scientific **data**

DATA DESCRIPTOR

Check for updates

Global soil moisture data fusion by Triple Collocation Analysis from 2011 to 2018

Qiuxia Xie [1,2], Li Jia [2 ✉], Massimo Menenti[2,3] & Guangcheng Hu[2]

Surface Soil Moisture (SSM) information is needed for agricultural water resource management, hydrology and climate analysis applications. Temporal and spatial sampling by the space-borne instruments designed to retrieve SSM is, however, limited by the orbit and sensors of the satellites. We produced a Global Daily-scale Soil Moisture Fusion Dataset (GDSMFD) with 25 km spatial resolution (2011~2018) by applying the Triple Collocation Analysis (TCA) and Linear Weight Fusion (LWF) methods. Using five metrics, the GDSMFD was evaluated against *in-situ* soil moisture measurements from ten ground observation networks and compared with the prefusion SSM products. Results indicated that the GDSMFD was consistent with *in-situ* soil moisture measurements, the minimum of root mean square error values of GDSMFD was only 0.036 cm³/cm³. Moreover, the GDSMFD had a good global coverage with mean Global Coverage Fraction (GCF) of 0.672 and the maximum GCF of 0.837. GDSMFD performed well in accuracy and global coverage fraction, making it valuable in applications to the global climate change monitoring, drought monitoring and hydrological monitoring.

Background & Summary

Information on Surface Soil Moisture (SSM) plays a key role for many practical applications, such as agricultural water management, global weather forecasts, hydrology and natural disasters monitoring. SSM is also an important variable in water and energy exchanges at the atmosphere/land-surface interface[1,2]. Since 1970s active and passive microwave remote sensing instruments and algorithms to retrieve near-surface soil moisture (top 5 cm depth), have improved to make it possible to monitor global SSM with sufficient accuracy and spatio-temporal resolution[3–5]. Active sensors mainly include the Advanced Scatterometer (ASCAT) onboard the Meteorological Operational Satellite-A/B (MetOp-A/B) (2007-present) and the radar onboard Soil Moisture Active and Passive mission (SMAP) (2015-present). Passive sensors mainly include the Tropical Rainfall Measuring Mission's (TRMM) Microwave Imager (TMI) (1997–2015), the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) onboard Aqua satellite (2002~2011), the Coriolis satellite WindSat (2003–2012), the Soil Moisture and Ocean Salinity (SMOS) (2010-present)[6], the MicroWave Radiation Imager (MWRI) of FenYun-3B (FY3-B) (2011–2019) and FenYun-3C (FY3-C) (2014-present) satellites, the Advanced Microwave Scanning Radiometer-2 (AMSR-2) of Global Change Observation Mission 1st - Water "SHIZUKU" (GCOM-W1) (2012-present) and the radiometer onboard SMAP (2015-present).

Based on observations of active/passive microwave remote sensing instruments, many global SSM products such as AMSR-E, AMSR-2, ASCAT, SMOS, SMAP and the European Space Agency-Climate Change Initiative (ESA-CCI) were produced and evaluated using *in-situ* soil moisture measurements and airborne observations[3,7–9]. Particularly, the *in-situ* soil moisture measurements from the International Soil Moisture Network (ISMN) and simulated SSM data by land surface models were used for this purpose. However, previous studies found that no global active/passive SSM products performed superior to other products everywhere and performance seemed to depend on land cover[10]. In addition, there are gaps in the global SSM products due to incomplete coverage of satellite orbits and retrieval errors. These shortcomings limited these soil moisture data for long-time applications in, e.g., agricultural drought monitoring and water resources management.

Data fusion to merge different global SSM data product seems the most effective solution to overcome the shortcomings identified above[11]. Data fusion can integrate multiple remote sensing data sets and knowledge into

[1]School of Surveying and Geo-Informatics, Shandong Jianzhu University, Jinan, 250101, China. [2]State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, 100101, China. [3]Delft University of Technology, Delft, The Netherlands. ✉e-mail: jiali@aircas.ac.cn
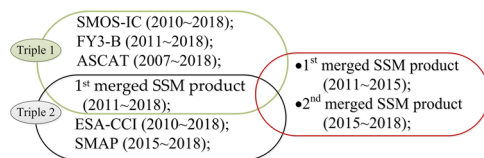
Fig. 1 Overview of the two-triplet merging approach from global-scale original SSM products (SMOS, FY3-B, ASCAT, ESA-CCI and SMAP) to final merged SSM products (1st and 2nd merged SSM products).

Products	Versions	Period	Sensor types	Used Band	SGR	Unit	Algorithms	Main reference
SMOS-IC	L3	2010~2018	Passive	1.41 GHz	25 km	cm³/cm³	L-MEB	Wigneron et al.[23]
FY3-B	L2	2011~2018	Passive	10.7 GHz	25 km	cm³/cm³	Qp	Shi et al.[27]
ASCAT	V3.0	2007~2018	Active	5.3 GHz	0.1°	%	TU-Wien	Wagner et al.[29]
SMAP	L3	2015~2018	Passive	1.41 GHz	9 km	cm³/cm³	SCA-V	Jackson[37]
ESA-CCI	V4.5	1978~2018	/	/	0.25°	cm³/cm³	Merged	Gruber et al.[20]

Table 1. Information on five active and passive microwave SSM products used in this study.

a better data product[12]. Examples of such algorithms are the Linear Weight Fusion (LWF), non-Linear WF, multiple linear regression, artificial neural network, Copula-based data fusion and least square method-based image fusion[13–16]. The LWF algorithm, combining data sets linearly based on the weight values of components to be fused, is the most practical and well understood fusion algorithm[17]. The key to the LWF algorithm is to calculate the weight values of the different components on the basis of the correlations or errors of different components. At present, Triple Collocation Analysis (TCA) method is one of the most popular methods for estimating random errors in soil moisture retrievals without reliable ground observations[18]. TCA was first proposed to estimate ocean wind speed errors and to correct ocean wind speed[19], and subsequently used to estimate uncertainties in remote sensing retrievals of, e.g., soil moisture and leaf area index[20]. Therefore, it will be potential by merging of existential soil moisture based on the combing algorithm of TCA and LWF methods for improvement of the global SSM product with 25 km SGR.

Here, we present a novel global daily-scale SSM product (i.e., GDSMFD) with 25 km SGR in good accuracy and global coverage using the easiest and well understood LWF method to merging SMOS, FY3-B, ASCAT, ESA-CCI and SMAP SSM products (i.e. SMOS SSM retrieved using L-MEB algorithm, FY3-B SSM retrieved using the soil moisture retrieval algorithm based on Qp roughness model, ASCAT SSM retrieved using the SWI algorithm based on change monitoring method, ESA-CCI SSM retrieved using the data fusion procedure and SMAP SSM retrieved using SCA-V algorithm). We calculated the pixel-wised weight values of the SM values from SMOS, FY3-B, ASCAT, SMAP SSM and ESA-CCI SSM products based on TCA. We applied a strategy that first SMOS, FY3-B and ASCAT SSM products from 2011 to 2018 were merged using the TCA LWF method to produce the 1st merged SSM product which is further merged with SMAP-SSM and ESA-CCI-SM products to obtain the 2nd merged SSM product for 2015~2018 (SMAP-SSM available from 2015) by using the same TCA LWF method (Fig. 1). The above two merged SM datasets composed our time series of fusion SM data between 2011–2018.

Methods

Global soil moisture fusion data preparation. Five satellite surface soil moisture products including three passive microwave soil moisture products (i.e., SMOS-IC, FY3-B and SMAP), one active microwave soil moisture products (i.e., ASCAT) and one merged soil moisture product (i.e., ESA-CCI) were used in this study to generate the merged dataset (Table 1), the details are described below.

SMOS satellite launched in November 2009 by ESA carries the Microwave Interferometric Radiometer with Aperture Synthesis (MIRAS), a synthetic aperture microwave radiometer to observe the dual-polarized, multi-angular brightness temperature in L band (1.4 GHz) and designed to retrieve global SSM at 3~5 cm soil depth with high accuracy[21]. In this study, we used the new global SMOS-IC L3 SSM product with 25 km EASE 2.0 SGR generated by the Institut National de la Recherche Agronomique-Centre d'Etudes Spatiales de la Biosphère (INRA-CESBIO) (http://bec.icm.csic.es/data/data-access/)[22,23]. The SMOS-IC L3 SSM retrieval algorithm was based on the L-MEB to make soil moisture retrieval as independent as possible from auxiliary data such as the Leaf Area Index (LAI) and Normalized Difference Vegetation Index (NDVI)[23,24]. The SMOS-IC L3 SSM algorithm considered pixels as homogeneous avoiding uncertainties related to auxiliary data sets that were used to characterize the pixel heterogeneity in e.g. forest areas, which different from SMOS L2 SSM[25]. Compared to SMOS L2 SSM products, the SMOS-IC L3 SSM product is more robust in reducing the effect of vegetation and surface roughness[23].

FY3-B launched on November 5, 2010, was the second polar-orbiting satellite of the FY3 meteorological satellite series and carried the MWRI operating at 5 frequencies, i.e., 10.65, 18.7, 23.8, 36.5 and 89 GHz, and Horizontal (H) and Vertical (V) polarization. MWRI is one of the eleven instruments onboard the FY3-B satellite and measures the exitance of the earth-atmosphere system[26]. In this study, we used the daily FY3-B SSM product with 25 km SGR and a nominal sensing soil depth of 2 cm and applying a retrieval algorithm including

an improved Qp surface roughness model developed by Shi *et al.*[27]. This product was provided by the National Satellite Meteorological Center (NSMC) of China (http://satellite.nsmc.org.cn/PortalSite/Default.aspx). The improved Qp model reduces the effect of surface roughness on the soil moisture retrieval. An empirical vegetation opacity model using Vegetation Water Content (VWC), assumed to depend only on LAI was applied to correct the effect of vegetation on FY3-B/MWRI brightness temperature observations[27].

ASCAT sensors onboard the MetOp-A and MetOp-B launched by ESA since 2006 are real aperture radar sensors operating in C band (5.255 GHz) using vertical polarization antennas to measure global radar backscatter at 25 km SGR with 1~3 days revisit time. In this study, we used the daily ASCAT SSM product with 10 km SGR released by the Copernicus Global Land Service (CGLS) based on the SWI algorithm developed by the Vienna University of Technology (https://land.copernicus.eu/global/products/swi)[28]. The SWI algorithm is based on a simple soil moisture infiltration model proposed by Wagner *et al.*[29] to estimate profile soil moisture from ASCAT backscatter observations[29]. ASCAT product has eight SWI layers according to different T values (i.e., 1, 5, 10, 15, 20, 40, 60 and 100). In the SWI algorithm, the T value is a function ($T = L/C$) of L (the depth of the soil layer) and C (the area-representative pseudo diffusivity constant) parameters, which means a high T describes a deeper soil layer if the soil water diffusivity is constant. Therefore, this study used the SWI data of the first layer ($T = 1$) so that each pixel value represents the degree of saturation (%) with a nominal sensing depth of 1~5 cm.

ESA-CCI V4.5 SSM product was developed by ESA in 2010 using the TCA-LWF algorithm applied to either backscatter coefficient or brightness temperature retrieved with three active sensors, i.e., European Remote Sensing-1/2 (ERS-1/2), MetOp-A/ASCAT, and MetOp-B/ASCAT, and seven passive sensors, i.e. Nimbus-7's Scanning Multichannel Microwave Radiometer (SMMR), Special Sensor Microwave-Imager (SSM/I), TRMM, AMSR-E, AMSR-2, WindSat, and SMOS[15,20,30–33]. Currently, it is one of the most widely used SSM data products in the world (https://esa-soilmoisture-cci.org/). The three soil moisture data products based on the measurements by active instruments were retrieved using the same soil moisture retrieval algorithm, i.e., the change detection method developed by the Vienna University of Technology (TU-Wien) WARP v5.5, from backscatter coefficient observed by ERS-1/2, MetOp-A/ASCAT, and MetOp-B/ASCAT, then were merged using the TCA-LWF algorithm. The seven SSM data products based on the measurements by microwave radiometers were retrieved using the same soil moisture retrieval algorithm, i.e., the LPRM developed by National Aeronautics and Space Administration (NASA), from brightness temperature observed by SMMR, SSM/I, TRMM, AMSR-E, AMSR2, WindSat and SMOS, then were merged using the TCA-based LWF algorithm. Finally, the active and passive merged data products were blended into one final dataset. In this study, we merged the final combined dataset with SMOS, ASCAT, FY3-B and SMAP SSM products.

SMAP was launched by NASA in January 2015 and includes a L-band (1.26 GHz) radar and a L-band (1.41 GHz) radiometer. This unique near-polar sun-synchronous satellite monitors global SSM in the top 5 cm soil and freeze/thaw state with higher accuracy, coverage and resolution than previous microwave systems[34]. Currently, there are three SMAP SSM products with 3 km, 9 km, and 36 km SGR available. The L-band radar with higher SGR (1~3 km) stopped working after 11 weeks of operation, however. Therefore, only SMAP SSM products with 9 km and 36 km SGR are available from 2015 to the present. In this study we used the enhanced L3 SMAP SSM product assuming a 5 cm soil sensing depth with 9 km EASE 2.0 SGR released by the Nation Snow & Ice Data Center (NSIDC) (http://nsidc.org/data/)[35]. This product is a daily composite of the enhanced SMAP L2 half-orbit SSM data with 9 km SGR, retrieved from SMAP L1 interpolated brightness temperature data with 36 km SGR using the SCA-V algorithm[36–38].

