https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

Open Access Earth System Science Data Discussions

A new dataset of satellite observation-based global surface soil moisture covering 2003~2018

Yongzhe Chen[1,2], Xiaoming Feng[1,*], Bojie Fu[1,2]

[1] State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, PR China.

[2] University of Chinese Academy of Sciences, Beijing 100049, PR China.

Correspondence to: Xiaoming Feng(fengxm@rcees.ac.cn)

Abstract. Soil moisture is an important variable linking the atmosphere and the terrestrial ecosystems. However, long-term satellite monitoring of surface soil moisture is still lacking at global scale. In this study, we conducted data calibration and fusion of 11 well-acknowledged microwave remote sensing-based soil moisture products since 2003 through neural network approach, with SMAP soil moisture data applied as the fundamental training target. The training efficiency proves to be high (R^2 =0.95) due to the selection of 9 quality impact factors of microwave soil moisture products and the elaborate organization structure of multiple various neural networks(5 rounds of simulation; 8 substeps; 74 independent neural networks; and >10^6 regional subnetworks). We achieved global satellite monitoring of surface soil moisture during 2003~2018 at 0.1° resolution. This new dataset, once validated against the International Soil Moisture Network (ISMN) records, is supposed to be superior to the existing products (ASCAT-SWI, GLDAS Noah, ERA5-Land, CCI/ECV and GLEAM), and is applicable to studying both the spatial and temporal patterns. It suggests an increase in global mean surface soil moisture, and reveals that the surface moisture decline on rainless days is highest in summers over the low-latitudes but highest in winters over most mid-latitude areas. Notably, the error propagation with the extension of the simulation period to the past is well controlled, indicating that the fusion algorithm will be more meaningful in future when more advanced sensors are in operation. The dataset can be accessed at https://doi.pangaea.de/10.1594/PANGAEA.912597 (Chen, 2020).

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

1 Introduction

Soil moisture plays an important role in modulating the exchanges of water, carbon and energy between the land surface and atmosphere, as well as linking the global water, carbon and energy cycles (Dorigo et al., 2012; Karthikeyan et al., 2017a), which has been endorsed by the Global Climate Observing System (GCOS) as an essential climate variable (Bojinski et al., 2014). It is also considered as the best indicator of ecological droughts (Martínez-Fernández et al., 2016; Samaniego et al., 2018) and is thus a direct driving factor of evapotranspiration, carbon uptake and many other ecological processes, yet surrogates such as precipitation and vapor pressure deficit (VPD) are often applied in modeling due to the large uncertainty in global-scale soil moisture data (Hashimoto et al., 2015; Stocker et al., 2019).

The reanalysis land surface model products (e.g., the Global Land Data Assimilation System (GLDAS, with spatial resolution of 0.25°) (Rodell et al., 2004), ECMWF ERA-interim (0.25°) (Balsamo et al., 2015) and its newly-published successors: ERA5 (0.25°) and the land product, ERA5-Land (0.1°)(Hoffmann et al., 2019)) are the most frequently used. Though these products can often predict temporal variations well due to the incorporation of high-quality precipitation data, the bias and root mean square error (RMSE) may be large (Bi et al., 2016; Gu et al., 2019). Moreover, the significant impacts of human activities such as irrigation and land cover changes on soil moisture can hardly be considered (Kumar et al., 2015; Qiu et al., 2016).

With the advance of remote sensing technology, soil moisture products derived from microwave remote sensing, which prove to be superior to those derived from other electromagnetic wave bands (Karthikeyan et al., 2017a), have become an alternative of surface soil moisture (the penetrability of microwaves is usually <5 cm of soil) monitoring (Feng et al., 2017; Jiao et al., 2016). Currently, global-scale soil moisture can be acquired from either passive (e.g., SMMR, SSM/I, TMI, WindSAT, AMSRE, AMSR2, SMOS, SMAP) or active sensors (e.g., ERS and ASCAT) but the valid temporal spans of all these sensors are limited, and the data quality and spatial coverage were unsatisfactory till the launch of AMSRE in June 2002 (Karthikeyan et al., 2017b; Kawanishi et al., 2003). Currently, the longest continuous record of global soil moisture retrieved frommicrowave remote sensing only is the ASCAT product (Bartalis et al., 2007), whose temporal span was from 2007 till now. Apart from the relatively short time series, the accuracy of satellite-based soil moisture is usually lower than

