}$ and $\mathbf{a}$ are not known in general¹¹. proposed an iterative optimization method, SNR Estimation (SNR-est). SNR-est can estimate $\mathbf{N}$ and $\mathbf{a}$ without the ground truth by assuming that the off-diagonal elements of $\mathbf{N}$ are small and the signal power $E(y^2)$ is known. In this study, $E(y^2)$ is obtained from the average of the parent FY3 satellite estimates during the period of the data merging. All the merging in the study are based on the anomalies of the FY3 satellite observations, after which we combined the merged anomalies and then add the ERA5_land multiyear climatology to produce the final result. For more details of the approach, including examples with soil moisture and land surface temperatures, readers are referred to Kim *et al.*¹¹.

The deep learning gap-filling approach. Several approaches have been proposed for filling gaps in satellite datasets in the last two decades based on statistical formalisms and machine learning/deep learning. In the last decade, the latter has become a preferred choice because it can handle more complex problems of gap-filling problems. In fact, some studies have even shown that deep learning approaches may be reliable for tackling climate modelling problems, such as those obtained from numerical simulations with climate models^{42–44}. Along these lines, Barth *et al.*¹⁷ proposed a neural network in the form of a convolutional encoder that we can train to reconstruct satellite observations with missing observations called the Data INterpolation Convolutional Auto-Encoder (DINCAE). This builds on a previous approach, which relies on empirical orthogonal function instead of deep learning. Both approaches aim to reduce the dimensional subspace of the input data, where the auto-encoder is a network able to handle complex nonlinear problems such as those found in soil moisture dynamics. The DINCAE reconstructs the data by inputting the satellite observation and its expected error variance and outputting the reconstructed data and its error variance, which is useful in data assimilation studies or any other application whose result dependent on the accuracy of the reconstructed data. Missing observations are estimated by the neural network trained by maximizing the likelihood of the observations, equivalent to minimizing the negative log-likelihood. Since this is performed on one physical parameter, soil moisture, we do not introduce uncertainties from other variables, such as those found with multivariate gap-filling approaches. In a recent study, the DINCAE was applied to CCI soil moisture and the merged data by Hagan *et al.*² to gap-fill their missing days, demonstrating a significantly high consistency between the original data and the gap-filled data⁴⁵. Due to computational constraints, we separated the globe into six continental regions following Hu *et al.*⁴⁵.

The Rvalue technique. Independent evaluation methods are useful for providing a fair assessment of the strengths and limitations of datasets. In this paper, we rely on the Rvalue technique initially developed by Crow & Zhan⁴⁶ and later adapted by Crow *et al.*⁴⁷ to provide a comprehensive independent areal evaluation of a temporal correlation-based skill of the satellite soil moisture anomalies. It is based on the relationship between precipitation events and the subsequent changes found in soil moisture. Rvalue relies on contrasts in the quality of rainfall datasets to assess the degree to which analysis increments from a sequential assimilation of soil moisture into a simple water balance model could accurately compensate for known rainfall errors. In this study, we follow the approach presented by Parinussa *et al.*⁴⁸, where we artificially deteriorate the CHIRPS precipitation product to generate these rainfall errors. Here, Rvalue closer to 0 indicates a poor performance while higher values indicate good skill, which represent not only the temporal dynamics of the soil moisture product, but also its sensitivity to rainfall events. Nonetheless, there are some well-known limitations of the Rvalue technique. In extremely arid climate regimes, the technique may be very unreliable due to an insufficient number of precipitation events in these regions. Thus, we mask these regions (NDVI < 0.1) in our analysis following earlier studies^{25,49}. Additionally, like satellite soil moisture retrievals, Rvalue skill also deteriorates under dense vegetation conditions. Here as well, we mask out soil moisture values in these regions in our analysis. More details on the mathematical formulation of the technique can be found in Crow *et al.*⁴⁷ and Parinussa *et al.*⁵⁰. Finally, to provide robust statistical results from this analysis, we only apply it to the FY3 datasets that have more than 3 years of coverage in our study period. Therefore, we only apply it to the ascending and descending products of FY-3B, FY-3C and the merged product between 2013 and 2016. The goal here is to verify how the merging scheme preserves the sensitivity to rainfall events in the merged product which are very important data assimilation applications. Thus, we focus on the similarities between the merged and parent products.

Along with response to precipitation, soil moisture drydown patterns, which represent changes in soil moisture following the infiltration of precipitation, are very important to soil moisture characterization. They serve as a good indicator of soil moisture dynamics, quantifying its response to the boundary layer processes such as changes evaporation, land cover type and soil drainage⁵¹. To assess these drydowns in the merged data, we

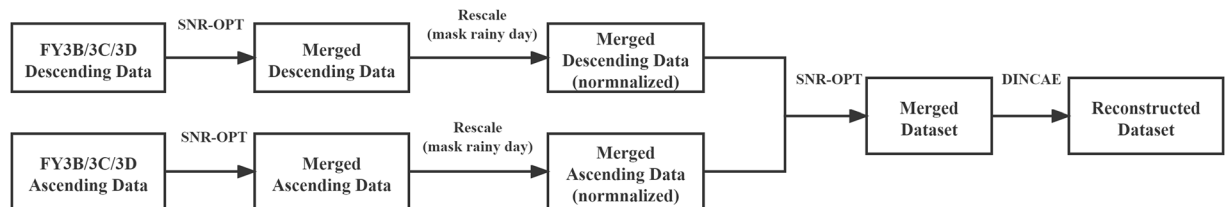


Fig. 1 A flowchart of FY-3 soil moisture data merging.

calculate the rate of soil moisture drydowns (τ) following the approach presented by McColl *et al.*⁵² which does not require any precipitation inputs, making it independent from the Rvalue analysis. τ is computed for the *in situ* observations, the merged and reconstructed data to understand, firstly how drydowns are captured in the FY3 products, and how the reconstruction would influence changes to the soil moisture dynamics following precipitation events.

The distance between indices of simulation and observation (DISO) evaluation. It is essential to accurately understand the strengths and uncertainties of the merged data developed in this study. The common practice for this is calculating error and performance metrics like the root mean square errors (RMSE), standard deviations and correlation coefficients (R), which quantify centered similarity and differences between the datasets and a chosen reference. Previous attempts have been made to develop metrics that harmonize the different error and performance metrics that provide summarized datasets evaluations⁵³. Recently, Hu *et al.*⁵⁴ developed a harmonized metric, the distance between indices of simulation and observation (DISO), which combines different statistical metrics, including R, average errors, and RMSE based on the distance between the simulated model and observed field. In this study, we rely on the DISO to quantify a holistic evaluation of the merged data. Details of the theoretical framework and assumptions of the DISO can be found in Hu *et al.*⁵⁴ and Zhou *et al.*⁵⁵.

Data preprocessing and architecture of the merging process. The architecture of the entire process is shown in Fig. 1. Firstly, soil moisture is retrieved using the X-band channel observations from the three FY3 satellites using the LPRM at a spatial resolution of 0.15°. Here, frozen conditions, pixels over open water locations and regions where NDVI is greater than 0.8 are all masked. We also mask out the whole of Greenland. From there, rainfall days are also temporally removed to compute the optimum weights for SNR-opt based on daily rainfall in the ERA5 precipitation data. Next, the ascending and descending observations with rainfall days are merged separately to obtain ascending-merged and descending-merged observations of the FY3 PMWs using the SNR-opt approach. From there, the two merged datasets are normalized to the daily average of the FY-3B (with masked-out rainfall days) and merged to obtain daily averaged soil moisture observations. Finally, we interpolate the daily FY3 soil moisture datasets to gap-fill the missing observations and cross-validate the reconstructed data with the gappy merged and *in situ* soil moisture. All the datasets used in this study have been resampled to match the spatial resolution of the FY3 satellite observations.