Soil moisture validation data preparation. *In-situ* soil moisture measurements collected at a total 311 sites of the International Soil Moisture Network (ISMN) (https://ismn.geo.tuwien.ac.at/en/), specifically 10 contributing networks, i.e. CTP_SMTMN, RSMN, AMMACATCH, DAHRA, BIEBRZA_S-1, MySMNet, REMEDHUS, HOBE, USCRN and OZNET, with dense sites in 9 countries, i.e. Poland, Spain, China, Malaysia, Romania, Benin, Niger, Mali, Senegal, USA, Australia and Denmark were used to verify the TCA assumptions and evaluate SMOS, FY3-B, ASCAT, ESA-CCI, SMAP satellite SSM products and the merged SSM product developed by this study using the TCA-LWF algorithm (Table 2 and Fig. 2)[39–49]. These 311 sites are part of 10 networks distributed in five continents: Asia, Europe, Africa, North America and Australia. Particularly, in the CTP_SMTMN in China, there are about 20 sites in a pixel of the SSM product with 25 km SGR. The mean value of the *in-situ* soil moisture data in a pixel of the SSM products was considered as the true value of soil moisture. Also, we used the ESA-CCI L4 global 2015 Land Cover (LC) map with 300 m SGR to identify the land cover type of each *in-situ* soil moisture measurement site (http://www.esa-landcover-cci.org/). The soil moisture measurements are obtained at different soil depths, i.e., 5 cm, 10 cm, 20 cm and 40 cm. In this study, we used the *in-situ* soil moisture measurements at 5 cm soil depth, approximately consistent with the nominal sensing depth of microwave SSM L-band instruments.

We identified the land cover type at the sites described above by applying the ESA-CCI L4 2015 global LC map with 300 m SGR (Fig. 2). The land cover types of ESA-CCI LC data were defined using the United Nations-Land Cover Classification System (UN-LCCS) which considers 40 classes. The 311 ISMN sites were distributed mainly in 8 classes (Fig. 2), with most sites being located in grassland and cropland areas, followed by shrub, tree and mosaic cropland/natural vegetation (tree, shrub and grass) (37). In the REMEDHUS network of Spain, there were 24 sites, among which 16 sites were located in herbaceous cover area. In the RSMN network of Romania, 20 sites were distributed mainly in the rainfed cropland and herbaceous cover area. In the BIEBRZA_S-1 network of Poland, 28 sites were mainly in rainfed cropland and grassland, others in shrubland or herbaceous cover, flooding, fresh/saline/brackish water area. In the CTP_SMTMN network, the land cover of 56 out of 57 sites was grassland. In the MySMNet network of Malaysia, 4 out of 7 sites were in the mosaic nature vegetation (tree, shrub, herbaceous >50%)/ cropland (<50%). In DAHRA and AMMA-CATCH networks of

Networks	Countries	N	Instruments	Obs. Depth	Period (dd/mm/yy)	References
CTP_SMTMN	China	57	5TM/EC-TM	5 cm	01/08/2010~19/09/2016	Yang et al.[40]
RSMN	Romania	20	5TM (0~1)	5 cm	09/04/2014~15/05/2020	Sandric et al.[41]
AMMACATCH	Benin, Niger, Mali	7	CS616	5 cm	01/01/2006~31/12/2018	Pellarin et al.[42]
DAHRA	Senegal	1	Theta Probe ML2X	5 cm	04/07/2002~01/01/2016	Tagesson et al.[43]
BIEBRZA_S-1	Poland	28	GS-3(0~1)	5 cm	23/04/2015~01/12/2018	Musial et al.[44]
MySMNet	Malaysia	7	Water Scout SM100	5 cm	31/05/2014~31/12/2015	Kang et al.[45]
REMEDHUS	Spain	24	Stevens Hydra Probe	5 cm	15/03/2005~01/01/2020	Martínez-Fernández et al.[46]
HOBE	Denmark	32	Decagon 5TE	5 cm	08/09/2009~13/03/2019	Bircher et al.[47]
USCRN	USA	115	Stevens Hydra probe II Sdi-12	5 cm	15/11/2000~26/10/2020	Bell et al.[48]
OZNET	Australia	20	Stevens Hydra Probe	5 cm	12/09/2001~27/08/2018	Young et al.[49]

Table 2. Information on *in-situ* soil moisture measurements used in this study (*N*: the number of sites).

Africa, out of 8 sites, 2 sites were in the broadleaf deciduous tree (15~40%), 2 sites in shrubland and 2 sites in grassland. In HOBE 11 out of 32 sites are located in raining cropland, while the remaining ones were concentrated in tree and mosaic cropland/natural vegetation in the USCRN, 45 out of 115 were in grassland and 24 sites were in shrub land. In OZNET, 8 out of 20 were in rained cropland, 4 in shrub land and 5 in mixed vegetation. Overall, 116 of the 311 ISMN sites were in grassland, 93 in cropland, 30 in shrub land, 21 in tree and the remaining ones were in mosaic cropland/natural vegetation (tree, shrub and grass).

Triple collocation analysis. At present, the TCA method is one of the most widely used evaluation methods for satellite SSM products in the absence of *in-situ* soil moisture measurement data[50]. The evaluation of errors in satellite SSM data products using TCA is based on three assumptions: a) soil moisture retrievals are linearly related to the true soil moisture value; b) the errors on each SSM retrieval are uncorrelated with the true soil moisture; and c) errors within each selected triplet of SSM retrievals are uncorrelated with each other[20].

The retrieved soil moisture (SM, cm^3/cm^3) in each pixel can then be expressed as a linear relationship between the true soil moisture value (t, cm^3/cm^3) and random error (ε) as follows[50]:

$$SM = \alpha + \beta * t + \varepsilon \tag{1}$$

Where α and β are the coefficients.

The covariance ($Cov(SM_i, SM_j)$) of two SSM products (SM_i and SM_j) is then:

$$Cov(SM_i, SM_j) = \beta_i\beta_j\sigma_t^2 + \beta_i Cov(t, \varepsilon_j) + \beta_j Cov(t, \varepsilon_i) + Cov(\varepsilon_i, \varepsilon_j) \tag{2}$$

$$Q_{ij} = \begin{bmatrix} Q_{11} & Q_{12} & Q_{13} \\ Q_{21} & Q_{22} & Q_{23} \\ Q_{31} & Q_{32} & Q_{33} \end{bmatrix}, \begin{cases} i = 1, 2, 3 \\ j = 1, 2, 3 \end{cases} \tag{3}$$

Where σ_t^2 is the variance of true soil moisture value (t, cm^3/cm^3); $Cov(SM_i, SM_j)$ is the covariance matrix of paired data products, with three pairs in each triplet when applying TCA (Eq. 3) (Table 3).

In each triplet, three soil moisture datasets were retrieved using different algorithms and from different data sources. Accordingly, it was assumed that zero error cross-correlation between each pair of soil moisture datasets was nihil, i.e., $Cov(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$, and that the error - on "true" soil moisture and the error on each soil moisture dataset i.e., $Cov(t, \varepsilon_i) = 0$ were orthogonal. Based on these assumptions, Eq. 2 can be rewritten as:

$$Q_{ij} = Cov(SM_i, SM_j) = \begin{cases} \beta_i\beta_j\sigma_t^2, & i \neq j \\ \beta_i\beta_j\sigma_t^2 + \sigma_{\varepsilon i}^2, & i = j \end{cases}, \begin{cases} i = 1, 2, 3 \\ j = 1, 2, 3 \end{cases} \tag{4}$$

Since there are six unique terms (Q_{11}, $Q_{12} = Q_{21}$, $Q_{13} = Q_{31}$, Q_{22}, $Q_{23} = Q_{32}$, Q_{33}) in the 3×3 covariance matrix (Eq. 3), we can obtain six equations but seven unknowns (β_1, β_2, β_3, $\sigma_{\varepsilon 1}^2$, $\sigma_{\varepsilon 2}^2$, $\sigma_{\varepsilon 3}^2$, σ_t^2). Therefore, there is no unique solution. A new variable $\theta_i = \beta_i\sigma_t$ is introduced in Eq. 4, which can be rewritten as:

$$Q_{ij} = \begin{cases} \theta_i\theta_j, & i \neq j \\ \theta_i^2 + \sigma_{\varepsilon i}^2, & i = j \end{cases}, \begin{cases} i = 1, 2, 3 \\ j = 1, 2, 3 \end{cases} \tag{5}$$

In Eq. 5, there are six equations and six unknowns. Therefore, a unique value for each unknown can be calculated. Finally, the error variance values on the three independent SSM products in each triplet (Table 3) ($\sigma_{\varepsilon 1}$, $\sigma_{\varepsilon 2}$, $\sigma_{\varepsilon 3}$) can be estimated as:
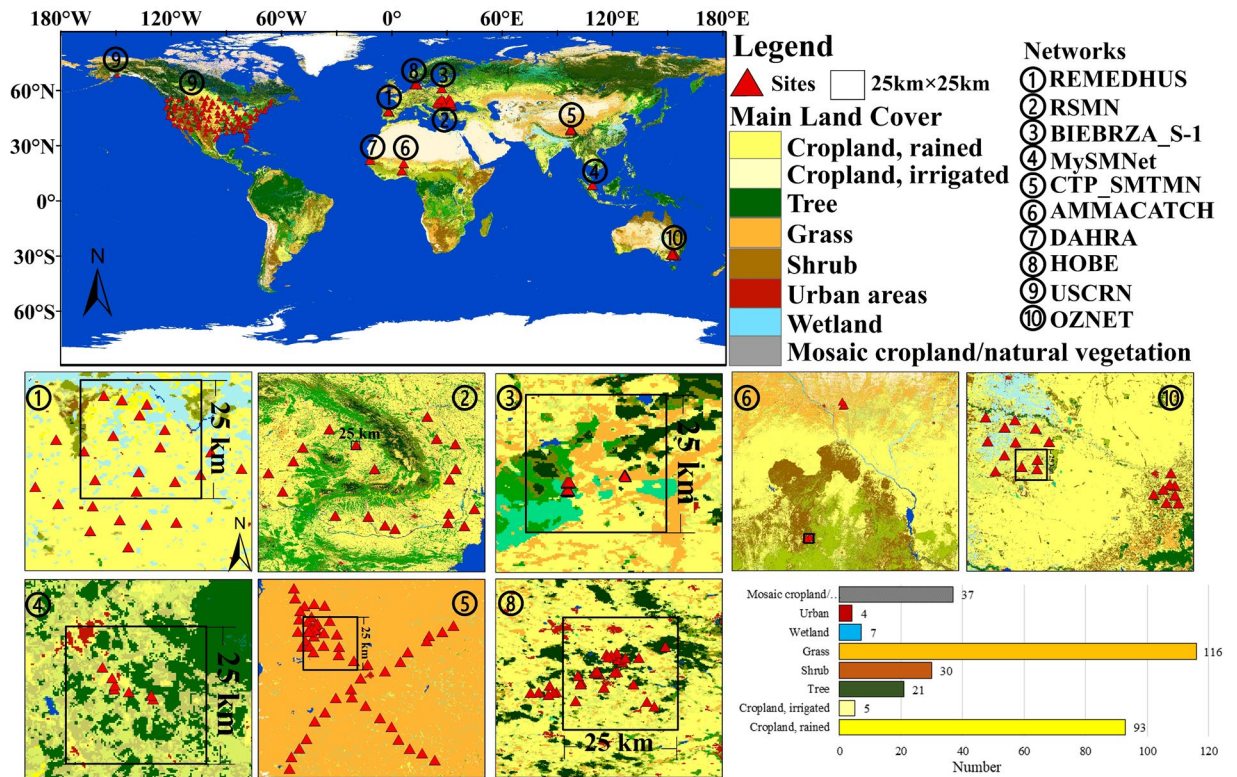
Fig. 2 Distribution of *in-situ* soil moisture measurement stations and LC characteristics of the networks listed in Table 2.