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

that of model products (Albergel et al., 2012; Chen et al., 2013) due to various disturbances. Moreover, large discrepancies exist among the soil moisture retrieved from various sensors, and that by using different algorithms (Kim et al., 2015; Mladenova et al., 2014). Although new sensors, SMOS (Stillman and Zeng, 2018) and SMAP (Entekhabi et al., 2010), can produce significantly improved estimates because L-band microwaves (1~2 GHz) have better penetrability (Burgin et al., 2017; Chen et al., 2018; Karthikeyan et al., 2017b; Kerr et al., 2016; Kim et al., 2018; Leroux et al., 2014; Stillman and Zeng, 2018), the applicability of both products is still limited. SMOS data have too much noise and too many missing values in Asia due to high radio frequency interference (RFI) (Oliva et al., 2012). While SMAP has the highest quality (the unbiased RMSE of the passive product can be close to its target of 0.04 m^3/m^3) and has filtered RFI (Chen et al., 2018; Colliander et al., 2017), the data are only available since March 2015.

Because both simulated and satellite-observed soil moisture have advantages and disadvantage, there are increasing interests in date integration. The European Space Agency (ESA) published a long-term surface soil moisture dataset called the Climate Change Initiative (CCI) or Essential Climate Variable (ECV), and v4.4 covers 1978~2018. Two aspects contribute to the combined product. The first involves rescaling all microwave sensors' retrievals against the reference data (GLDAS Noah product) by cumulative distribution function (CDF) matching. The second step is to merge the rescaled products together by selecting the best product in each subperiod or averaging the products weighted by the estimated errors (Dorigo et al., 2017; Gruber et al., 2017; Gruber et al., 2019; Liu et al., 2012). Although CCI covers more than 40 years, the data before June 2002 have many missing values and are with lower quality. Upon rescaling, the spatial patterns of the satellite products are almost replaced by those of GLDAS. Although temporal patterns are retained, the merging algorithm is too simple (Liu et al., 2012) to harmonize the discrepancy among the temporal variations in various products (Feng et al., 2017). Because the temporal variation in soil moisture is often better captured by model simulations than remote sensing inversions, CCI may undesirably combine the disadvantages of both. Another popular data product, GLEAM (Global Land Evaporation Amsterdam Model) soil moisture, is produced by the assimilation of the CCI data (Burgin et al., 2017; Martens et al., 2017; Miralles et al., 2011). Currently, anomalies of CCI (the deviationsto the seasonal climatology that indicate whether the soil moisture at a time point is more humid or drier than the multi-year average (Lievens et al., 2017)) are assimilated instead.

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020


Open Access Earth System Science Data Discussions

Therefore, the observed spatial information is ignored, while the temporal changes are mainly driven by model simulation as well, meaning that remote sensing data are not optimally utilized.

Hence, the probably best practice for global long-term soil moisture mapping is to first develop a long-term surface soil moisture dataset only using the satellite data and then assimilate this dataset into model simulations. The first step is the major task. Rescaling using CDF matching followed by data averaging, which is adopted in CCI, can hardly unify the temporal variations. Targeting at using the information acquired by one sensor to produce soil moisture compatible with that retrieved from another, three methods have been proposed. Based on physical-based equations (Wigneron et al., 2004), the regression between SMOS soil moisture and the dual-polarized Tb (brightness temperature) data of AMSRE is used to calibrate the AMSRE soil moisture time series (R-square, R^2=0.36) (Al-Yaari et al., 2016). An example of the second method is to use LPRM (Land Parameter Retrieval Model (Owe et al., 2008)) to retrieve soil moisture from SMOS and then match it with the AMSRE-LPRM product by calibrating the LPRM parameters and applying a linear regression (van der Schalie et al., 2017). Because machine learning can better characterize the nonlinear relationship between surface soil moisture and Tb (Rodriguez-Fernandez et al., 2015), researchers built a neural network linking SMOS soil moisture with the Tb and polarized reflectivity of AMSRE to produce a calibrated soil moisture data covering 9 years (2003~2011) (Rodríguez-Fernández et al., 2016). Among these three approaches, machine learning proves to be probably the best choice (Van der Schalie et al., 2018).