Data Records

The dataset is available at Zenodo⁵⁶. It comprises two main outputs: the reconstructed merged daily averages (FY3_Reconstructed_<year>) and uncertainties (FY3_ErVar_<year>) for each time and grip of the reconstructed data which can be very useful for data assimilation applications. For each data, the datasets are provided in NetCDF formats for each year, stored in zip formats for each year for each data due to the large sizes of the datasets. The current data spans 2011 to 2020 and will be updated to the present year at the end of 2024.

Technical Validation

Evaluation of the output of the merging process: the merged ascending and descending FY3 observations. Before proceeding to the final merged daily FY3 observations, we first evaluate the interim merged ascending and descending observations that provide potential sub-daily soil moisture estimates since uncertainties at this stage will feed into the final merged observations. Figure 2 presents the time series of the ascending (top panel) and descending (bottom panel) observations from the individual satellites and merged soil moisture for the entire period for a point (43.5°N, 119.5°W). Both panels demonstrate that the merged product covers the entire period and combines the three FY3 datasets (blue/inblack dots) within a reasonable range to capture soil moisture dynamics. The temporal coverage begins with FY-3B from July 2011 to September 2013, where FY-3C observations (orange/inblack dots) become available, and extends the coverage to 2019, where observations from all three PMWs are available briefly. Beyond that, FY-3B stops being operational and data availability depends on FY-3C and FY-3D (green/inblack dots) until mid-2020, after which only FY-3D provides observations for soil moisture retrieval. Thus, there is a sufficient continuity between the three products, which was unfortunately absent in the AMSRE-AMSR2 framework in 2011, which is helpful when developing consistent long-term datasets. We also note that because the SNR-opt can handle any number of input data to merge, the different availabilities are easily incorporated into the merging scheme, as shown in Fig. 2.

Evaluation with *in situ* soil moisture. Relative qualities in satellite soil moisture retrievals have been shown to vary across different climate conditions²². Here, we indicate the different climate conditions based on precipitation and temperature climatology variability²⁶, which provides independent climate zones from the FY3 soil

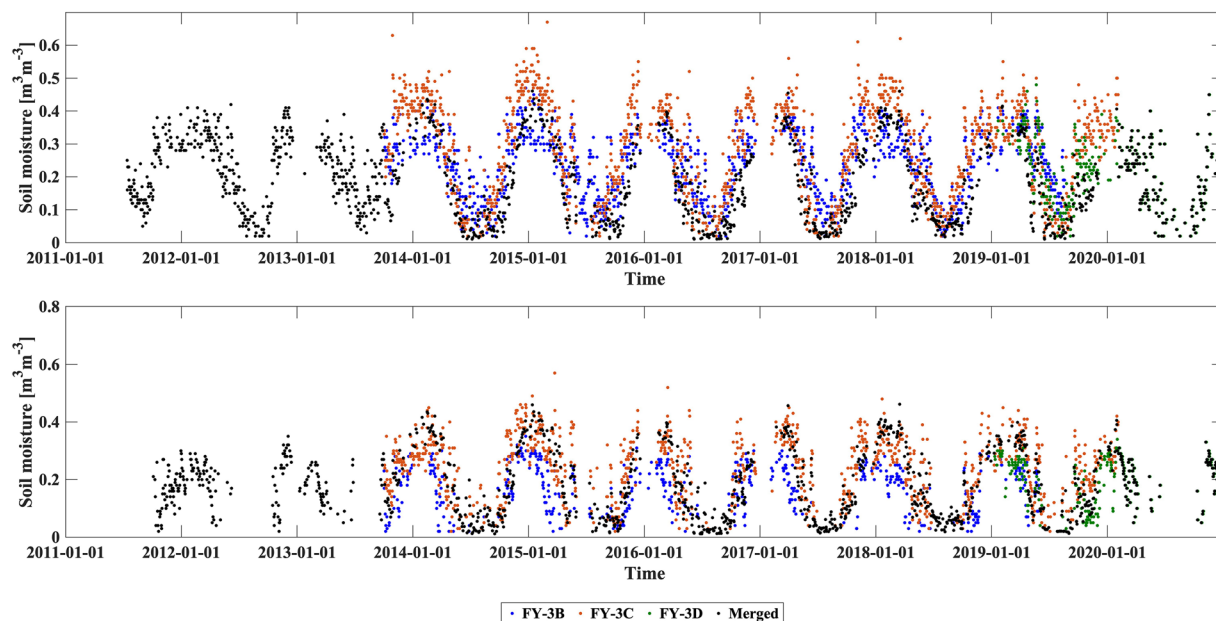


Fig. 2 Time series of a selected location (43.5°N, 119.5°W) showing the (a) descending observations of the three FY3 and their merged soil moisture product based on SNR-opt(top panel), and (b) same as (a) but for ascending observations (bottom panel).

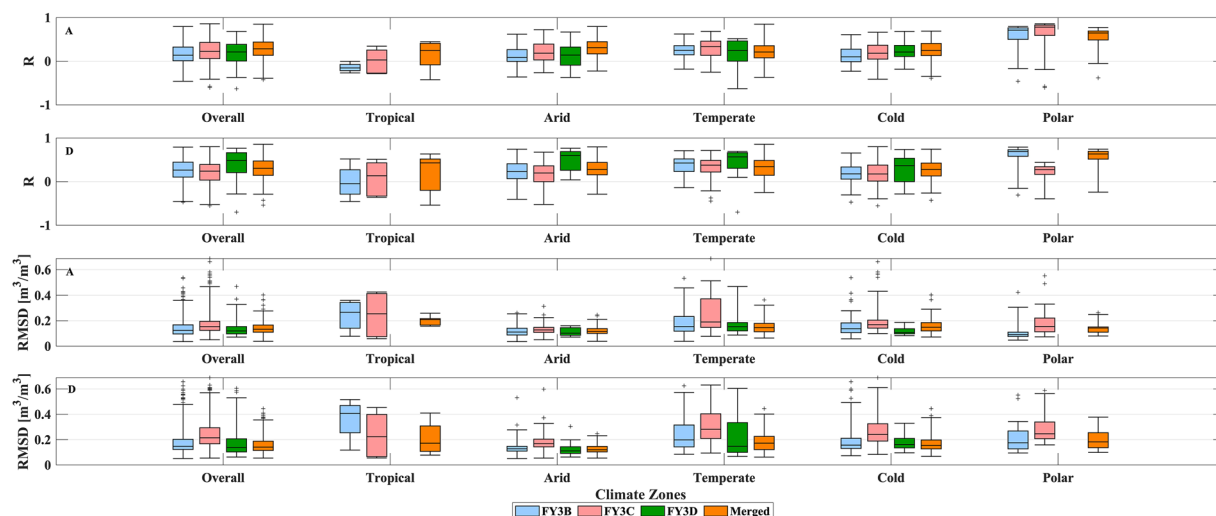


Fig. 3 Inter-comparisons between the ISMN *in situ* soil moisture and the FY3 merged and parent (FY-3B, FY-3C, FY-3D) products for ascending (A) and descending (D) observations showing (a,b) correlations of A and D (top panels) and (c,d) RMSE of A and D (bottom panels) for five global climate zones: Tropical, arid, temperature, cold and polar regions.

moisture data. Figure 3 shows a comparison between the ascending and descending FY3 Merged and parent (FY-3B, FY-3C, FY-3D) soil moisture products with the ISMN *in situ* datasets across five climate zones: tropical, arid, temperature cold and polar regions. For each climate zone, *in situ* stations found there are matched to the closest pixels of the satellite datasets. Where more than one station is found in a pixel, they are averaged. Here we computed the Pearson correlation coefficient (R) and the RMSE (root mean square error) between the *in situ* data and their matched satellite data time series from 2013 to 2019 where more than one product is available at a time. Overall, the results show that the merged data has harnessed the strengths of the parent product to obtain a harmonized product with an overall median correlation coefficient of about 0.5 and median RMSE of 0.08 m³/m³, which are of higher qualities relative to the parent products. In the tropical, arid and cold regions, the ascending merged product shows the best correlations with the *in situ* observations, while in the descending results, the merged results consistently show higher median correlations than FY-3B and FY-3C. We note that the small sample size of FY-3D could explain why FY-3D appears to have higher correlations than the merged. While the merged product covers the entire period from 2013 to 2019, FY-3D only covers 2019. Additionally,