Steps	Pairs	Results
Triplet 1	SMOS, FY3-B	1st merged SSM product
	SMOS, ASCAT	
	FY3-B, ASCAT	
Triplet 2	1st merged SSM, ESA-CCI	2nd merged SSM product
	1st merged SSM, SMAP	
	ESA-CCI, SMAP	

Table 3. Pair strategy of SMOS, FY3-B, ASCAT, SMAP and ESA-CCI for covariance estimation in the 1st and 2nd merging procedures.

$$\begin{cases} \sigma_{\varepsilon 1}^2 \\ \sigma_{\varepsilon 2}^2 \\ \sigma_{\varepsilon 3}^2 \end{cases} = \begin{bmatrix} Q_{11} - \dfrac{Q_{12}Q_{13}}{Q_{23}} \\[2ex] Q_{22} - \dfrac{Q_{12}Q_{23}}{Q_{13}} \\[2ex] Q_{33} - \dfrac{Q_{23}Q_{13}}{Q_{12}} \end{bmatrix} \tag{6}$$

The linear weight fusion (LWF) method. The key in the LWF method is to estimate the weight value of each pixel of each SSM product. When the soil moisture values of three SSM products in 1st or 2nd step of the merging procedure are available, the weight values of three SSM products (w_1, w_2, w_3) can be calculated by using the estimated error variance values on three SSM products from Eq. 6 as:

$$\begin{cases} w_1 \\ w_2 \\ w_3 \end{cases} = \begin{cases} \dfrac{1/\sigma_{\varepsilon1}^2}{1/\sigma_{\varepsilon1}^2 + 1/\sigma_{\varepsilon2}^2 + 1/\sigma_{\varepsilon3}^2} \\[2mm] \dfrac{1/\sigma_{\varepsilon2}^2}{1/\sigma_{\varepsilon1}^2 + 1/\sigma_{\varepsilon2}^2 + 1/\sigma_{\varepsilon3}^2} , \ w_1 + w_2 + w_3 = 1 \\[2mm] \dfrac{1/\sigma_{\varepsilon3}^2}{1/\sigma_{\varepsilon1}^2 + 1/\sigma_{\varepsilon2}^2 + 1/\sigma_{\varepsilon3}^2} \end{cases} \tag{7}$$

When the soil moisture values of only two of three SSM products in 1st or 2nd step of the merging procedure are available, the weight values of two of three SSM products are re-estimated by using the estimated error variance values on two of the three SSM products from Eq. 6 as:

$$\begin{cases} w_1 \\ w_2 \end{cases} = \begin{cases} \dfrac{\frac{1}{\sigma_{\varepsilon1}^2}}{\frac{1}{\sigma_{\varepsilon1}^2} + \frac{1}{\sigma_{\varepsilon2}^2}} \\[2mm] \dfrac{\frac{1}{\sigma_{\varepsilon2}^2}}{\frac{1}{\sigma_{\varepsilon1}^2} + \frac{1}{\sigma_{\varepsilon2}^2}} \end{cases}, \ w_1 + w_2 = 1 \ \text{or} \ \begin{cases} w_1 \\ w_3 \end{cases} = \begin{cases} \dfrac{\frac{1}{\sigma_{\varepsilon1}^2}}{\frac{1}{\sigma_{\varepsilon1}^2} + \frac{1}{\sigma_{\varepsilon3}^2}} \\[2mm] \dfrac{\frac{1}{\sigma_{\varepsilon3}^2}}{\frac{1}{\sigma_{\varepsilon1}^2} + \frac{1}{\sigma_{\varepsilon3}^2}} \end{cases}, \ w_1 + w_3 = 1 \ \text{or} \ \begin{cases} w_2 \\ w_3 \end{cases} = \begin{cases} \dfrac{\frac{1}{\sigma_{\varepsilon2}^2}}{\frac{1}{\sigma_{\varepsilon2}^2} + \frac{1}{\sigma_{\varepsilon3}^2}} \\[2mm] \dfrac{\frac{1}{\sigma_{\varepsilon3}^2}}{\frac{1}{\sigma_{\varepsilon2}^2} + \frac{1}{\sigma_{\varepsilon3}^2}} \end{cases}, \ w_2 + w_3 = 1 \tag{8}$$

When only one soil moisture values of the three SSM data products in 1st step or 2nd step is available, the weight value of one of three SSM products is equal to 1 and weight = 0 for the two remaining SSM products. In this study, the scheme to estimate the weight values of three SSM products in 1st and 2nd steps is summarized in Table 4. When the temporal coverage of one or more data sets in the triplet is lower, TCA results are often considered unreliable. Therefore, the minimum number of valid samples in each pixel for each triplet in the time-series was set at 100[51].

The TCA–LWF method was used to merge the SSM data products in each triplet (Table 3). This method is simple and feasible, and the error attributes of every SSM product that are estimated using TCA algorithm are considered. By using the calculated weight values of three SSM products (w_1, w_2, w_3) (Table 4), three SSM products (SM_1, SM_2, SM_3) can be merged by calculating their weighted average as:

$$SM_f = w_1 * SM_1 + w_2 * SM_2 + w_3 * SM_3 \tag{9}$$

Based on preliminary assessment analysis of the satellite SSM products, the ESA-CCI SSM product has better performance in accuracy and effective coverage area than AMSR-2, AMSR-E, WindSat and TRMM/TMI SSM products[3,9,10,26,52]. Therefore, this was the reason that the ESA-CCI SSM product was selected and used in the 2nd triplet to obtain the 2nd merged (final) SSM. To limit the complexity of the merging procedure, i.e., the number of unknowns, we split it into two steps, each dealing with three SSM data products to get finally a single merged product from the SMOS, FY3-B, ASCAT, ESA-CCI, and SMAP SSM products. In the first step, the SMOS, FY3-B, and ASCAT SSM products are merged by using a LWF algorithm to obtain the 1st merged SSM product from 2011~2018. Second, the latter is merged with the ESA-CCI and SMAP SSM products by using the same algorithm i.e., the LWF algorithm from 2015~2018. The specific flow chart of SSM product fusion using the LWF algorithm is shown in Fig. 3.

Prior to the fusion of the SMOS, FY3-B, ASCAT, ESA-CCI and SMAP datasets using TCA-based LWF algorithm, unit conversion, abnormal pixel removal, masking and resampling were applied to these five SSM products. All SSM products, i.e., SMOS, FY3-B, ASCAT, ESA-CCI, and SMAP SSM, were resampled to a consistent SGR i.e., 25 km by interpolation of adjacent pixels. In addition, the soil moisture units of SMOS, FY3-B, ESA-CCI and SMAP are the same, i.e., volumetric soil water content (cm^3/cm^3), but ASCAT SSM is expressed as saturation (%). To keep consistent units. i.e. volumetric soil water content, cm^3/cm^3, we used the global porosity data with 1-degree SGR, released by the Goddard Earth Sciences Data and Information Services Center (GES DISC), were used to convert the ASCAT SWI product to volumetric soil water content (cm^3/cm^3)[53]. After resampling and unit conversion, the range of ASCAT SSM was 0–1.0 cm^3/cm^3, i.e., the same as SMAP. For SMOS product, the abnormal SMOS SSM values >1.0 or <0 cm^3/cm^3, caused by the L-MEB retrieval algorithm, were removed, so that the range of SMOS SSM was 0~1.0 cm^3/cm^3, i.e., the same as ASCAT and SMAP. The ranges of ESA-CCI and FY3-B SSM values were 0~0.6 cm^3/cm^3 and 0~0.5 cm^3/cm^3. Moreover, in order to keep soil moisture information from SMOS, FY3-B, ASCAT, ESA-CCI and SMAP SSM products as possible, the rescaling and masking processes were not implemented for these five original SSM products, but only for the final 1st merged and 2nd merged SSM products. The SSM values >0.6 cm^3/cm^3 in the 1st merged and 2nd merged SSM products were replaced using the saturated soil water content data released by the Land-Atmosphere Interaction Research Group at Sun Yat-sen University. The water bodies and permanent snow and ice according to the ESA-CCI L4 global LC map were masked in the final 1st merged and 2nd merged SSM products.

Metrics for evaluation. In this study, the Mean Absolute Error (MAE, cm^3/cm^3), Root Mean Square Error (RMSE, cm^3/cm^3), Relative Bias (RB), correlation coefficient (R) and Global Coverage Fraction (GCF) were used to compare and assess the 1st and 2nd merged SSM products, as well as the SMOS, FY3-B, ASCAT, ESA-CCI, SMAP:

Merge Steps	SMOS	FY3-B	ASCAT	Weight values	Result
Triplet 1	×	×	×	$w_1 = w_2 = w_3 = 0$	1st merged SSM product
	O	×	×	$w_1 = 1, w_2 = w_3 = 0$	
	×	O	×	$w_2 = 1, w_1 = w_3 = 0$	
	×	×	O	$w_3 = 1, w_1 = w_2 = 0$	
	O	O	×	w_1 and w_2 by Eq. 8, $w_3 = 0$	
	×	O	O	w_2 and w_3 by Eq. 8, $w_1 = 0$	
	O	×	O	w_1 and w_3 by Eq. 8, $w_2 = 0$	
	O	O	O	w_1, w_2 and w_3 by Eq. 7	
	1st merged SSM	**ESA-CCI**	**SMAP**	**Weight values**	**Result**
Triplet 2	×	×	×	$w_1 = w_2 = w_3 = 0$	2nd merged SSM product
	O	×	×	$w_1 = 1, w_2 = w_3 = 0$	
	×	O	×	$w_2 = 1, w_1 = w_3 = 0$	
	×	×	O	$w_3 = 1, w_1 = w_2 = 0$	
	O	O	×	w_1 and w_2 by Eq. 8, $w_3 = 0$	
	×	O	O	w_2 and w_3 by Eq. 8, $w_1 = 0$	
	O	×	O	w_1 and w_3 by Eq. 8, $w_2 = 0$	
	O	O	O	w_1, w_2 and w_3 by Eq. 7	

Table 4. Scheme to assign the weight values of three SSM products in 1st and 2nd steps (×: soil moisture value of SSM product is not available; O: soil moisture value of SSM product is available).
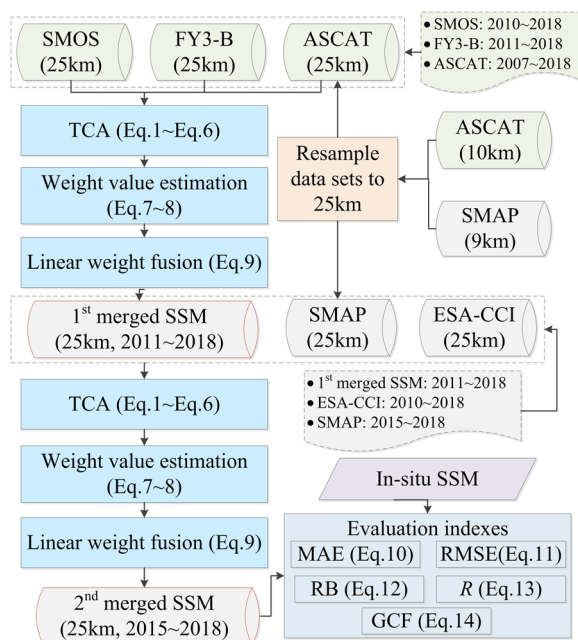
Fig. 3 Flowchart of work-flow to merge the SMOS, FY3-B, ASCAT, ESA-CCI, and SMAP SSM products using TCA method.

$$MAE = \frac{\sum_{D=1}^{N}(SM_D^E - SM_D^O)}{N} \tag{10}$$

$$RMSE = \sqrt{\frac{\sum_{D=1}^{N}(SM_D^E - SM_D^O)^2}{N}} \tag{11}$$

$$RB = \frac{\sum_{D=1}^{N}(SM_D^E - SM_D^O)}{\sum_{D=1}^{N}SM_D^O} \tag{12}$$

$$R = \frac{\text{cov}(\text{SM}_D^E, \text{SM}_D^O)}{\sigma_{\text{SM}_D^E}\sigma_{\text{SM}_D^O}}, \quad \begin{cases} \text{cov}: \text{covariance} \\ \sigma: \text{standard deviation} \end{cases} \tag{13}$$

$$\text{GCF} = \frac{\text{Area of successful SM}_t^E \text{ retrievals}}{\text{The total area of land surface}} \tag{14}$$

where SM_D^E (cm³/cm³) is the estimated soil moisture value on day D; SM_D^O (cm³/cm³) is the *in-situ* measured soil moisture on day D; N is the number of measurements.

Data Records

Global daily-scale soil moisture fusion dataset (GDSMFD) based on TCA-LWF (2011–2018) is publicly available at the National Tibetan Plateau Data Center (https://doi.org/10.11888/Terre.tpdc.271935)[54]. This dataset contains global daily-scale soil moisture fusion data with a spatial grid resolution of 25 km, in cm³/cm³, from January 1, 2011 to December 31, 2018. These data are stored in TIFF format with one file per day. Each file is named as "Fusion_SMOS_FY3B_ASCAT_ESACCI_SMAP_V1_yymmdd.tif", where "yy" represents the year, "mm" represents the month and "dd" represents the day.

Technical Validation

Verification of TCA assumptions.   The SMOS, FY3-B, and ASCAT were used to generate the 1st merged data product in the first step (Table 3), then combined the latter with SMAP and ESA-CCI in the second step (Table 3). TCA assumes independent errors of each pair of datasets. So, in each triplet, we selected soil moisture products generated with algorithms as different as possible, because algorithm similarity may lead to correlated errors on the SSM datasets (Table 3). In triplet 1, SMOS SSM product was retrieved using the L-MEB algorithm from the MIRAS passive microwave brightness temperature observations with L band (1.4 GHz) and multi-angles (0~55°), FY3-B SSM product was estimated using the Qp- algorithm from the MWRI passive microwave brightness temperature observations with X band (10.65 GHz) and single incidence angle (55°), while the ASCAT SSM product was retrieved using the SWI algorithm from active microwave backscattering observations in C band (5.255 GHz) and multi-angular observations (25~65°). In triplet 2, the 1st merged SSM product was estimated using TCA-based LWF algorithm from SMOS, FY3-B and ASCAT SSM products, while SMAP SSM product was retrieved using the SCA-V algorithm from L band (1.4GHz) and single incidence angle (40°) and the ESA-CCI SSM product was derived from seven passive microwave SSM products (SMMR, SSM/I, TRMM, AMSR-E, AMSR-2, and SMOS) using the LPRM algorithm, and three active microwave SSM products (ERS-1/2, MetOp-A/ASCAT, and MetOp-B/ASCAT) using the change detection algorithm.

In Eq. 2 to Eq. 3, $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ when $i \neq j$ (zero error cross-correlation) and $\text{Cov}(t, \varepsilon_i) = 0$ (error-orthogonality) were assumed. To verify quantitatively the assumptions of TCA when the pairing strategy of SMOS, FY3-B, ASCAT, SMAP and ESA-CCI (i.e., Table 3) was used to apply the TCA algorithm, the *in-situ* soil moisture measurement data of CTP_SMTMN, RSMN, DAHRA, BIEBRZA_S-1, MySMNet and REMEDHUS networks were considered the "true" SSM values, i.e., the t in Eqs. 1~3). To calculate the errors on the SSM products, i.e., the ε in Eq. 1, first, the linear function $SSM = \alpha + \beta*t$ between the *in-situ* soil moisture measurement and soil moisture product was fitted and the values of the α and β parameters were obtained. Second, the rescaled "true" values (SSM′) were calculated using *in-situ* soil moisture measurements based on the fitted linear equation. Finally, according to Eq. 1, the ε values of each soil moisture product were expressed as $\varepsilon = SSM - SSM'$ (Table 5) i.e., the difference between the original satellite SSM soil moisture product value and the rescaled "true" values.