Global long-term observational-based soil moisture has been developed recently by building a simple neural network between the SMOS product and the Tb of AMSRE (2003~September 2011) and AMSR2 (July 2012~2015) (Yao et al., 2017), with environmental factors including land surface temperature (LST) derived from TB(V) at 36.5 GHz (Holmes et al., 2009) and microwave vegetation index (MVI, an indicator of vegetation amount) incorporated. The training R^2 is only 0.45 (R=0.67), and the validation against measurements shows a temporal R of 0.52 (temporal RMSE=0.084). In addition, the gap between the temporal spans of AMSRE and AMSR2 and the lack of SMOS data in Asia resulted in large quantities of missing data. As SMAP observations become more and more available, it has been chosen as the training target, improving the training $R^2$to 0.55, while the overall R and RMSE against measurements are 0.44 and 0.113(Yao et al., 2019) . Another

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

study rebuilt a soil moisture time series over the Tibetan Plateau by using SMAP data as the reference of a random forest (Qu et al., 2019). For environmental factors, while vegetation cover is not considered, elevation (DEM), land use type, MODIS LST, precipitation and the day of a year (DOY) are chosen as ancillary predicators. The training R^2in this region reached 0.9, with a temporal accuracy higher than other products (R=0.7; RMSE=0.07 in the unfrozen season). However, this data is regional, with a large temporal gap, and cannot be seen as observational-based only since precipitation data is incorporated.

To be concluded, while recent studies have focused on developing long-term satellite-based surface moisture using machine learning, substantial success has not been achieved yet. The major concerns are as follows: 1) the high-quality microwave observations are not fully utilized, leading to too large temporal and spatial gaps, and limited training efficiency; 2) it remains unclear whatenvironmental factors should or should not be incorporated, and why; and 3) the training designed for soil moisture estimation at global scale ought to be more complex than that for only a specific region to ensure a satisfying training efficiency. By taking these matters into account, this study more efficiently calibrated and fused the well-known microwave soil moisture products since 2003 to developa global long-term soil moisture dataset with higher accuracy.

2 Data and methods

2.1 The satellite-based surface soil moisture data products

SMAP is currently the best of all remote sensing-based soil moisture products and is thus chosen as the primary training target. The SMAP Enhanced L3 Radiometer Global Daily 9 km EASE-Grid Soil Moisture V002 (SPL3SMP_E_002, which is referred to as SMAP_E for short), which is developed by an improved spatial interpolation of the original 36 km resolution SMAP soil moisture data(Chan et al., 2018), was adopted in this study. The nominal depth of SMAP_E is ~5 cm.

Previous studies used Tb observations at various bands as network inputs (Rodríguez-Fernández et al., 2016). However, because the primary goal of this study is to calibrate and then fuse the existing popular microwave soil moisture products, while the Tb observations at multiple bands contain too much information that is not related to soil moisture, which will weaken the training efficiency and probably lead to over-fitting, the well-acknowledged surface soil moisture products

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

retrieved through mature algorithms(Table 1) are directly applied here, instead of Tb. Though the drawback is that the final soil moisture products may inherit the uncertainties associated with each retrieval method, this problem can be generally solved by including quality impact factors (see section 2.2 for details).