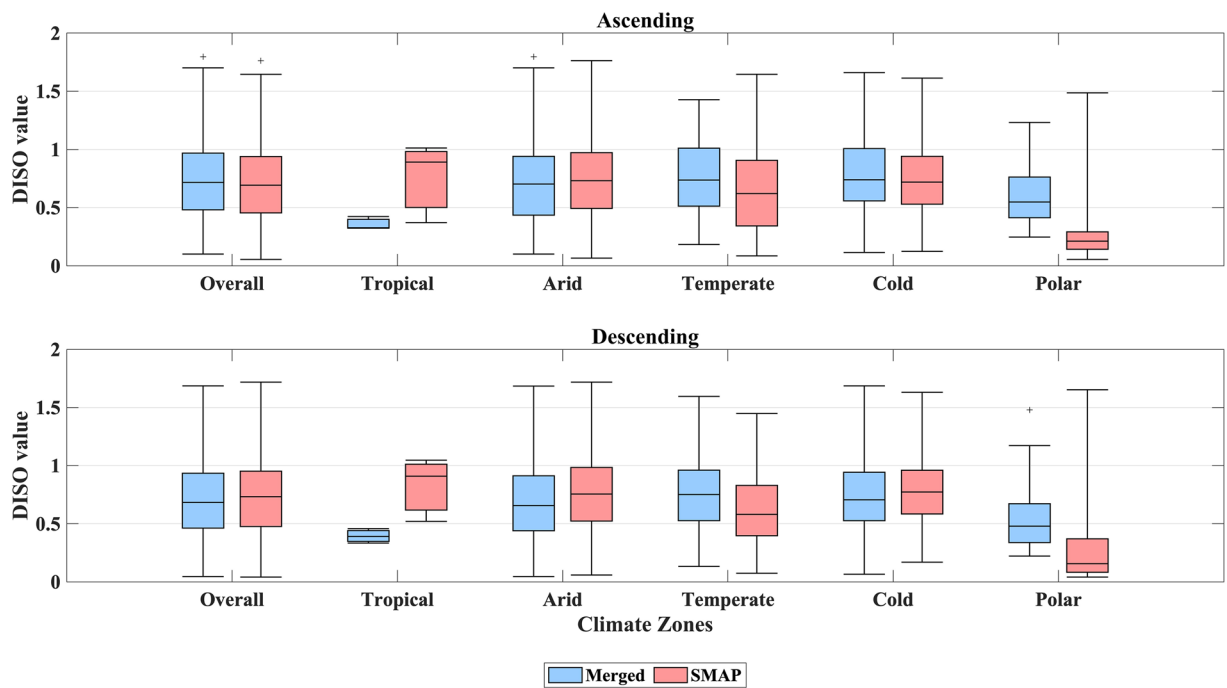


Fig. 4 Evaluation of FY3 Merged soil moisture and SMAP soil moisture based on the DISO approach with the ISMN *in situ* soil moisture for (a) ascending (A, top panel) and (b) descending (D, bottom panel) observations for five global climate zones: Tropical, arid, temperate, cold and polar regions.

the missing results for FY-3C are because the sample size was below the set threshold used in this study to ensure statistical rigour. For both groups of observations in Fig. 3a,b, the order of performance of the merged product, and relatively for the parent products as well, increases from the tropical, arid, temperate, cold to polar.

Figure 3c,d show the RMSE across the different climate zones. The lowest errors are found in the arid regions, while the largest errors are found in the temperate and tropical regions. The results indicate that the merged product consistently shows the smallest error spread and thus, the lowest uncertainties. While the RMSE range in the parent products is between 0.05 m³/m³ and 0.6 m³/m³, the RMSE range for the merged falls between 0.05 m³/m³ and 0.4 m³/m³. These results also demonstrate that the merging scheme leverages the strengths of the parent products to obtain a merged product with a higher skill. Preliminary comparisons of the Merged and the three parent products where the sample used for the Merged is the same as the individual parent products (not shown) have also shown higher qualities in the Merged.

As a further step, we provide comparisons of the merged with SMAP soil moisture ascending and descending observations. Figure 4 shows DISO results of both satellite datasets with the ISMN *in situ* soil moisture measurements across different climate zones globally. DISO values closer to 0 indicate better performance. In general, the boxplots in Fig. 4a,b show a similar pattern in the ascending and descending observations. Both datasets show very similar qualities across different climate regions with the largest differences found in the tropical and polar regions, where the merged has higher qualities in the former and SMAP has higher qualities in the latter. The similarities and differences in their performances demonstrate how these two datasets can be used to complement each other. The mean DISO values, although have very small differences, also show that SMAP generally shows higher qualities in the ascending (Fig. 4a) while the Merged shows higher qualities in the descending observations (Fig. 4b). Nonetheless, the smaller errors, especially over the polar regions, might also be partially influenced by the ERA5 climatologies in the merged data.

Evaluation of merging sensitivity to rainfall. To provide more comprehensive areal extent uncertainties in the merged ascending and descending observations, we use the independent Rvalue metric in this section to assess the consistency of the sensitivity of the merged product anomalies to rainfall in a cross-comparison with the parent products. Additionally, we compare the performance of the Merged with SMAP soil moisture, which is a trusted existing satellite product. Here, we aim to understand the potential usability of the merged product in, for example, land-atmosphere interactions and data assimilation studies based on the Rvalue metric (Fig. 5), and also to identify regions where the merged product here could complement the use of existing products such as the SMAP soil moisture (Fig. 4).

Figure 5 shows the cross-comparison of the Rvalue of the Parent products (FY-3B and FY-3C) and the Merged product. The cross-comparison aims to assess the potential deterioration of the sensitivity to rainfall of the merging scheme. Rainfall events occur at certain times of the day, which may or may not be captured in satellite observations. Generally, averaging multiple products could potentially mask or reduce the sensitivity to rainfall, especially where the parent products come at different times of the day. Since this quality is very important when these datasets are used in application studies, it is vital to quantify the extent to which an

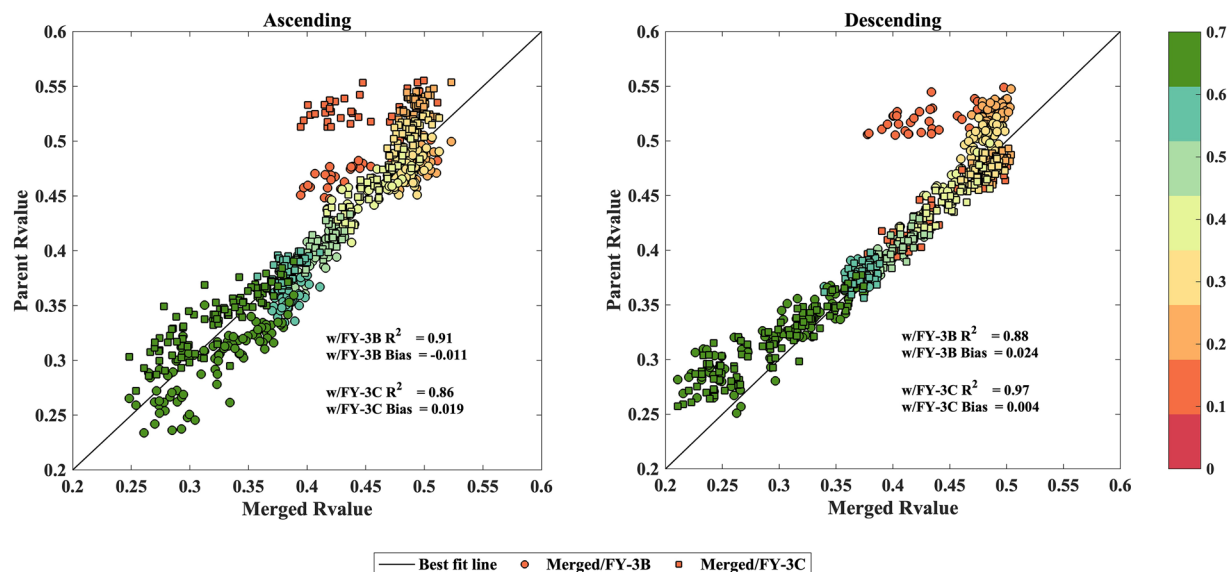


Fig. 5 Rvalue cross-comparison of the Merged FY3 product with parent products (FY-3B, FY-3C) for the (a) ascending (left panel) and (b) descending overpasses (right panel) across different vegetation densities (green shades).