From Table 5, all absolute $\text{Cov}(t, \varepsilon_i)$ and $\text{Cov}(t, \varepsilon_j)$ values of triplet 1 and 2 are less than 0.00029 cm⁶/cm⁶, and most of $\text{Cov}(t, \varepsilon_i)$ and $\text{Cov}(t, \varepsilon_j)$ only are approximately 0. All $\text{Cov}(\varepsilon_i, \varepsilon_j)$ values of triplet 1 and 2 are less than 0.00352 cm⁶/cm⁶. The minimum value of $\text{Cov}(\varepsilon_i, \varepsilon_j)$ equal to 0.00049 cm⁶/cm⁶, occurred in the pair ESA-CCI and SMAP SSM. Comparing the absolute $\text{Cov}(t, \varepsilon_i)$ (or $\text{Cov}(t, \varepsilon_j)$) and $\text{Cov}(\varepsilon_i, \varepsilon_j)$ values, the absolute $\text{Cov}(t, \varepsilon_i)$ (or $\text{Cov}(t, \varepsilon_j)$) are significantly lower than the $\text{Cov}(\varepsilon_i, \varepsilon_j)$ values i.e., the influence of error non-orthogonality is smaller than the error cross-correlation. In addition, the R^2 values between errors of each pair of soil moisture products in triplet 1 and 2 are lower. Only the R^2 values between ESA-CCI and SMAP, SMOS and ASCAT are higher, 0.215 and 0.139 respectively. Therefore, in this study the estimated $\text{Cov}(\varepsilon_i, \varepsilon_j)$, $\text{Cov}(t, \varepsilon_i)$ and $\text{Cov}(t, \varepsilon_j)$ values using *in-situ* measurements when the pair strategy of SMOS, FY3-B, ASCAT, SMAP and ESA-CCI (i.e., Table 3) was used to apply the TCA algorithm were negligible suggesting the TCA assumptions did apply to our data.

Global-scale weight comparison.   First, we calculated the weight values of SMOS, FY3-B and ASCAT SSM products based on TCA (Eqs. 1~8), and merged these three SSM products by applying the linear weight fusion method with the weight values to get the 1st merged SSM product (Table 4). Then, we used the first merged SSM product, ESA-CCI, and SMAP SSM in the same way. The global-scale weight distribution maps of SMOS, FY3-B, ASCAT, ESA-CCI, and SMAP SSM products are compared and shown in Fig. 4.

In different areas, the weight values of SMOS, FY3-B, and ASCAT have a large spatial variability (Fig. 4). There are some areas without data because of less sample points used in TCA calculation in triplet 1 or 2 and permanent snow, ice and water bodies. In Australia and northern Africa, the weight value of FY3-B is higher than that of SMOS and ASCAT, which means that FY3-B is a dominant SSM product in Australia and North Africa when merging SMOS, FY3-B, and ASCAT SSM products. The RMSE of FY3-B SSM calculated by TCA

	Pairs (i, j)	Cov(t, ε_i)	Cov(t, ε_j)	Cov(ε_i, ε_j)	R^2
	SMOS, FY3-B	$2.87*10^{-04}$	$-2.48*10^{-05}$	$4.88*10^{-04}$	0.005
Triplet 1	SMOS, ASCAT	$2.87*10^{-04}$	$-2.77*10^{-07}$	$2.66*10^{-03}$	0.139
	FY3-B, ASCAT	$-2.48*10^{-05}$	$-2.77*10^{-07}$	$1.74*10^{-03}$	0.081
	1st merged SSM, ESA-CCI	$-7.46*10^{-08}$	$8.37*10^{-08}$	$8.46*10^{-04}$	0.014
Triplet 2	1st merged SSM, SMAP	$-7.46*10^{-08}$	$1.71*10^{-07}$	$2.15*10^{-03}$	0.054
	ESA-CCI, SMAP	$8.37*10^{-08}$	$1.71*10^{-07}$	$3.52*10^{-03}$	0.215

Table 5. Covariance (cm^6/cm^6) and correlation values between errors of each pair of soil moisture products in triplet 1 and 2 (i.e., the Cov(ε_i, ε_j) and R^2 values of Eq. 2), and covariance values between *in-situ* soil moisture measurements and errors of soil moisture products (i.e., the Cov(t, ε_i) and Cov(t, ε_j) values of Eq. 2).

(Eq. 6) is lower than that of SMOS and ASCAT. In northern Asia, northern America and eastern South America, the weight values of SMOS are higher than that of FY3-B and ASCAT and SMOS is a dominant SSM product in the merging process. In general, SMOS and FY3-B are dominant SSM products in the first merging process. Comparing weight values of the 1st merged SSM product, ESA-CCI, and SMAP SSM, the weight values of ESA-CCI and SMAP are higher than that of the 1st merged SSM product. Especially, in Australia, Asia, and most of America, the weight values of ESA-CCI are higher but lower than SMAP in most of northern Africa. The weight values of 1st merged SSM product are lower in most of areas (Fig. 4D). Lower weight value does represent that the RMSE values of 1st merged SSM product estimated based on TCA are higher than RMSE values of ESA-CCI and SMAP SSM products. There maybe are more potential errors in 1st merged SSM data. In 2nd merged SSM data, the large errors of 1st merged SSM data may be reduced by using lower weight values of 1st merged SSM data.

Global-scale comparison of SSM products. We compared the spatial distribution of global SMOS, FY3-B, ASCAT, ESA-CCI, SMAP, the 1st and 2nd merged SSM products on August 1, 2015 (Fig. 5). The areas of permanent snow, ice and water bodies were filtered and masked in the global SSM products using the ESA-CCI L4 2015 global LC map with 300 m SGR. Besides, we quantitatively compared the GCF values (Eq. 14) of SMOS, FY3-B, ASCAT, ESA-CCI, SMAP, the 1st and 2nd merged SSM products from 2011 to 2018 (Table 6 and Fig. 6).

Missing values in the 2nd merged SSM product are significantly less than in SMOS, FY3-B, ASCAT, ESA-CCI, and SMAP SSM (Fig. 5). The 2nd merged SSM product has almost complete coverage around the globe. In northern Asia and America, the values of the 2nd merged SSM product are lower than the 1st merged SSM. In northern Africa the 2nd merged SSM is lower than the ESA-CCI SSM. In northern Asia, most values of SMOS, ESA-CCI, SMAP, and 2nd merged SSM are in the range of 0.05~0.35 cm^3/cm^3, while most values of the ASCAT and 1st merged SSM are >0.35 cm^3/cm^3. There are many missing values in the FY3-B SSM in northern Asia (Fig. 5B). In northern Africa, the values of the 2nd merged SSM and SMAP SSM data products are between 0.03 and 0.15 cm^3/cm^3, while the ESA-CCI SSM is in the range of 0.05~0.20 cm^3/cm^3.

The GCF of the 1st and the 2nd merged SSM is higher than that of SMOS, FY3-B, ASCAT, ESA-CCI, and SMAP (Fig. 6), since gaps in a component data set are filled using retrievals in the remaining components of a triplet. The mean and maximum GCF of the 2nd merged SSM product are the highest ones, i.e., 0.67 and 0.84 respectively (Table 6). Clearly, the GCF of all SSM data products is highest in the rainy season from June to September and lowest in the dry season in January, especially for ASCAT, ESA-CCI, SMAP, 1st and 2nd merged SSM. The intra-annual GCF amplitude of SMOS and FY3-B SSM is smaller than that of ASCAT, ESA-CCI, SMAP, 1st and 2nd merged SSM. The GCF of SMOS SSM fluctuates up and down around 0.25. Comparing with GCF values of SMOS, ASCAT and SMAP retrieved from observations of single satellite sensor, the GCF of FY3-B were the lowest, and the max GCF of FY3-B was only 0.31. Overall, in terms of global coverage, the 1st and 2nd merged SSM products were improved using the TCA- LWF algorithm, particularly the 2nd merged SSM product.

Comparison with *in-situ* measurements. We used the mean value of *in-situ* measurements within a satellite SSM pixel as the reference to evaluate the satellite SSM products. We compared the SMOS, FY3-B, ASCAT, ESA-CCI, SMAP, the 1st and 2nd merged SSM with *in-situ* soil moisture measurements from 2014 to 2018 (Figs. 7, 8). In CTP_SMTMN (China), AMMA-CATCH (Benin, Niger, Mali, and DAHRA (Senegal) the soil moisture measurements from 2015 to 2016 show a clear seasonality with higher SM in the rainy season and lower in the dry season. The seasonality was also very clear in BIEBRZA_S-1 (Poland), HOBE (Denmark) and REMEDHUS (Spain), where the soil moisture was higher in winter and lower in summer especially in BIEBRZA_S-1. In RSMN (Romania) and MySMNet (Malaysia) there was no clear seasonality from 2014 to 2017. In RSMN and HOBE the soil moisture from 2014 to 2018 fluctuated around 0.2 cm^3/cm^3 and 0.25 cm^3/cm^3 respectively. In USCRN (USA) and OZNET (Australia), there were many situations when the soil moisture increased sharply from 2015 to 2018, possibly in response irrigation. Overall, the 2nd merged SSM product could capture the soil moisture dynamic characteristics caused by rainfall or irrigation, especially in CTP_SMTMN, REMEDHUS and OZNET, and was basically consistent with the temporal variability of *in-situ* soil moisture.

In CTP_SMTMN and RSMN, the 2nd merged SSM product was more consistent with *in-situ* soil moisture measurements especially from April to September. In winter from October to March of next year, however, the 2nd merged SSM underestimated the soil moisture compared with the *in-situ* measurements. The 1st merged SSM underestimated soil moisture in winter and largely overestimated in summer. In these areas the ESA-CCI
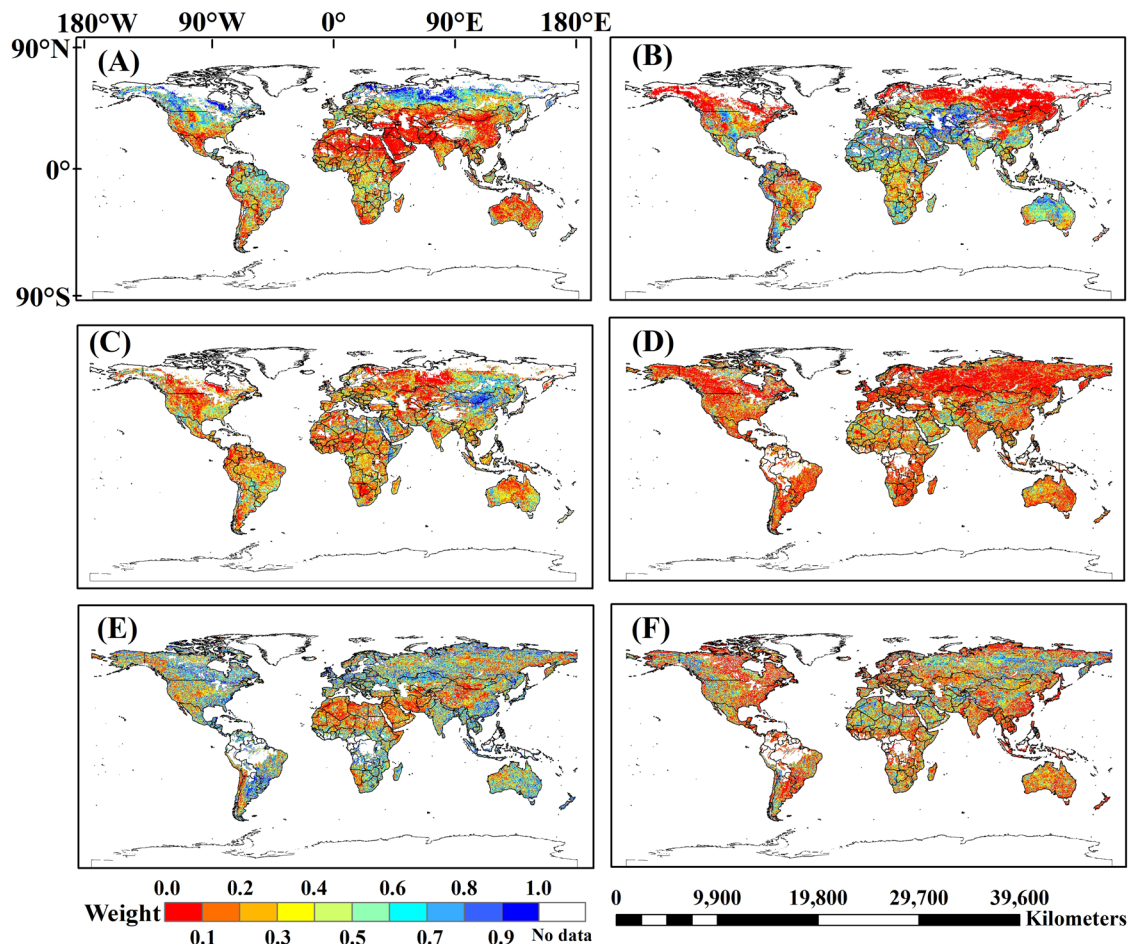
Fig. 4 Weight values of SMOS (**A**), FY3-B (**B**), ASCAT (**C**), the 1st merged SSM (**D**), ESA-CCI (**E**), and SMAP (**F**) SSM products based on TCA.