To reduce noises and fill the gaps between satellite tracks, for every soil moisture product, both the daytime and nighttime observations within each 10-day period (20 observations) are combined by data averaging (the diurnal variation of soil moisture is ignored in this study, while the relative superiority of daytime and nighttime retrievals is supposed uncertain). The ASCAT soil moisture index product (ASCAT-SWI) was developed by the ESA (Albergel et al., 2008; Wagner et al., 1999). The saturation degree of the top soil layer (SWI_001) was converted to volumetric soil moisture by multiplication with soil porosity data included in the SMAP L4 Global Surface and Root Zone Soil Moisture Land Model Constants V004 dataset (hereinafter the SMAP Constant). AMSR2-JAXA is the AMSR2 soil moisture retrieved by the Japan Aerospace Exploration Agency (JAXA) using Tb at X-band (10.65 GHz) (Fujii et al., 2009), and version 3 data on the Global Portal System (G-Portal) were used. AMSR2-LPRM-X stands for the AMSR2 soil moisture produced by applying the LPRM algorithm at the X-band (Parinussa et al., 2014) (C-band data such as AMSR2-LPRM-C or AMSRE-LPRM-C, was not incorporated due to high RFI (Njoku et al., 2005)), and is obtained from NASA's Earthdata Search web. SMOS-IC is a new SMOS soil moisture product made by INRA and CESBIO with the main goal of being as independent as possible from auxiliary data, including the simulated soil moisture (Fernandez-Moran et al., 2017a; Fernandez-Moran et al., 2017b; Wigneron et al., 2007). The accuracy of SMOS-IC proves to be higher than other SMOS products (Al-Yaari et al., 2019; Ma et al., 2019), and the data version 105 offered by Centre Aval de Traitement des Données SMOS (CATDS) is adopted. TMI-LPRM-X is the X-band LPRM product of TMI and was created by the NASA Goddard Space Flight Center (GSFC). Fengyun 3B is a Chinese meteorological satellite with a Microwave RadiationImager (MWRI) onboard (Yang et al., 2011; Yang et al., 2012). The National Satellite Meteorological Center product is retrieved using Tb at 10.7 GHz, which is denoted by 'FY-3B-NSMC'. WindSat is flown on the Coriolis satellite (Gaiser et al., 2004), the soil moisture retrieved by LPRM at X-band (Parinussa et al., 2012) is provided by NASA. Three AMSRE products are used, including the NASA product (AE_Land3) created by the National Snow and Ice Data Center (AMSRE-NSIDC)(Njoku et al., 2003), the JAXA product

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

(AMSRE-JAXA) (Fujii et al., 2009; Koike et al., 2004) obtained from G-Portal and the LPRM product (AMSRE-LPRM) available at the NASA Earthdata Search. All data are re-projected to the WGS-1984 coordinate system and resampled to 0.1°.

2.2 The quality impact factors of soil moisture retrievals

Environmental factors, including DEM, precipitation, LST and vegetation (NDVI, MVI), were used as neural network inputs for improved soil moisture estimation and data fusion (Lu et al., 2015; Qu et al., 2019; Yao et al., 2017). However, the mechanism why these factors are helpful remains controversial: are these factors used as direct spatial predictors of soil moisture or just because they are related to the errors of satellite soil moisture retrievals (i.e., the quality impact factors of soil moisture)? We insist on the latter,proposing two main reasons for the incorporation of environmental factors: 1) the bias of soil moisture estimates derived from a certain sensor or a specific algorithm can be correlated with the degree of disturbances from various environmental factors. For example, in vegetated areas, LST are overestimated by LPRM (Ma et al., 2019), whereas soil moisture is underestimated by JAXA (Kim et al., 2015), and the magnitudes of biasesare determined by vegetation amount or VOD. It indicates that the environmental factors are essential for a better calibration of various products, especially when soil moisture, which contains errors associated with the retrieval method, is directly applied instead of Tb; and 2) the relative performances of different products are also controlled by environmental factors; for example, the ASCAT product is preferable to AMSRE-LPRM in vegetated areas (Dorigo et al., 2010), while LST influences the relative superiority of the LPRM and JAXA algorithms (Kim et al., 2015). Therefore, for improved data fusion, the weights assigned to different predictor soil moisture (or Tb) data available at the same time should be determined by referring to these factors.

Based on the two criteria above, we determinedwhich environmental factors should be included. The first is the ‘vegetation factor’ (i.e., vegetation water content, VWC). Plants can absorb or scatter the radiation from soil and emit radiation, reducing the sensitivities of both radiometer and radar to soil moisture (Du et al., 2000; Owe et al., 2001). However, L-band microwaves can penetrate the vegetation layer better due to its longer wavelengths (Konings et al., 2017). On the other hand,