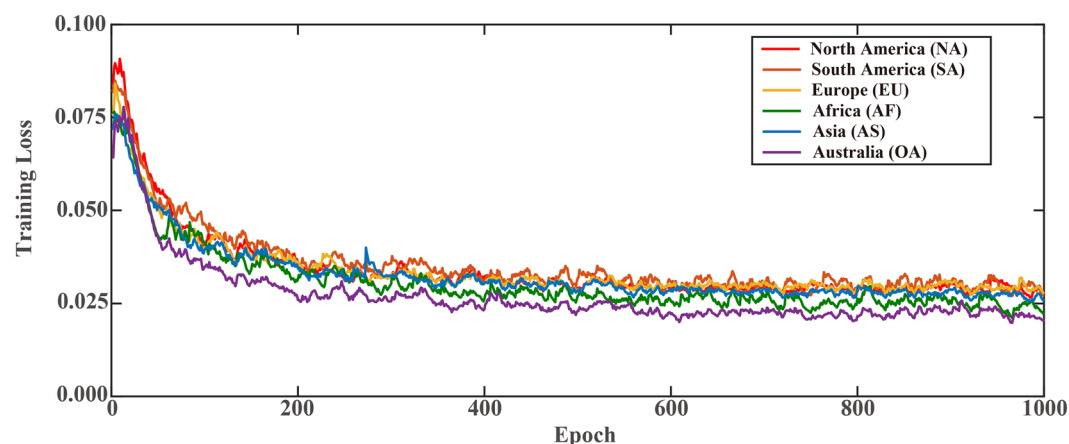


Fig. 6 Training losses of the FY-3 soil moisture.

averaging or merging scheme might have reduced the sensitivity to rainfall events. Figure 5 demonstrates that the merging does not negatively impact the Rvalue skill of the parent data for both the ascending and descending paths. The comparisons show R^2 values ranging from 0.88 to as high as 0.97, and difference between the Merged and each parent product also ranging between absolute values of 0.004 to 0.024. The descending observations appear to show the highest consistencies with the Merged (Fig. 5b). We also find that low (high) Rvalue are generally found over high (low) vegetation density regions (markers with dark green fills) for both sets of observations, which is consistent with findings of Rvalue patterns in earlier studies^{25,49}. Thus, these results provide confidence in potential applications with the merged product where sensitivity to rainfall is necessary.

Merging sub-dailies into daily averages and gap-filling. *Evaluation of the gap-filling process.* Daily soil moisture estimates are useful for long-term applications such as land-atmosphere interactions, trend and drought analysis. However, simple averages of sub-daily estimates often do not yield the most optimal daily estimates, in which case the SNR-opts approach could provide more optimal averages¹¹. So far in this study, the merged FY3 sub-daily soil moisture observations show that the qualities in their spatial and temporal distributions are relatively similar across the globe, allowing them to be easily merged into one framework of daily soil moisture estimates (Fig. 1). After merging the FY3 ascending and descending observations, we proceed to understand the limitations of the merged daily estimates. The results show that more observations are available in the higher northern and southern latitudes which range from about 30% to about 60% over Northern Europe. On the

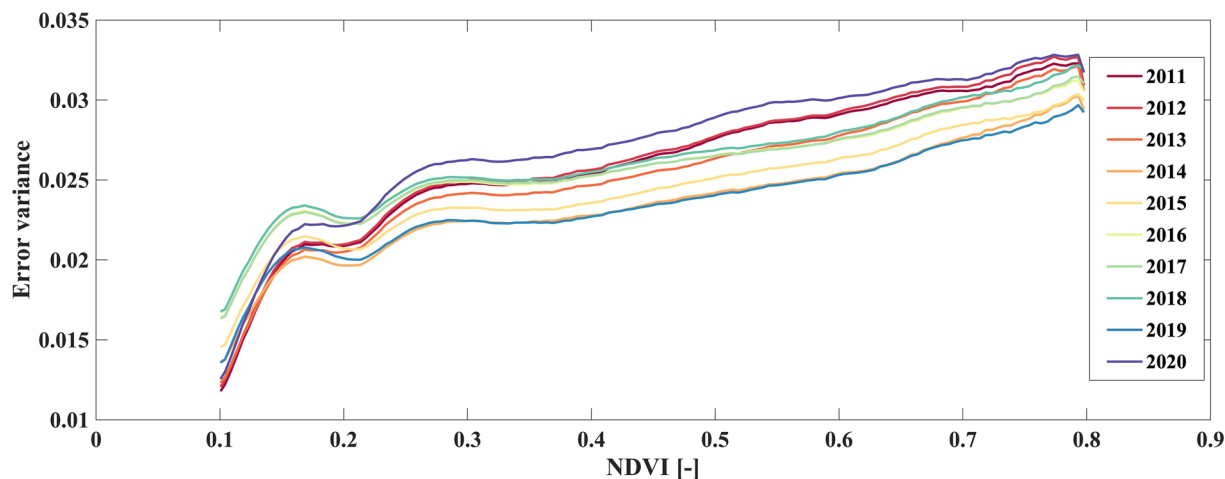


Fig. 7 Error variance for the reconstructed soil moisture plotted by years from 2011 to 2020 binned across NDVI scenarios. NDVI < 0.1 are masked out since satellite soil moisture has little to no sensitivity due to insufficient precipitation. NDVI > 0.8 are also masked out since satellite soil moisture loses skill resulting in no noisy sensitivity.

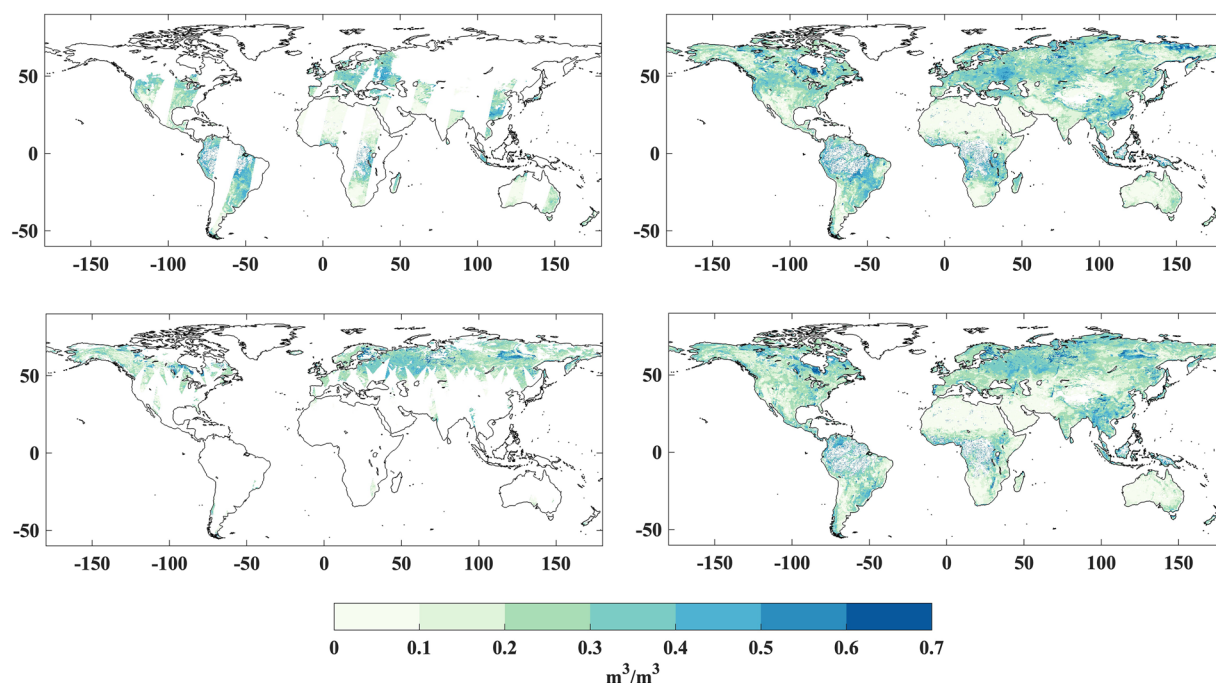


Fig. 8 The merged daily FY3 soil moisture estimates with missing observations and the reconstructed FY3 daily soil moisture estimates with the DINCAE for (a,b) 2017-03-20 respectively (top panels) and (c,d) 2019-09-22 respectively (bottom panels).

other hand, the tropical regions appear to have less than 20% of available observations. This implies every location has missing observations and would benefit from gap-filling to obtain spatio-temporally consistent observations.