SSM product performed better than SMOS, FY3-B, ASCAT, SMAP and the 1st merged SSM. In AMMA-CATCH and DAHRA most SSM retrievals from 2014 to 2017 were largely overestimated, especially in the rainy season, except SMOS and SMAP SSM. In these areas, most SMOS SSM retrievals underestimated soil moisture. The SMAP SSM was more consistent with *in-situ* soil moisture measurements, although part of the SMAP retrievals were overestimated. In the BIEBRZA_S-1 area, almost all SSM retrievals from 2015 to 2017 were underestimated. In the BIEBRZA_S-1 area from 2015 to 2017, the mean *in-situ* soil moisture was larger than 0.4 cm³/cm³ and the highest soil moisture measurement was as high as about 0.8 cm³/cm³. The *in-situ* soil moisture measurements, therefore, may not be reliable here. In the MySMNet area, most SMOS SSM retrievals were underestimated. There were many missing soil moisture retrievals in FY3-B SSM data product in the MySMNet area. The reason may be the limited FY3-B MWRI observation range and the soil moisture retrieval algorithm applied to the FY3-B radiometric data. The ESA-CCI and SMAP SSM largely overestimated the soil moisture compared with the *in-situ* measurements.

Overall, in the MySMNet area from 2014 to 2016 all SSM retrievals performed badly against *in-situ* soil moisture measurements, while performed better in the REMEDHUS area. ESA-CCI, SMAP and the 2nd merged SSM were more consistent with *in-situ* soil moisture measurements. The SMOS SSM product largely underestimated soil moisture from 2014 to 2018, especially in the rainy season. In the same period the ASCAT SSM underestimated in the rainy season and overestimated soil moisture in the dry season. Overall, in the CTP_SMTMN, RSMN, and REMEDHUS areas, the ESA-CCI, SMAP and 2nd merged SSM was more consistent with *in-situ* soil moisture measurements.

We applied the MAE (Eq. 10), RMSE (Eq. 11), RB (Eq. 12) and *R* (Eq. 13) metrics to evaluate the SMOS, FY3-B, ASCAT, ESA-CCI, SMAP, the 1st and 2nd merged SSM against the *in-situ* soil moisture measurements at the selected sites from the ISMN (Table 7).

The 2nd merged and the SMAP SSM were more consistent with the CTP_SMTMN and RSMN soil moisture measurements. The lowest MAE, RMSE, RB values, and highest *R*-value were 0.028 cm³/cm³, 0.036 cm³/cm³, 0.019, and 0.861 respectively. In the AMMA-CATCH and DAHRA networks of Africa, the FY3-B, SMAP, and SMOS SSM retrievals were more consistent with *in-situ* soil moisture measurements than ESA-CCI, 1st and 2nd merged SSM. The MAE value was lowest for SMAP, i.e., 0.024 cm³/cm³. The RMSE and RB values of SMOS were the lowest, i.e., 0.038 cm³/cm³ and 0.083 respectively. The *R*-value was highest for FY3-B/MWRI, i.e., 0.770.
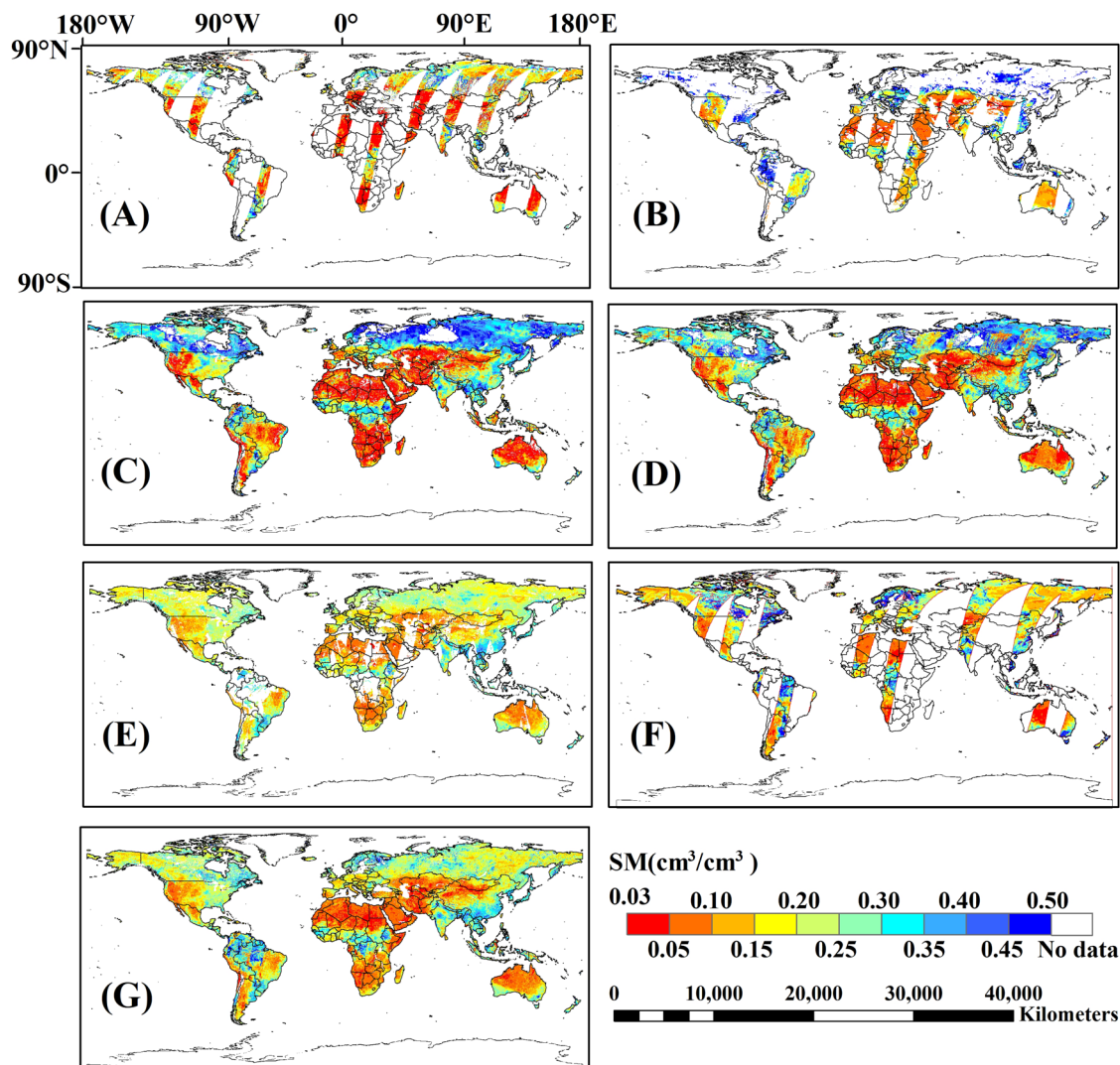
Fig. 5 Global SSM data products on August 1st, 2015: SMOS (**A**), FY3-B (**B**), ASCAT (**C**), 1st merged SSM (**D**), ESA-CCI (**E**), SMAP (**F**) and 2nd merged SSM (**G**).

GCF	SMOS	FY3-B	ASCAT	1st merged SSM	ESA-CCI	SMAP	2nd merged SSM
Min	0.049	0.002	0.328	0.360	0.170	0.022	0.278
Max	0.363	0.310	0.745	0.820	0.596	0.501	0.837
Average	0.291	0.195	0.567	0.616	0.398	0.308	0.672

Table 6. Min, max and mean GCF of global SMOS, FY3-B, ASCAT, ESA-CCI, SMAP, 1st and 2nd merged SSM products from 2011 to 2018.

The 2nd merged SSM was more consistent than ESA-CCI with the AMMA-CATCH and DAHRA soil moisture measurements. In BIEBRZA_S-1 and MySMNet networks, the difference between soil moisture measurements and all SSM retrievals was very large. Especially, in the BIEBRZA_S-1 network, all SSM products performed very badly, largely underestimating measured soil moisture. In the REMEDHUS network, the ESA-CCI, SMAP, and 2nd merged SSM were more consistent with soil moisture measurements. The MAE and RMSE were lowest for ESA-CCI, i.e., 0.038 cm³/cm³ and 0.049 cm³/cm³ respectively. The RB value was lowest for the 2nd merged SSM, i.e., −0.013, while the R-value was highest for SMAP, i.e., 0.803. In HOBE, USCRN and OZNET networks, although the MAE and RMSE values of 2nd merged SSM were slightly larger than ESA-CCI SSM, the MAE and RMSE values of 2nd merged SSM were clearly less than SMOS, FY3-B, ASCAT, 1st merged and SMAP SSM.

Overall, in the 10 networks CTP_SMTMN, RSMN, AMMA-CATCH, DAHRA, BIEBRZA_S-1, MySMNet, REMEDHUS, HOBE, USCRN and OZNET, the mean values of MAE, RMSE and BR values of the 2nd merged SSM were lower than SMOS, FY3-B, ASCAT, ESA-CCI, SMAP and 1st merged SSM (Table 7). For example, the mean of all MAE values of 2nd merged SSM was 0.094 cm³/cm³ less than 0.120 cm³/cm³ of SMOS, 0.132 cm³/
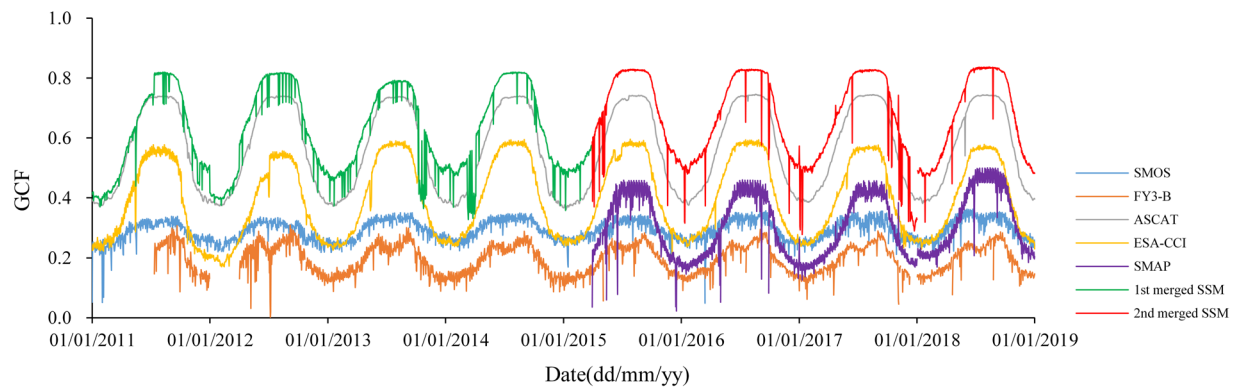
Fig. 6 Daily GCF during 2011~2018 of global SMOS, FY3-B, ASCAT, ESA-CCI, SMAP, 1st and 2nd merged SSM (2015~2018 only).
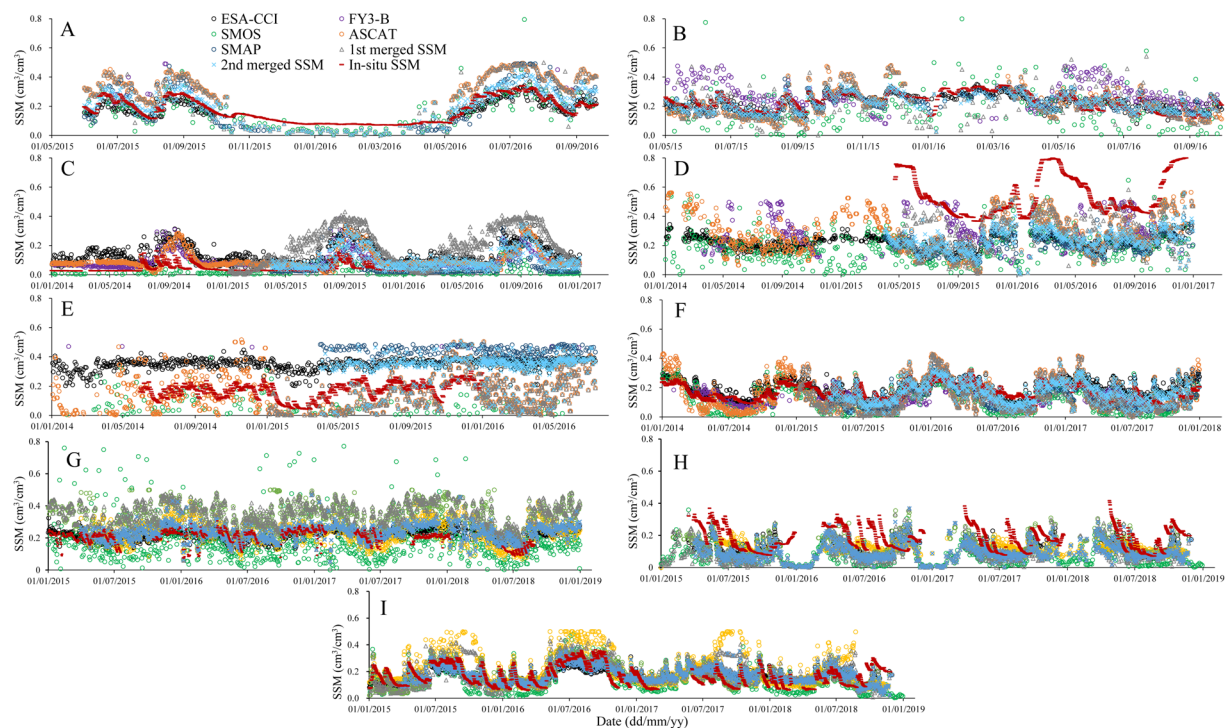

Fig. 7 Temporal evolution of *in-situ*, SMOS, FY3-B, ASCAT, ESA-CCI, SMAP, the 1st and 2nd merged SSM in the seven reference areas ((**A**) CTP_SMTMN; (**B**) RSMN; (**C**) AMMA-CATCH and DAHRA; (**D**) BIEBRZA_S-1; (**E**) MySMNet; (**F**) REMEDHUS; (**G**) HOBE; (**H**) USCRN; (**I**) OZNET).

cm^3 of FY3-B, 0.114 cm^3/cm^3 of ASCAT, 0.110 cm^3/cm^3 of 1st merged SSM, 0.105 cm^3/cm^3 of ESA-CCI and 0.105 cm^3/cm^3 of SMAP. Although, the mean value (0.633) of all R values of 2nd merged SSM was less than ESA-CCI (0.645) and SMAP (0.697), it was higher than SMOS (0.502), FY3-B (0.260), ASCAT (0.573) and 1st merged SSM (0.564). In addition, except the MySMNet and BIEBRZA_S-1 networks, the 2nd merged SSM was more consistent with soil moisture measurements than the 1st merged SSM. It should be noted that in AMMA-CATCH and DAHRA networks, the SMOS and SMAP SSM retrievals from observations by a single microwave radiometer were more consistent with *in-situ* soil moisture measurements than the ESA-CCI, 1st and 2nd merged SSM. Overall, the 2nd merged SSM was more consistent than ESA-CCI with *in-situ* soil moisture measurements at most sites.