The DINCAEC is used to gap-fill the merged FY3 daily soil moisture observations for the entire period and each pixel. In the gap-filling approach, we use a neural network with the structure of a convolutional auto-encoder derived from the U-Net architecture¹⁷ to train the gap-filling. The input data given the network correspond to three consecutive time instances and their corresponding presence-absence mask, as well as the longitude and latitude of every pixel. The year-day (days since the start of the year) is also included as input to account for the strong seasonality of the input data. The configuration of the neural network from Barth *et al.*¹⁷ was also found to be suitable for the present use case. The neural network is composed of 5 convolutional layers followed by average pooling (encoder network) and two fully connected layers (with drop-out during training) and 5 convolutional layers followed by up-sampling layers (decoder network). Skip-connections between the encoder and decoder avoid excessive smoothing of the input data during the reconstruction. The output of the

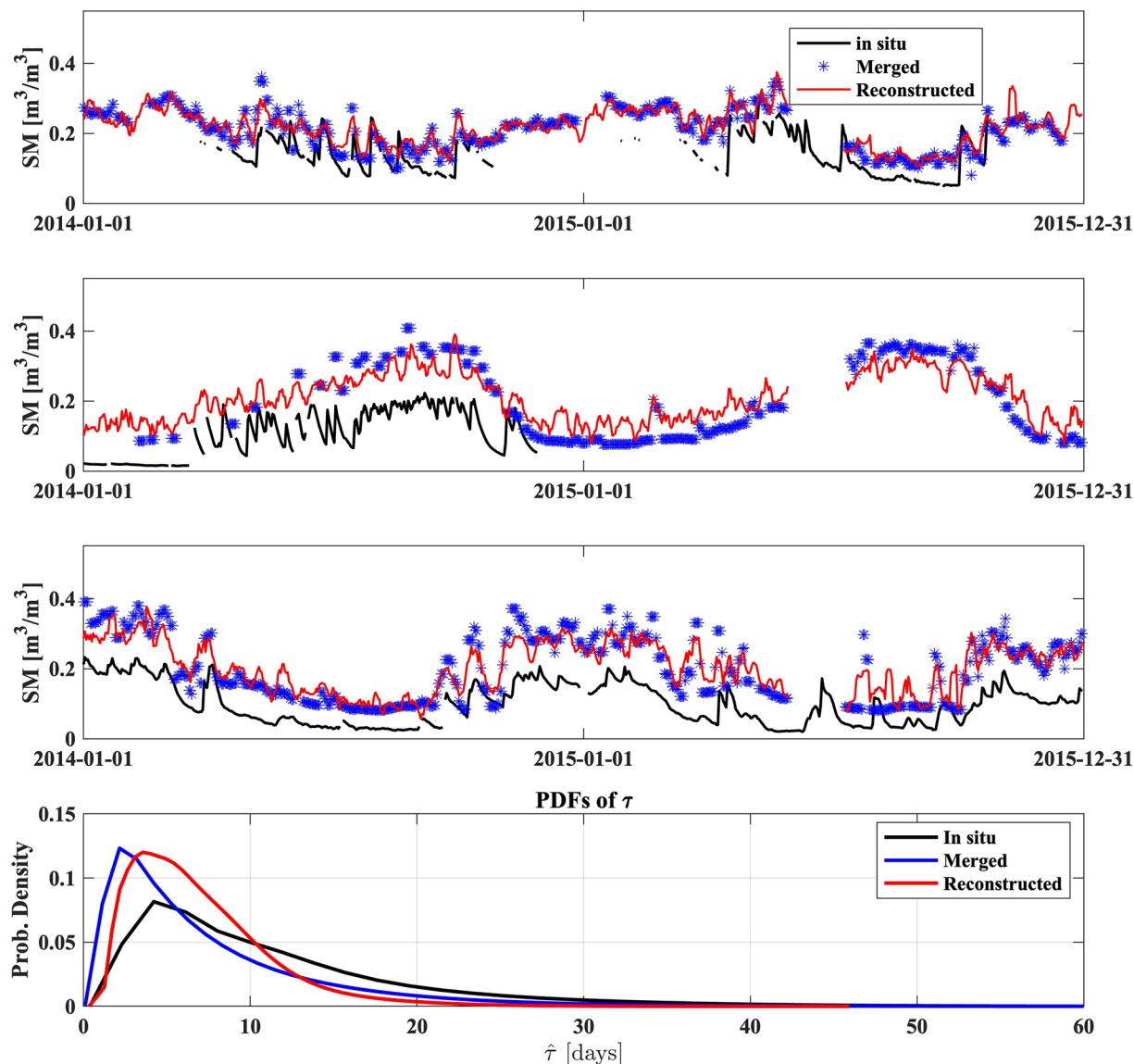


Fig. 9 Time series of selected locations (a) Weld, Colorado, United States (40.86°N, 104.74°E), (b) Bellefoungou, Benin (9.79°N, 1.71°W) and (c) Barranco de las Vacas, Gran Canaria, Spain (41.34°N, 5.22°E) of the *in situ* datasets, merged FY3 data and reconstructed data (unit: m³m⁻³). (d) Estimated probability density function (PDF) of median soil moisture drydowns ($\hat{\tau}$) of matched FY3 pixels with ISMN stations (bottom panel).

neural network is the full gap-free image and a corresponding error estimate for every pixel. The cost function is based on the likelihood of the observed values as described in more detail in Barth *et al.*¹⁷. The training dataset is split into batches of 50 instances (so-called mini-batches). The gradient of the cost-function is computed for every mini-batch. The neural network is trained for 1000 epochs with a learning rate of 0.001 using the ADAM optimizer⁵⁷. To do this, epochs are chosen from 1 to 1000 to determine the most suitable epoch for which we obtain high qualities in the training set. Figure 6 shows the training loss for different epochs. As all weights of the neural network are initialized randomly expected, initially the training loss is relatively large and gradually decreases much lower epochs obtain higher training losses up to about 300 epochs where we obtain minimum losses for all the regions. Regions over Australia and South-east Asia (OA) obtain the lowest losses followed by the African continent (AF). The other regions appear to have higher losses, especially the North American (NA) region. Based on these analysis, a minimum epoch is selected for each region which we use to gap-fill the data. Regions of dense vegetation (NDVI>0.8) are masked out of the gap-filled FY3 soil moisture since the noise-to-signal ratio in this region is relatively high and does not reflect soil moisture conditions at the land surface. These regions are mostly found in the Amazon and Congo basins. We further examine the error variances from the interpolation process of the Merged FY3 datasets for each year of the study from 2011 to 2020 as shown in Fig. 7. The results indicate that the errors increase as a function vegetation density, which has been noted in earlier studies like Parinussa *et al.*⁵⁸. Qualities in satellite soil moisture datasets from LPRM, and other retrieval models, are generally found to decrease from arid (low vegetation) to humid (high vegetation) regions^{2,18}. Thus, these are not necessarily artifacts of the interpolation approach, but the retrieval model, LPRM, used here. In

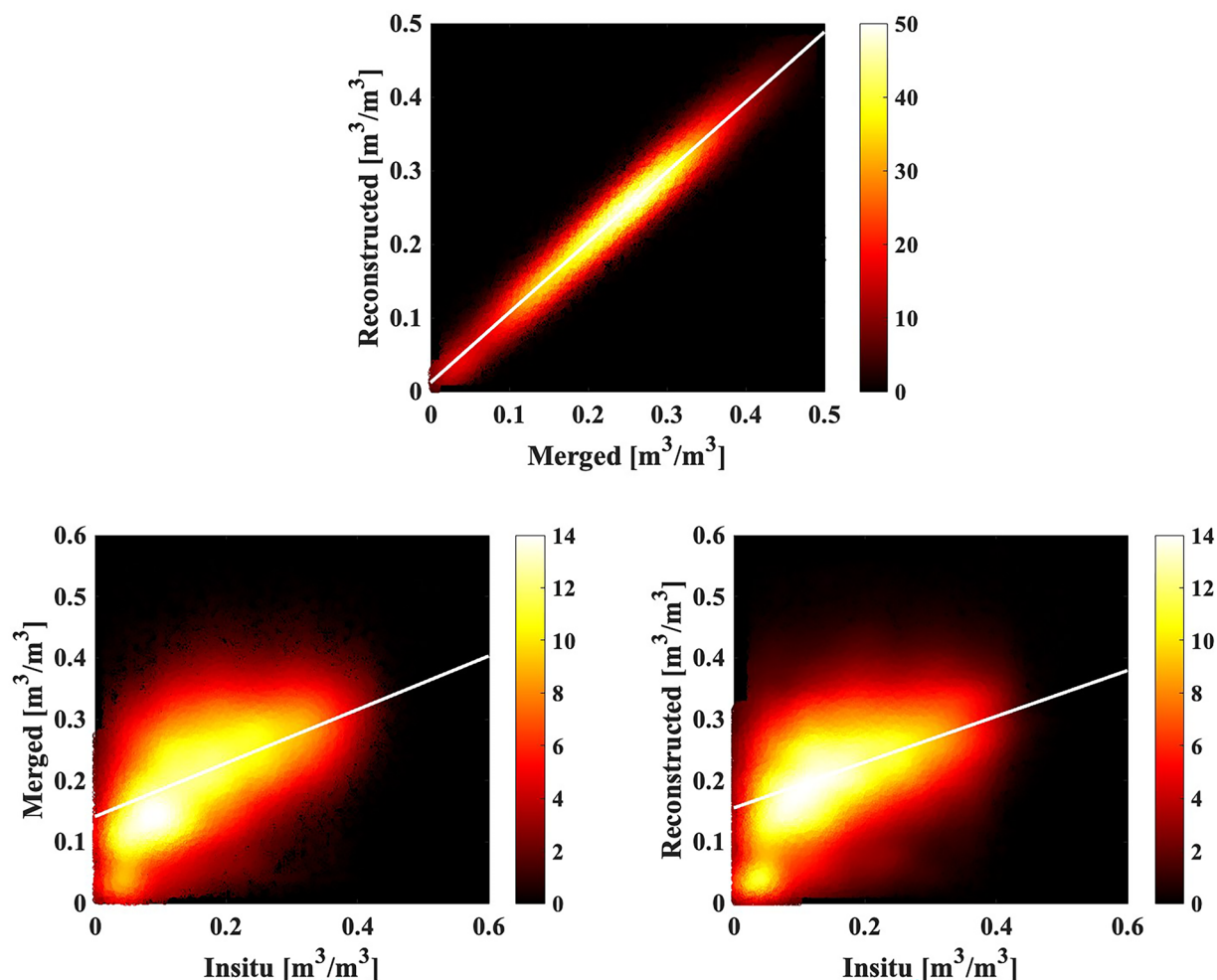


Fig. 10 Cross-validation of the merged soil moisture and the reconstructed soil moisture for with (a) each other and (b,c) the *in situ* data. The color gradient represents the data cluster density.

Fig. 7, 2011, 2012 and 2020 have the largest errors because these are years where there was only one parent product in the Merged. As a result, the product does not benefit from leveraging from another source where there are lower retrieval qualities. The other years which have 2 or more inputs have higher qualities, especially for 2019 where there were three inputs. Since these errors are provided for each grid point for each day, they can be used in data assimilation schemes when assimilating this merged product in climate models.

Intercomparison of the daily merged and reconstructed merged soil moisture. To obtain a first glance at the success of the reconstructed FY3 soil moisture, we select two time slots (2017-03-20 and 2019-09-22) to display their global results as shown in Fig. 8 for the original merged (Fig. 8a,c) and reconstructed (Fig. 8b,d). The results show that The DINAEC has reconstructed the FY3 soil moisture estimates to obtain global soil moisture distribution patterns. The high-latitude regions of the northern hemisphere and other humid regions such as South East China have the highest soil moisture content, while the desert areas have a very low soil moisture content. Transition regions between arid and humid climate conditions are also well captured in Fig. 8b,d from very limited observations in left panels in Fig. 8a,c. However, this comes at a cost of increased variances in the reconstructed time series which could introduce artificial spikes impact soil moisture drydown variations.

Additionally, we show time series plots of the soil moisture for both original and reconstructed merged FY3 daily soil moisture for three points where *in situ* observations are available in Fig. 9c. Here, the daily averaged *in situ* soil moisture are obtained by averaging the hourly soil moisture measurements to obtain daily estimates. Figure 9a–c show that firstly, the daily merged FY3 soil moisture observation capture daily soil moisture sufficiently well indicated by the alignment of the black plots with the green plots. After gap filling with the DINCAE, the reconstructed series (blue) also shows very close alignment with the originally merged observations and the *in situ* observations, although it inherits a little more variance due to the gap-filling process. This shows that the reconstructed data is consistent with the merged FY3 daily soil moisture observations. The temporal patterns in Fig. 9a–c also agree with expected seasonal variations in soil moisture in the region. Finally, the PDF plots of the median soil moisture drydowns also demonstrate that the FY3 datasets have comparably similar drydown patterns with the *in situ* time series (Fig. 9d). All soil moisture estimates show dominant median drydowns from 2 to 4 days.

Cross-validation with *in situ* soil moisture. We further explore the consistency between the reconstructed FY3 soil moisture data in a cross-validation analysis as shown in Fig. 10. Figure 10a is a direct cross-validation that demonstrates that the two datasets are very consistent with each other. This implies that negligible changes have appeared in the regions that initially had data available. Almost all points in both datasets lie on the 1:1 line in Fig. 10a. This demonstrates that the climate conditions found in the gappy merged data are fully present in the reconstructed with dry places remaining dry and wet places remaining wet while climate zones in between also fit each other. Both datasets range from 0.02 m³/m³ to about 0.48 m³/m³. In Fig. 10b,c the two datasets are cross-validated against *in situ* datasets. Here, we observe that the deviations from the line of best fit are very similar. Figure 10b,c show that some semi-arid regions in the satellite datasets correspond to arid regions in the *in situ* datasets. These are actually the tropical regions in Fig. 3, where we find low correlations and large errors. We also observe that fewer very arid conditions are present in the reconstructed datasets as the density of arid points reduces in Fig. 10c. The implications of this for drought analysis is that in arid places, based on the *in situ* data, the FY3 may underestimate dry conditions while overestimating in humid regions. This limitation was also found in the drydown results, especially in the reconstructed data. Nonetheless, both datasets show very high consistency even with this independent *in situ* data, which demonstrates the success of the interpolation process.

Usage Notes

These gap-filled merged and its error variances FY3 global soil moisture datasets are readily downloadable by potential users⁵⁶. The datasets serve as a proxy for consistent satellite-only soil moisture datasets from 2011 and eventually, to the present day after the updates. They are useful for both research and operational purposes such as drought and flood analysis in any location of the global land which require finer than the usual 0.25° degree spatial resolution. The gap-filled soil moisture data could also be used as baseline for assessing data assimilation routines targeted at gap filling soil moisture states. We have used the merged gappy data as the baseline for verifying that the gap-filled data preserves all the intrinsic qualities of the original observations. Should a need arise for using the original merged gappy data, which is not the final product of this study, users can download that here: <https://doi.org/10.5281/zenodo.11500736>.