To improve our validation study on GDSMFD product, we have expanded significantly the soil moisture reference data set used for this purpose (see the Supplementary file). We used soil moisture retrievals based on the area-scale airborne radiometer observation by SCA-V (Single Channel Algorithm-Vertical polarization) soil moisture retrieval algorithm. Also, we used two global-scale merged soil moisture data, i.e., the NNsm and RSSSM data products based on the neural network fusion algorithm. We deemed potentially confusing to mix the validation against actual soil moisture measurements with the comparison with retrievals and other data products. Accordingly, the comparison of our merged soil moisture data of this study with the airborne
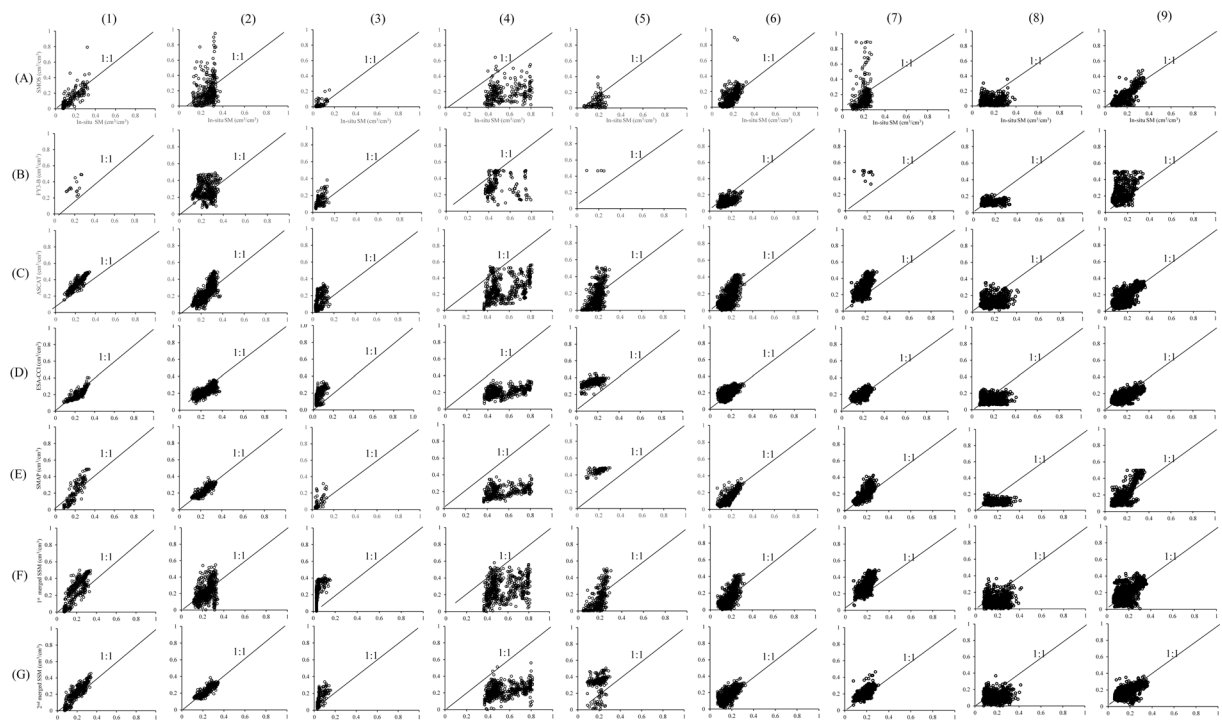
Fig. 8 Scatter plots of SSM retrievals: SMOS (**A**), FY3-B (**B**), ASCAT (**C**), ESA-CCI (**D**), SMAP (**E**), 1st merged (**F**) and 2nd merged (**G**) versus *in-situ* soil moisture measurements vs at: (1) CTP_SMTMN, (2) RSMN, (3) AMMA-CATCH and DAHRA, (4) BIEBRZA_S-1, (5) MySMNet, (6) REMEDHUS, (7) HOBE, (8) USCRN, (9) OZNET.

retrievals and the NNsm and RSSSM data products are presented in the document on supplementary materials. Comparing with the airborne radiometer observation-based soil moisture data, the 2nd merged soil moisture product had better performance in term of consistence with the airborne radiometer observation-based soil moisture data than FY3-B, ASCAT, ESA-CCI and 1st merged soil moisture products. In addition, the spatial distribution pattern of GDSMFD product has basically consistent spatial pattern with vegetation coverage data in global scale, and also is same with NNsm and RSSSM products.

Usage Notes

We present a global daily-scale soil moisture fusion dataset generated through the TCA-LWF. It is possible that in any given pixel only one or two of SMOS, FY3-B and ASCAT SSM is available for 1st merged SSM data, or only one or two of 1st merged SSM, ESA-CCI and SMAP SSM data is available for the 2nd merged SSM data. Therefore, it is helpful to note the spatial distribution of SSM data products for use and further improvement of the merged SSM generated by this study in future. Taking August 1st, 2015 as an example, we produced the distribution flags of SMOS, FY3-B and ASCAT SSM data for the 1st merged SSM product (Fig. 9A), and the distribution flags of the 1st merged SSM, ESA-CCI and SMAP SSM data for the 2nd merged SSM product (Fig. 9B). The areas where SMOS, FY3-B and ASCAT SSM data overlap are smaller than the areas where either SMOS and FY3-B or FY3-B and ASCAT SSM data do (Fig. 9A). In addition, there are many areas where only ASCAT SSM is available (Yellow areas in Fig. 9A). In these areas, the errors in the 1st merged SSM data are only due to errors in the ASCAT SSM data. In the 2nd merged SSM data, there are many areas with overlapping coverage (Deep blue areas in Fig. 9B) by the 1st merged SSM, ESA-CCI and SMAP and by the 1st merged SSM and ESA-CCI (Green areas in Fig. 9B). In northern South America and central Africa (Red areas in Fig. 9B), there is only the 1st merged SSM data available i.e., in these areas the errors of 2nd merged SSM data only come from the 1st merged SSM data.

In addition, although the final 2nd merged SSM product has better performance in accuracy than SMOS, FY3-B, ASCAT, SMAP and even than ESA-CCI SSM (Fig. 6 and Table 7), there are overestimations or underestimations against *in-situ* soil moisture measurements especially with the data collected in the AMMA-CATCH and DAHRA, BIEBRZA_S-1 and MySMNet networks. According to the LWF algorithm (Eqs. 7~9), the errors leading to overestimations or underestimations in the final 2nd merged SSM are from SMOS, FY3-B, ASCAT, ESA-CCI and SMAP SSM. Therefore, for further improvement of the merged SSM product generated by this study, it is also necessary to discuss the potential error sources in SSM data products, which is dealt with in the following sections.

Errors derived from ESA-CCI SSM. We found that the SMOS, FY3-B and SMAP SSM retrievals from observations by a single microwave radiometer were more consistent with *in-situ* soil moisture measurements in AMMA-CATCH and DAHRA than the ESA-CCI SSM (Figs. 7 and 8). Most soil moisture values of ESA-CCI SSM

Indies	Networks	SMOS	FY3-B	ASCAT	1st merged SSM	ESA-CCI	SMAP	2nd merged SSM
MAE	CTP_SMTMN	0.102	0.120	0.102	0.098	0.070	0.074	0.064
	RSMN	0.107	0.100	0.109	0.075	0.032	0.028	0.028
	ACD	0.033	0.053	0.063	0.059	0.088	0.024	0.048
	BIEBRZA_S-1	0.381	0.156	0.298	0.276	0.338	0.303	0.287
	MySMNet	0.091	/	0.087	0.094	0.181	0.207	0.152
	REMEDHUS	0.075	0.065	0.071	0.071	0.038	0.049	0.040
	HOBE	0.134	0.282	0.135	0.141	0.060	0.071	0.072
	USCRN	0.094	0.138	0.095	0.095	0.079	0.099	0.085
	OZNET	0.064	0.138	0.067	0.080	0.061	0.086	0.072
	Mean	*0.120*	*0.132*	*0.114*	*0.110*	*0.105*	*0.105*	*0.094*
RMSE	CTP_SMTMN	0.124	0.140	0.115	0.114	0.080	0.090	0.075
	RSMN	0.126	0.120	0.150	0.098	0.041	0.036	0.036
	ACD	0.038	0.065	0.086	0.075	0.104	0.043	0.069
	BIEBRZA_S-1	0.410	0.221	0.340	0.315	0.360	0.331	0.313
	MySMNet	0.109	/	0.108	0.109	0.186	0.216	0.161
	REMEDHUS	0.093	0.082	0.082	0.083	0.049	0.058	0.051
	HOBE	0.188	0.291	0.155	0.161	0.074	0.085	0.087
	USCRN	0.112	0.158	0.112	0.114	0.092	0.111	0.101
	OZNET	0.082	0.175	0.083	0.099	0.076	0.109	0.088
	Mean	*0.142*	*0.157*	*0.137*	*0.130*	*0.118*	*0.120*	*0.109*
RB	CTP_SMTMN	0.382	0.290	0.258	0.258	0.180	0.058	0.053
	RSMN	0.437	0.200	0.400	0.048	0.065	0.019	0.068
	ACD	0.083	0.553	0.573	0.561	0.671	0.185	0.493
	BIEBRZA_S-1	−1.895	−0.379	−1.122	−0.907	−1.596	−1.280	−1.372
	MySMNet	−0.468	/	0.075	0.272	0.516	0.522	0.382
	REMEDHUS	−0.449	−0.449	−0.187	−0.257	0.125	−0.283	−0.013
	HOBE	−0.119	0.572	0.315	0.330	0.010	0.056	0.126
	USCRN	−0.274	0.273	−0.076	0.001	0.112	0.053	0.081
	OZNET	−0.039	0.459	0.189	0.229	0.133	0.255	0.141
	Mean	*−0.260*	*0.190*	*0.047*	*0.059*	*0.024*	*−0.046*	*−0.005*
R	CTP_SMTMN	0.639	0.560	0.806	0.819	0.790	0.824	0.766
	RSMN	0.497	0.110	0.266	0.493	0.772	0.821	0.861
	ACD	0.644	0.770	0.595	0.590	0.699	0.750	0.718
	BIEBRZA_S-1	0.293	−0.028	0.371	0.319	0.445	0.528	0.331
	MySMNet	0.385	/	0.646	0.635	0.599	0.450	0.398
	REMEDHUS	0.637	0.372	0.744	0.729	0.708	0.803	0.691
	HOBE	0.155	−0.162	0.540	0.499	0.541	0.696	0.633
	USCRN	0.522	0.015	0.455	0.366	0.528	0.627	0.573
	OZNET	0.743	0.443	0.736	0.626	0.720	0.775	0.730
	Mean	*0.502*	*0.260*	*0.573*	*0.564*	*0.645*	*0.697*	*0.633*

Table 7. MAE (cm³/cm³), RMSE (cm³/cm³), RB and *R* values of SMOS, FY3-B, ASCAT, ESA-CCI, SMAP, the 1st and 2nd merged SSM against *in-situ* soil moisture measurements from ISMN. (ACD: AMMA-CATCH and DAHRA networks).

are overestimated compared with *in-situ* soil moisture measurements in AMMA-CATCH and DAHRA especially in the rainy season. The calculated RMSE value (0.104 cm³/cm³) of ESA-CCI SSM (Table 7) is higher than the RMSE values of SMOS, FY3-B and SMAP, i.e., 0.038, 0.065 and 0.043 cm³/cm³ respectively. In addition, most ESA-CCI SSM weight values were lower, less than 0.3 or even less than 0.1 in northern Africa, than the SMAP SSM weight values, which were higher than 0.3 (Fig. 4). The weight values were calculated by Eqs. 7 and 8 with the RMSE values estimated using TCA and decrease with increasing RMSE. In the ESA-CCI SSM data product, all the SSM retrievals with passive microwave data such as AMSR-E, SMOS and AMSR-2 SSM were done with the same retrieval algorithm i.e., LPRM. Some authors indicated however, that the AMSR-E, AMSR-2 and SMOS SSM retrievals with LRPM overestimated the soil moisture, and were inconsistent with *in-situ* soil moisture measurements[9,26,55–57]. For example, Cui *et al*.[26] found that the AMSR-2 SSM retrieved using LPRM largely overestimated the soil moisture with bias >0.09 cm³/cm³ in the Little Washita Watershed and REMEDHUS networks. It was also indicated that the vegetation optical depth was overestimated by LPRM leading to overestimation of soil moisture retrieved using LPRM[26]. In AMMA-CATCH and DAHRA network, the main land cover type is shrubland with sparse vegetation and low vegetation optical depth. Accordingly, one of the reasons of overestimation of ESA-CCI SSM data in northern Africa is the overestimation of vegetation optical depth by LPRM. Therefore, in
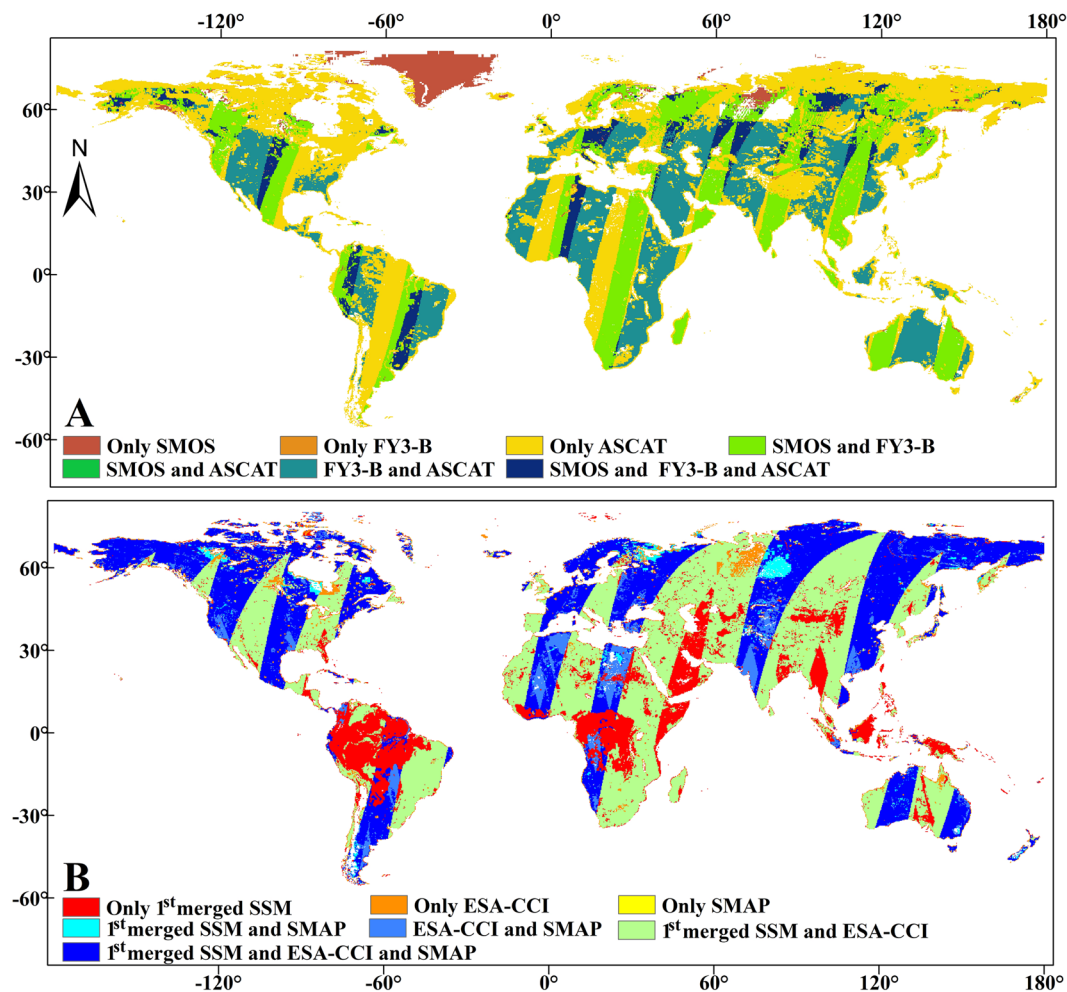
Fig. 9 Distribution flags of SMOS, FY3-B, ASCAT, ESA-CCI and SMAP SSM data on August 1, 2015 in: (**A**) the 1st merged SSM data; (**B**) the 2nd merged SSM data.

areas with low vegetation, the LPRM soil moisture retrieval may not be the most accurate. The LPRM may lead to errors introduced into the ESA-CCI product, because LPRM does not perform well in global SSM retrieval with microwave radiometers. So, it is not unlikely that errors caused by LPRM in the ESA-CCI SSM data product were introduced in our 2nd merged SSM data.

Errors derived from ASCAT SSM. The GCF of ASCAT SSM is higher than SMOS, FY3-B, SMAP and even than ESA-CCI (Fig. 6). There are many areas where only ASCAT SSM is available (Fig. 9). Therefore, ASCAT SSM plays an important role in our 1st and 2nd merged SSM data, especially in the 1st merged SSM. From ASCAT SSM product retrieved using SWI algorithm with C-band observations (5.3 GHz) gave a worse performance (Table 7) than SMOS, FY3-B and SMAP SSM and overestimated soil moisture compared with the *in-situ* measurements in the CTP_SMTMN, RSMN, AMMA-CATCH and DAHRA, MySMNet networks. Other studies drew similar conclusions, i.e. although ASCAT SSM was correlated with *in-situ* soil moisture data, there was an issue of overestimation against *in-situ* soil moisture measurements[9]. One possible cause leading to overestimation in the ASCAT SSM could be errors in the global soil porosity data with 1-degree SGR used to convert the ASCAT SSM from degree of saturation (%) into volumetric soil moisture content (cm³/cm³)[10,58,59]. In this study, the adopted soil porosity data (Fig. 10) was estimated using equations developed by Saxton and Rawls (2006), who took sand, clay, silt, and organic matter as input. In most areas of the globe soil porosity is between the 0.4 and 0.5 (Fig. 10). There are a few areas such as the northern Asia and northern North America where soil porosity is higher than 0.5 and there almost no areas where soil porosity is less than 0.35 (Fig. 10). In the future, if the soil porosity data will be corrected using *in-situ* clay, silt, and organic matter fractions, ASCAT SSM might be improved and might also lead to improvement in our 1st and 2nd merged SSM product.

Errors derived from SMOS, FY3-B and SMAP SSM. Although the soil moisture retrieval algorithms of SMOS, FY3-B and SMAP SSM products are different, they are based on the same radiative transfer equation (Table 8), i.e. the zero-order radiative transfer τ-w model[60]. The difference between these soil moisture retrieval algorithms is that the applied parameters of the zero-order radiative transfer model to get best SSM retrieval are
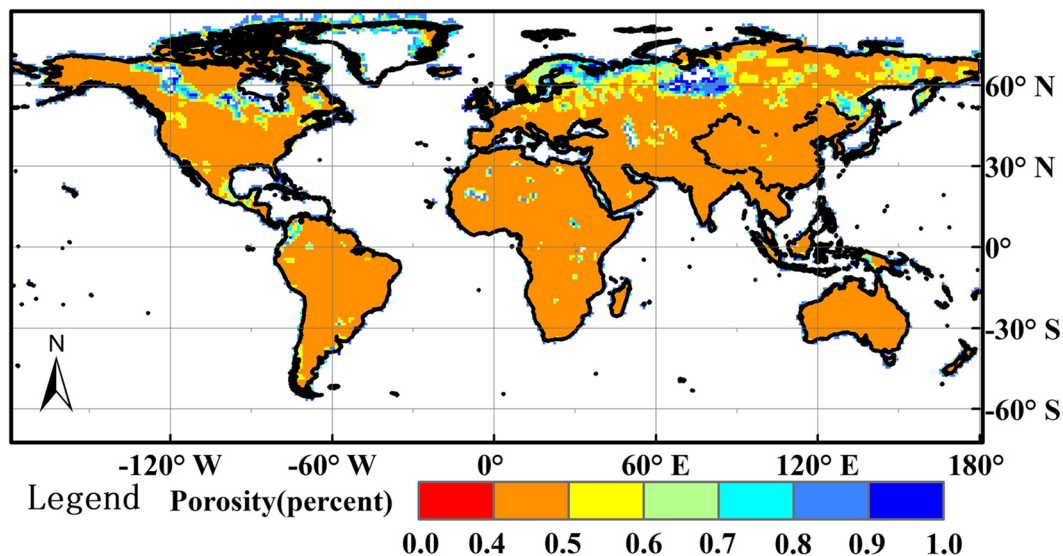
Fig. 10 Global soil porosity map[53].

Foundational model	Zero-order radiative transfer model (i.e.,$\tau-w$ model)		
Algorithm	L-MEB for SMOS	Qp model for FY3-B	SCA-V for SMAP
Brightness temperature (TB)	1.4 GHz with muti-angles (2.5 ± 2.5 to 62.5 ± 2.5°)	10.65 and 36.5 GHz	Interpolated TB at 1.41GHz using the Backus-Gilbert method
Soil roughness model	QHN model: $Q=0$; $N_{H-p}=N_{V-p}=-1$ (low vegetation); $N_{H-p}=1, N_{V-p}=-1$ (V-p), (forest); H: from LUT based on IGBP ($0.02 \leq H \leq 0.3$)	Qp model: The coefficients of Qp model are estimated using simulated data from AIEM	QHN model: $Q=0$; $N=2$; H: from LUT based on IGBP ($0.083 \leq H \leq 0.160$)
Vegetation optical depth (τ)	τ is retrieved in L-MEB with soil moisture simultaneously	$\tau = b \times VWC/\cos\theta$: VWC: from an empirical function of LAI; θ: the incident angle.	$\tau = b \times VWC$: b: from LUT based on IGBP ($0 \leq b \leq 0.13$); VWC: from the nonlinear function of NDVI
Soil temperature (Ts)	from ECMWF (Level 1~3)	from an empirical linear function of TB at 36.5 GHz	from the GMAO-GEOS-FP (Level 1: 5~15cm and Level 2: 15–35 cm)
Canopy temperature (Tv)	from ECMWF (Level 1: top 0–7 cm)	$Tv = Ts$	$Tv = Ts$
Single scattering albedo (w)	from LUT based on IGBP ($0.06 \leq w \leq 0.12$)	$w = 0$	from LUT based on IGBP ($0 \leq w \leq 0.07$)
Soil dielectric model	Mironov *et al.*	/	Mironov *et al.*

Table 8. Main inputs of the retrieval algorithms of SMOS, FY3-B and SMAP SSM. *Q, H and N are the parameters of QHN model; H-p: Horizontal-polarization; V-p: Vertical-polarization; IGBP: the International Geosphere-Biosphere Programme; ECMWF: the European Center for Medium range Weather Forecasting; AIEM: the Advanced Integral Equation Method; VWC: the Vegetation Water Content; LAI: the Leaf Area Index; NDVI: the Normalized Difference Vegetation Index; GMAO-GEOS-FP: the Goddard Modelling and Assimilation Office-Goddard Earth Observing System (model)-Forward Processing; LUT: Look-Up-Table.

different. The disparities in brightness temperature, soil roughness model, vegetation optical depth, soil temperature, vegetation canopy temperature, vegetation single scattering albedo, soil dielectric model lead to the difference between SMOS, FY3-B and SMAP SSM[56]. These main inputs to the retrieval algorithms of SMOS, FY3-B and SMAP SSM also produce errors, which are transferred to our 1st and 2nd merged SSM data. Further detailed analysis of the error sources in the SMOS, FY3-B and SMAP SSM data would be helpful to improve the merged SSM data products described in this study.

Code availability

The codes used for Global Daily-scale Soil Moisture Fusion Dataset (GDSMFD) are available in National Tibetan Plateau Data Center (https://doi.org/10.11888/RemoteSen.tpdc.271988)[61].

References

1. Das, K. & Paul, P. K. Present status of soil moisture estimation by microwave remote sensing. *Cogent Geosci.* **1**, 1–21, https://doi.org/10.1080/23312041.2015.1084669 (2015).

2. Petropoulos, G. P., Ireland, G. & Barrett, B. Surface soil moisture retrievals from remote sensing: Current status, products & future trends. *Phys. Chem. Earth* **83–84**, 36–56, https://doi.org/10.1016/j.pce.2015.02.009 (2015).

3. 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. *Adv. Water Resour.* **109**, 236–252, https://doi.org/10.1016/j.advwatres.2017.09.010 (2017).

4. Xie, Q., Menenti, M. & Jia, L. Improving the AMSR-E/NASA soil moisture data product using *in-situ* measurements from the Tibetan Plateau. *Remote Sens.* **11**, 7085–7088, https://doi.org/10.3390/rs11232748 (2019).

5. Mohanty, B. P., Cosh, M. H., Lakshmi, V. & Montzka, C. Soil Moisture Remote Sensing: State-of-the-Science. *Vadose Zo. J.* **16**, vzj2016.10.0105, https://doi.org/10.2136/vzj2016.10.0105 (2017).

6. Kerr, Y. H. *et al.* Overview of SMOS performance in terms of global soil moisture monitoring after six years in operation. *Remote Sens. Environ.* **180**, 40–63, https://doi.org/10.1016/j.rse.2016.02.042 (2016).

7. Kim, H. *et al.* Global scale error assessments of soil moisture estimates from microwave-based active and passive satellites and land surface models over forest and mixed irrigated/dryland agriculture regions. *Remote Sens. Environ.* **251**, 112052, https://doi.org/10.1016/j.rse.2020.112052 (2020).

8. Kim, H. *et al.* Global-scale assessment and combination of SMAP with ASCAT (active) and AMSR2 (passive) soil moisture products. *Remote Sens. Environ.* **204**, 260–275, https://doi.org/10.1016/j.rse.2017.10.026 (2018).