Code availability

The codes for the merging scheme of the datasets can be found here: <https://github.com/steelpl/snr-opt>. Codes for the gap-filling can also be downloaded here: <https://github.com/gher-uliege/DINCAE.jl>. Further questions can be directed to Daniel F.T. Hagan (daniel.hagan@ugent.be).

Received: 14 June 2024; Accepted: 22 May 2025;

Published online: 11 June 2025

References

- Dorigo, W. *et al.* ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment* **203**, 185–215, <https://doi.org/10.1016/j.rse.2017.07.001> (2017).
- Hagan, D. F. T. *et al.* Maximizing Temporal Correlations in Long-Term Global Satellite Soil Moisture Data-Merging. *Remote Sensing* **12**(13), 2164 (2020).
- Karthikeyan, L., Pan, M., Wanders, N., Kumar, D. N. & Wood, E. F. Four decades of microwave satellite soil moisture observations: Part 2. Product validation and inter-satellite comparisons. *Advances in Water Resources* **109**, 236–252, <https://doi.org/10.1016/j.advwatres.2017.09.010> (2017).
- Liu, Y. Y. *et al.* Trend-preserving blending of passive and active microwave soil moisture retrievals. *Remote Sensing of Environment* **123**, 280–297, <https://doi.org/10.1016/j.rse.2012.03.014> (2012).
- Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W. & Dorigo, W. Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology. *Earth Syst. Sci. Data* **11**(2), 717–739, <https://doi.org/10.5194/essd-11-717-2019> (2019).
- Kerr, Y. H. *et al.* Overview of SMOS performance in terms of global soil moisture monitoring after six years in operation. *Remote Sensing of Environment* **180**, 40–63, <https://doi.org/10.1016/j.rse.2016.02.042> (2016).
- Kerr, Y. H. *et al.* The SMOS Mission: New Tool for Monitoring Key Elements of the Global Water Cycle. *Proceedings of the IEEE* **98**(5), 666–687 (2010).
- Entekhabi, D. *et al.* The Soil Moisture Active Passive (SMAP) Mission. *Proceedings of the IEEE* **98**(5), 704–716, <https://doi.org/10.1109/JPROC.2010.2043918> (2010).
- Yang, J. *et al.* Improvements on global meteorological observations from the current Fengyun 3 satellites and beyond. *International Journal of Digital Earth* **5**(3), 251–265, <https://doi.org/10.1080/17538947.2012.658666> (2012).
- Wang, G. *et al.* Towards Consistent Soil Moisture Records from China's FengYun-3 Microwave Observations. *Remote Sensing* **14**(5), 1225 (2022).
- Kim, S., Sharma, A., Liu, Y. Y. & Young, S. I. Rethinking Satellite Data Merging: From Averaging to SNR Optimization. *IEEE Transactions on Geoscience and Remote Sensing* **60**, 1–15, <https://doi.org/10.1109/TGRS.2021.3107028> (2022).
- van der Schalie, R. *et al.* L-Band Soil Moisture Retrievals Using Microwave Based Temperature and Filtering. Towards Model-Independent Climate Data Records. *Remote Sensing* **13**(13), 2480 (2021).
- Parinussa, R. M., Yilmaz, M. T., Anderson, M. C., Hain, C. R. & de Jeu, R. A. M. An intercomparison of remotely sensed soil moisture products at various spatial scales over the Iberian Peninsula. *Hydrological Processes* **28**(18), 4865–4876, <https://doi.org/10.1002/hyp.9975> (2014).
- Wang, G., Garcia, D., Liu, Y., de Jeu, R. & Johannes Dolman, A. A three-dimensional gap filling method for large geophysical datasets: Application to global satellite soil moisture observations. *Environmental Modelling & Software* **30**, 139–142, <https://doi.org/10.1016/j.envsoft.2011.10.015> (2012).
- Meng, X. *et al.* A fine-resolution soil moisture dataset for China in 2002–2018. *Earth Syst. Sci. Data* **13**(7), 3239–3261, <https://doi.org/10.5194/essd-13-3239-2021> (2021).
- Rao, P. *et al.* Daily soil moisture mapping at 1 km resolution based on SMAP data for desertification areas in northern China. *Earth Syst. Sci. Data* **14**(7), 3053–3073, <https://doi.org/10.5194/essd-14-3053-2022> (2022).
- Barth, A., Alvera-Azcárate, A., Licer, M. & Beckers, J. M. DINCAE 1.0: a convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations. *Geosci. Model Dev.* **13**(3), 1609–1622, <https://doi.org/10.5194/gmd-13-1609-2020> (2020).