9. Zeng, J. *et al.* Evaluation of remotely sensed and reanalysis soil moisture products over the Tibetan Plateau using *in-situ* observations. *Remote Sens. Environ.* **163**, 91–110, https://doi.org/10.1016/j.rse.2015.03.008 (2015).

10. Al-Yaari, A. *et al.* Assessment and inter-comparison of recently developed/reprocessed microwave satellite soil moisture products using ISMN ground-based measurements. *Remote Sens. Environ.* **224**, 289–303, https://doi.org/10.1016/j.rse.2019.02.008 (2019).

11. Srivastava, P. K., Han, D., Rico-Ramirez, M. A., Al-Shrafany, D. & Islam, T. Data Fusion Techniques for Improving Soil Moisture Deficit Using SMOS Satellite and WRF-NOAH Land Surface Model. *Water Resour. Manag.* **27**, 5069–5087, https://doi.org/10.1007/s11269-013-0452-7 (2013).

12. Kim, S., Parinussa, R. M., Liu, Y. Y., Johnson, F. M. & Sharma, A. Merging alternate remotely-sensed soil moisture retrievals using a non-static model combination approach. *Remote Sens.* **8**, 1–16, https://doi.org/10.3390/rs8060518 (2016).

13. Das, N. N., Entekhabi, D., Kim, S., Yueh, S. & O'Neill, P. Combining SMAP and Sentinel data for high-resolution Soil Moisture product. in *International Geoscience and Remote Sensing Symposium (IGARSS)* 129–131, https://doi.org/10.1109/IGARSS.2016.7729024 (2016).

14. Kim, S., Parinussa, R. M., Liu, Y. Y., Johnson, F. M. & Sharma, A. A framework for combining multiple soil moisture retrievals based on maximizing temporal correlation. *Geophys. Res. Lett.* **42**, 6662–6670, https://doi.org/10.1002/2015GL064981 (2015).

15. Liu, Y. Y. *et al.* Trend-preserving blending of passive and active microwave soil moisture retrievals. *Remote Sens. Environ.* **123**, 280–297, https://doi.org/10.1016/j.rse.2012.03.014 (2012).

16. Santi, E. *et al.* On the synergy of SMAP, AMSR2 AND SENTINEL-1 for retrieving soil moisture. *Int. J. Appl. Earth Obs. Geoinf.* **65**, 114–123, https://doi.org/10.1016/j.jag.2017.10.010 (2018).

17. Srivastava, P. K., Han, D., Ramirez, M. R. & Islam, T. Machine Learning Techniques for Downscaling SMOS Satellite Soil Moisture Using MODIS Land Surface Temperature for Hydrological Application. *Water Resour. Manag.* **27**, 3127–3144, https://doi.org/10.1007/s11269-013-0337-9 (2013).

18. Yilmaz, M. T., Crow, W. T., Anderson, M. C. & Hain, C. An objective methodology for merging satellite- and model-based soil moisture products. *Water Resour. Res.* **48**, 0–45, https://doi.org/10.1029/2011WR011682 (2012).

19. Stoffelen, A. Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *J. Geophys. Res. C Ocean.* **103**, 7755–7766, https://doi.org/10.1029/97jc03180 (1998).

20. Gruber, A., Dorigo, W. A., Crow, W. & Wagner, W. Triple Collocation-Based Merging of Satellite Soil Moisture Retrievals. *IEEE Trans. Geosci. Remote Sens.* **55**, 6780–6792, https://doi.org/10.1109/TGRS.2017.2734070 (2017).

21. Kerr, Y. H. *et al.* The SMOS soil moisture retrieval algorithm. *IEEE Trans. Geosci. Remote Sens.* **50**, 1384–1403, https://doi.org/10.1109/TGRS.2012.2184548 (2012).

22. Wigneron, J. P. *et al.* SMOS-IC: Current status and overview of soil moisture and VOD applications. *International Geoscience and Remote Sensing Symposium (IGARSS)* **1**, 1451–1454, https://doi.org/10.1109/IGARSS.2018.8519382 (2018).

23. Wigneron, J. P. *et al.* Modelling the passive microwave signature from land surfaces: A review of recent results and application to the L-band SMOS & SMAP soil moisture retrieval algorithms. *Remote Sens. Environ.* **192**, 238–262, https://doi.org/10.1016/j.rse.2017.01.024 (2017).

24. Mattar, C. *et al.* A combined optical-microwave method to retrieve soil moisture over vegetated areas. *IEEE Trans. Geosci. Remote Sens.* **50**, 1404–1413, https://doi.org/10.1109/TGRS.2011.2179051 (2012).

25. Wigneron, J. P. *et al.* First evaluation of the simultaneous SMOS and ELBARA-II observations in the Mediterranean region. *Remote Sens. Environ.* **124**, 26–37, https://doi.org/10.1016/j.rse.2012.04.014 (2012).

26. Cui, C. *et al.* Soil moisture mapping from satellites: An intercomparison of SMAP, SMOS, FY3B, AMSR2, and ESA CCI over two dense network regions at different spatial scales. *Remote Sens.* **10**, 1–19, https://doi.org/10.3390/rs10010033 (2018).

27. Shi, J. *et al.* Physically based estimation of bare-surface soil moisture with the passive radiometers. *IEEE Trans. Geosci. Remote Sens.* **44**, 3145–3152, https://doi.org/10.1109/TGRS.2006.876706 (2006).

28. Naeimi, V., Scipal, K., Bartalis, Z., Hasenauer, S. & Wagner, W. An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations. *IEEE Trans. Geosci. Remote Sens.* **47**, 1999–2013, https://doi.org/10.1109/TGRS.2008.2011617 (2009).

29. Wagner, W., Lemoine, G. & Rott, H. A method for estimating soil moisture from ERS Scatterometer and soil data. *Remote Sens. Environ.* **70**, 191–207, https://doi.org/10.1016/S0034-4257(99)00036-X (1999).

30. Wagner, W. *et al.* Fusion of active and passive microwave observations to create an essential climate variable data record on soil moisture. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **1**, 315–321, https://doi.org/10.5194/isprsannals-I-7-315-2012 (2012).

31. Dorigo, W. A. *et al.* Evaluation of the ESA CCI soil moisture product using ground-based observations. *Remote Sens. Environ.* **162**, 380–395, https://doi.org/10.1016/j.rse.2014.07.023 (2015).

32. Dorigo, W. *et al.* ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sens. Environ.* **203**, 185–215, https://doi.org/10.1016/j.rse.2017.07.001 (2017).

33. 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**, 717–739, https://doi.org/10.5194/essd-11-717-2019 (2019).

34. Entekhabi, D., Reichle, R. H., Koster, R. D. & Crow, W. T. Performance metrics for soil moisture retrievals and application requirements. *J. Hydrometeorol.* **11**, 832–840, https://doi.org/10.1175/2010JHM1223.1 (2010).

35. O'Neill *et al. SMAP Enhanced L3 Radiometer Global Daily 9 km EASE-Grid Soil Moisture, Version* **4**. https://doi.org/10.5067/T90W6VRLCBHI (2020).

36. Santi, E. *et al.* Integration of passive and active microwave data from SMAP, AMSR2 and Sentinel-1 for Soil Moisture monitoring. *Int. Geosci. Remote Sens. Symp.* 2016-Novem, 5252–5255, https://doi.org/10.1109/IGARSS.2016.7730368 (2016).

37. Jackson, T. J. III. Measuring surface soil moisture using passive microwave remote sensing. *Hydrol. Process.* **7**, 139–152, https://doi.org/10.1002/hyp.3360070205 (1993).

38. Chan, S. K. *et al.* Development and assessment of the SMAP enhanced passive soil moisture product. *Remote Sens. Environ.* **204**, 931–941, https://doi.org/10.1016/j.rse.2017.08.025 (2018).
39. Dorigo, W. A. *et al.* 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 (2011).
40. Yang, K. *et al.* A multiscale soil moisture and freeze-thaw monitoring network on the third pole. *Bull. Am. Meteorol. Soc.* **94**, 1907–1916, https://doi.org/10.1175/BAMs-d-12-00203.1 (2013).
41. Sandric, I. *et al.* Validation and upscaling of soil moisture satellite products in Romania. in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* vols XLI-B2 313–317, https://doi.org/10.5194/isprsarchives-XLI-B2-313-2016 (2016).
42. Pellarin, T. *et al.* Hydrological modelling and associated microwave emission of a semi-arid region in South-western Niger. *J. Hydrol.* **375**, 262–272, https://doi.org/10.1016/j.jhydrol.2008.12.003 (2009).
43. Tagesson, T. *et al.* Ecosystem properties of semiarid savanna grassland in West Africa and its relationship with environmental variability. *Glob. Chang. Biol.* **21**, 250–264, https://doi.org/10.1111/gcb.12734 (2015).
44. Musial, J. *et al.* Derivation and validation of the high resolution satellite soil moisture products: a case study of the Biebrza Sentinel-1 validation sites. *Geoinf. Issues* **8**, 37–53 (2016).
45. Kang, C. S., Kanniah, K. D., Kerr, Y. H. & Cracknell, A. P. Analysis of *in-situ* soil moisture data and validation of SMOS soil moisture products at selected agricultural sites over a tropical region. *Int. J. Remote Sens.* **37**, 3636–3654, https://doi.org/10.1080/01431161.2016.1201229 (2016).
46. Martínez-Fernández, J. & Ceballos, A. Mean soil moisture estimation using temporal stability analysis. *J. Hydrol.* **312**, 28–38, https://doi.org/10.1016/j.jhydrol.2005.02.007 (2005).
47. Bircher, S., Skou, N., Jensen, K. H., Walker, J. P. & Rasmussen, L. A soil moisture and temperature network for SMOS validation in Western Denmark. *Hydrol. Earth Syst. Sci.* **16**, 1445–1463, https://doi.org/10.5194/hess-16-1445-2012 (2012).
48. Bell, J. E. *et al.* U.S. climate reference network soil moisture and temperature observations. *J. Hydrometeorol.* **14**, 977–988, https://doi.org/10.1175/JHM-D-12-0146.1 (2013).
49. Young, R. *et al. Soil moisture and meteorological observations from the Murrumbidgee catchment.* http://www.oznet.org.au/documentation/ (2008).
50. McColl, K. A. *et al.* Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target. *Geophys. Res. Lett.* **41**, 6229–6236, https://doi.org/10.1002/2014GL061322 (2014).
51. Dorigo, W. *et al.* Merging active and passive microwave soil moisture data to construct long-term time series in support of climate studies. *ESA Living Planet Symp.* **12**, 7 (2010).
52. Ma, H. *et al.* Satellite surface soil moisture from SMAP, SMOS, AMSR2 and ESA CCI: A comprehensive assessment using global ground-based observations. *Remote Sens. Environ.* **231**, 111215, https://doi.org/10.1016/j.rse.2019.111215 (2019).
53. Saxton, K. E. & Rawls, W. J. Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. *Soil Sci. Soc. Am. J.* **70**, 1569–1578, https://doi.org/10.2136/sssaj2005.0117 (2006).
54. Jia, L., Xie, Q. & Hu, G. Global daily-scale soil moisture fusion dataset based on Triple Collocation Analysis (2011-2018). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Terre.tpdc.271935 (2021).
55. Fu, H., Zhou, T. & Sun, C. Evaluation and Analysis of AMSR2 and FY3B Soil Moisture Products by an *In Situ* Network in Cropland on Pixel Scale in the Northeast of China. *Remote Sens.* **11**, 868, https://doi.org/10.3390/rs11070868 (2019).
56. Liu, J. *et al.* Evaluation of SMAP, SMOS-IC, FY3B, JAXA, and LPRM Soil moisture products over the Qinghai-Tibet Plateau and Its surrounding areas. *Remote Sens.* **11**, 792–792, https://doi.org/10.3390/rs11070792 (2019).
57. Ma, W. & Ma, Y. The evaluation of AMSR-E soil moisture data in atmospheric modeling using a suitable time series iteration to derive land surface fluxes over the Tibetan Plateau. *PLoS One* **14**, 1–10, https://doi.org/10.1371/journal.pone.0226373 (2019).
58. Wagner, W. *et al.* The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications. *Meteorol. Zeitschrift* **22**, 5–33, https://doi.org/10.1127/0941-2948/2013/0399 (2013).
59. Su, Z., De Rosnay, P., Wen, J., Wang, L. & Zeng, Y. Evaluation of ECMWF's soil moisture analyses using observations on the Tibetan Plateau. *J. Geophys. Res. Atmos.* **118**, 5304–5318, https://doi.org/10.1002/jgrd.50468 (2013).
60. Mo, T., Choudhury, B. J., Schmugge, T. J., Wang, J. R. & Jackson, T. J. A model for microwave emission from vegetation-covered fields. *J. Geophys. Res.* **87**, 11229–11237, https://doi.org/10.1029/JC087iC13p11229 (1982).
61. Xie, Q., Jia, L. & Hu, G. Codes of the global daily-scale soil moisture fusion dataset based on Triple Collocation Analysis (2011–2018). *National Tibetan Plateau Data Center* https://doi.org/10.11888/RemoteSen.tpdc.271988 (2022).

Author contributions

Global daily-scale soil moisture fusion dataset (GDSMFD) was produced and completed through collaboration among all the authors. Qiuxia Xie prepared and processed the SMOS, FY3-B, ASCAT, ESA-CCI, SMAP surface soil moisture and *in-situ* measurements datasets, did the main analysis, wrote this manuscript and produced fusion dataset. Guangcheng Hu revised and edited this manuscript. Li Jia and Massimo Menenti supervised this study and revised and edited this manuscript, and provided many key comments. Li Jia and Massimo Menenti polished the language of this manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-022-01772-x.

Correspondence and requests for materials should be addressed to L.J.