18. van der Schalie, R. *et al.* Global SMOS Soil Moisture Retrievals from The Land Parameter Retrieval Model. *International Journal of Applied Earth Observation and Geoinformation* **45**, 125–134, <https://doi.org/10.1016/j.jag.2015.08.005> (2016).
19. Beck, H. E. *et al.* Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrol. Earth Syst. Sci.* **23**(1), 207–224, <https://doi.org/10.5194/hess-23-207-2019> (2019).
20. Albergel, C. *et al.* ERA-5 and ERA-Interim driven ISBA land surface model simulations: which one performs better? *Hydrol. Earth Syst. Sci.* **22**(6), 3515–3532, <https://doi.org/10.5194/hess-22-3515-2018> (2018).
21. Beck, H. E. *et al.* Evaluation of 18 satellite- and model-based soil moisture products using in situ measurements from 826 sensors. *Hydrol. Earth Syst. Sci. Discuss.* **2020**, 1–35, <https://doi.org/10.5194/hess-2020-184> (2020).
22. Hagan, D. F. T., and Parinussa, R. M., Wang, G., and Draper, Clara. An Evaluation of Soil Moisture Anomalies from Global Model-Based Datasets over the People's Republic of China, *Water* **12** (1), <https://doi.org/10.3390/w12010117> (2020).
23. Li, Mingxing, Wu, Peili, Ma, Zhuguang, A comprehensive evaluation of soil moisture and soil temperature from third-generation atmospheric and land reanalysis data sets, *International Journal of Climatology* **n/a**(n/a), <https://doi.org/10.1002/joc.6549> (2020).
24. Dorigo, W. A. *et al.* Global Automated Quality Control of In Situ Soil Moisture Data from the International Soil Moisture Network. *Vadose Zone Journal* **12**(3), vzj2012.0097, <https://doi.org/10.2136/vzj2012.0097> (2013).
25. Parinussa, R. *et al.* The Evaluation of Single-Sensor Surface Soil Moisture Anomalies over the Mainland of the People's Republic of China. *Remote Sensing* **9**, 149 (2017).
26. Beck, H. E. *et al.* Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data* **5**(1), 180214, <https://doi.org/10.1038/sdata.2018.214> (2018).
27. Funk, Chris *et al.* The climate hazards infrared precipitation with stations-a new environmental record for monitoring extremes. *Scientific Data* **2**(1), 150066, <https://doi.org/10.1038/sdata.2015.66> (2015).
28. Ullah, W. *et al.* Comparing Multiple Precipitation Products against In-Situ Observations over Different Climate Regions of Pakistan. *Remote Sensing* **11**(6), 628 (2019).
29. Owe, M., de Jeu, R., Holmes, T. Multisensor historical climatology of satellite-derived global land surface moisture, *Journal of Geophysical Research: Earth Surface* **113**(F1), <https://doi.org/10.1029/2007jf000769> (2008).
30. Holmes, T. R. H., De Jeu, R. A. M., Owe, M., Dolman, A. J., Land surface temperature from Ka band (37 GHz) passive microwave observations, *Journal of Geophysical Research: Atmospheres* **114**(D4), <https://doi.org/10.1029/2008jd010257> (2009).
31. Song, P. *et al.* An Improved Soil Moisture Retrieval Algorithm Based on the Land Parameter Retrieval Model for Water-Land Mixed Pixels Using AMSR-E Data. *IEEE Transactions on Geoscience and Remote Sensing* **57**(10), 7643–7657 (2019).
32. de Jeu, R. A. M. *et al.* Global Soil Moisture Patterns Observed by Space Borne Microwave Radiometers and Scatterometers. *Surveys in Geophysics* **29**(4), 399–420, <https://doi.org/10.1007/s10712-008-9044-0> (2008).
33. Owe, M., Jeu, R. de & Walker, J. A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. *IEEE Transactions on Geoscience and Remote Sensing* **39**(8), 1643–1654, <https://doi.org/10.1109/36.942542> (2001).
34. Parinussa, R. M. *et al.* Improved surface soil moisture anomalies from Fengyun-3B over the Jiangxi province of the People's Republic of China. *International Journal of Remote Sensing* **39**(23), 8950–8962, <https://doi.org/10.1080/01431161.2018.1500729> (2018).
35. Parinussa, R. M. *et al.* Global surface soil moisture from the Microwave Radiation Imager onboard the Fengyun-3B satellite. *International Journal of Remote Sensing* **35**(19), 7007–7029, <https://doi.org/10.1080/01431161.2014.960622> (2014).
36. van der Schalie, R. *et al.* The merging of radiative transfer based surface soil moisture data from SMOS and AMSR-E. *Remote Sensing of Environment* **189**, 180–193, <https://doi.org/10.1016/j.rse.2016.11.026> (2017).
37. Gruber, A., Dorigo, W. A., Crow, W. & Wagner, W. Triple Collocation-Based Merging of Satellite Soil Moisture Retrievals. *IEEE Transactions on Geoscience and Remote Sensing* **55**(12), 6780–6792, <https://doi.org/10.1109/TGRS.2017.2734070> (2017).
38. Kim, S., Parinussa, R., Liu, Y., Johnson, F. & Sharma, A. Merging Alternate Remotely-Sensed Soil Moisture Retrievals Using a Non-Static Model Combination Approach. *Remote Sensing* **8**, 518, <https://doi.org/10.3390/rs8060518> (2016).
39. Kim, S., Parinussa, R. M., Liu, Yi. Y., Johnson, F. M. & Sharma, Ashish A framework for combining multiple soil moisture retrievals based on maximizing temporal correlation. *Geophysical Research Letters* **42**(16), 6662–6670, <https://doi.org/10.1002/2015gl064981> (2015).
40. Zeng, Y. *et al.* Blending Satellite Observed, Model Simulated, and in Situ Measured Soil Moisture over Tibetan Plateau. *Remote Sensing* **8**(3), 268 (2016).
41. Kim, S., Liu, Y. Y., Johnson, F. M., Parinussa, R. M. & Sharma, Ashish A global comparison of alternate AMSR2 soil moisture products: Why do they differ? *Remote Sensing of Environment* **161**, 43–62, <https://doi.org/10.1016/j.rse.2015.02.002> (2015).
42. Martin, Z. K., Barnes, E. A. & Maloney, E. Using Simple, Explainable Neural Networks to Predict the Madden-Julian Oscillation. *Journal of Advances in Modeling Earth Systems* **14**(5), e2021MS002774, <https://doi.org/10.1029/2021MS002774> (2022).
43. Reichstein, M. *et al.* Deep learning and process understanding for data-driven Earth system science. *Nature* **566**(7743), 195–204, <https://doi.org/10.1038/s41586-019-0912-1> (2019).
44. Roundy, J. K. & Wood, E. F. The Attribution of Land-Atmosphere Interactions on the Seasonal Predictability of Drought. *Journal of Hydrometeorology* **16**(2), 793–810, <https://doi.org/10.1175/jhm-d-14-0121.1> (2015).
45. Hu, Y. *et al.* Reconstructing long-term global satellite-based soil moisture data using deep learning method, *Frontiers in Earth Science* **11**, <https://doi.org/10.3389/feart.2023.1130853> (2023).
46. Crow, W. T. & Zhan, X. Continental-Scale Evaluation of Remotely Sensed Soil Moisture Products. *IEEE Geoscience and Remote Sensing Letters* **4**(3), 451–455, <https://doi.org/10.1109/LGRS.2007.896533> (2007).
47. Crow, W. T., Miralles, D. G. & Cosh, M. H. A Quasi-Global Evaluation System for Satellite-Based Surface Soil Moisture Retrievals. *IEEE Transactions on Geoscience and Remote Sensing* **48**(6), 2516–2527, <https://doi.org/10.1109/TGRS.2010.2040481> (2010).
48. Parinussa, R. M. *et al.* A Quasi-Global Approach to Improve Day-Time Satellite Surface Soil Moisture Anomalies through the Land Surface Temperature Input. *Climate* **4**(4), 50 (2016).
49. Van der Schalie, R. *et al.* The Effect of Three Different Data Fusion Approaches on the Quality of Soil Moisture Retrievals from Multiple Passive Microwave Sensors. *Remote Sensing* **10**(1), 107 (2018).
50. Parinussa, R. M., Holmes, T. R. H., Yilmaz, M. T. & Crow, W. T. The impact of land surface temperature on soil moisture anomaly detection from passive microwave observations. *Hydrol. Earth Syst. Sci.* **15**(10), 3135–3151, <https://doi.org/10.5194/hess-15-3135-2011> (2011).
51. Ruscica, R. C. *et al.* Spatio-Temporal Soil Drying in Southeastern South America: The Importance of Effective Sampling Frequency and Observational Errors on Drydown Time Scale Estimates. *International Journal of Remote Sensing* **41**(20), 7958–92, <https://doi.org/10.1080/01431161.2020.1767825> (2020).
52. McColl, K. A. *et al.* Global characterization of surface soil moisture drydowns. *Geophysical Research Letters* **44**, 3682–3690, <https://doi.org/10.1002/2017GL072819> (2017).
53. Taylor, K. E. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres* **106**(D7), 7183–7192, <https://doi.org/10.1029/2000JD900719> (2001).
54. Hu, Z., Chen, X., Zhou, Q., Chen, D. & Li, J. DISO: A rethink of Taylor diagram. *International Journal of Climatology* **39**(5), 2825–2832, <https://doi.org/10.1002/joc.5972> (2019).
55. Zhou, Q., Chen, D., Hu, Z. & Chen, X. Decompositions of Taylor diagram and DISO performance criteria. *International Journal of Climatology* **41**(12), 5726–5732, <https://doi.org/10.1002/joc.7149> (2021).

56. Hagan, D. F. T. *et al.* Seamless High-resolution Soil Moisture from the Synergistic Merging of the Fengyun-3 Satellite Observations Series. *Zenodo*, <https://doi.org/10.5281/zenodo.11500736> (2024).
57. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization, *CoRR* abs/1412.6980 (2014).
58. Parinussa, R. M. *et al.* Error Estimates for Near-Real-Time Satellite Soil Moisture as Derived From the Land Parameter Retrieval Model. *IEEE Geoscience and Remote Sensing Letters* 8(4), 779–783, <https://doi.org/10.1109/LGRS.2011.2114872> (2011).

Acknowledgements

This study is supported by the National Natural Science Foundation of China (U23A2006, 42275028), the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX21_0935) and the European Research Council Consolidator Grant, HEAT(101088405). The authors also appreciate Dr. Robert Parinussa and Dr. Robin van der Schalie for all their work and contribution to the development of this dataset.

Author contributions

D.F.T.H. and G.W. conceived the idea. D.F.T.H., S.K. and X.M. conducted the research, completed the original draft, and revised it. Y.L. and A.B. gave constructive suggestions for the research design, data processing and manuscript revision. Y.H., H.L. and W.U. contributed to data processing and outputting manuscript figures. G.W., the corresponding author, contributed to designing, reviewing of the manuscript, funding acquisition, and project administration. All coauthors reviewed the manuscript and contributed to the writing process.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to G.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025