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

though vegetation effects can be somewhat corrected (Jackson and Schmugge, 1991), different methods have different efficiencies. First-order radiativetransfer models such as LPRM can hardly describe the radiationattenuation by dense canopy (Crow et al., 2010), but the TU-Wien change detection algorithm applied to ASCAT can reduce vegetation impacts due to the implicitaccount of higher-order scattering effects (Bartalis et al., 2007). Microwave vegetation indexes may contain large uncertainty and are in coarse resolution (Liu et al., 2011; Shi et al., 2008). NDVI becomes saturated at high vegetation cover (Huete et al., 2002). Because the leaf area index (LAI) stands for the total leaf area per unit land, which is closely related to VWC assuming a relatively stable leaf equivalent water thickness (Yilmaz et al., 2008), LAI is a suitable surrogate. The Copernicus global 1 km resolution LAI (GEOV2-LAI) data are adopted here due to high accuracy and full coverage (Baret et al., 2013; Camacho et al., 2013; Verger et al., 2014). Because the sensor conversion from SPOT-VGT to PROBA-V in 2014 led to an LAI data discontinuity in specific areas (Cammalleri et al., 2019), which may reduce neural network training and simulation efficiency, the GLASS (Global LAnd Surface Satellite) LAI product (Xiao et al., 2014; Xiao et al., 2016) from 2007~2017 is also used. The LAIs are averaged on a monthly scale and aggregated to 0.1° resolution. The second is the 'water factor' (i.e., water body fraction). Water bodies dramatically lower the Tb, leading to overestimation of soil moisture. Because there are different methods of detection and correction of water bodies (Entekhabi et al., 2010; Kerr et al., 2001; Mladenova et al., 2014; Njoku et al., 2003), soil moisture calibration and weight assignment based on the water body fraction make sense. In this study, daily water fraction derived from the Surface WAter Microwave Product Series (SWAMPS) v3.2 dataset (Schroeder et al., 2015) was applied. The third is the 'heat factor' (i.e., LST). Soil moisture retrieval from passive microwave sensors is based on the correlation between soil dielectric conductivity, that is influenced by soil moisture, and the emissivity estimated as the ratio of Tb to soil physical temperature (Ts) (Karthikeyan et al., 2017a). Ts is approximate to LST and can be derived from the Tb at 36.5 GHz (Holmes et al., 2009; Parinussa et al., 2011), or from reanalysis datasets including ECMWF, MERRA and NCEP, or set as a constant of 293 K (Koike, 2013). Active microwave products are independent of LST (Ulaby et al., 1978). Because different LST estimates have different biases compared to the actual values, the actual LST may determine the bias of each LST product, and finally the relative performances of various soil moisture products (Kim et al., 2015). In this study, we averaged the MODIS monthly LST

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

acquired from the ascending and descending passes of both TERRA and AQUA. The 4th is the 'land cover factor', which is added because the parameters essential for soil moisture retrieval (vegetation effect correction) are set based on land use types (Griend and Wigneron, 2004; Jackson and Schmugge, 1991; Jackson et al., 1982; Panciera et al., 2009). Additionally, landscape heterogeneity influences the retrieval accuracy (Lakhankar et al., 2009; Lei et al., 2018; Ma et al., 2019). Here, both the annual MODIS land use cover maps and MEaSUREs vegetation continuous fields (i.e. the cover fractions of trees, non-tree vegetation and bare ground (Hansen and Song, 2018)) are adopted. Apart from the above dynamic factors, there are also two static factors: the 'topographic factor' (i.e., topographic complexity or surface roughness) and the 'soil factor' (i.e., soil texture indicated by sand and clay fractions)(Neill et al., 2011). Both can influence the relationship between soil moisture and emissivity or dielectric constant (Dobson et al., 1985; Karthikeyan et al., 2017a; Njoku and Chan, 2006), but are characterized and corrected differently, leading to different relative performances of soil moisture products (Das and O'neill, 2010; Gao et al., 2006; Kim et al., 2015). For topographic complexity, the static layer of the Copernicus ASCAT-SWI product (hereinafter the ASCAT Constant) is adopted while for soil texture, the SMAP Constant is used. Precipitation and DOY are not included as ancillary inputs (see Text S1 for explanations). Therefore, there are 9 quality impact factors: LAI, water body fraction, LST, land use cover, tree cover fraction, non-tree vegetation fraction, topographic complexity, and sand and clay fractions, which are essential inputs of all neural networks in this study.

2.3 The basic flow of soil moisture calculation

The global long-term surface soil moisture data production includes three basic parts, which are as follows. 1) Preprocessing: the production of high-quality neural network inputs; 2) neural network operation: the network training and soil moisture simulation and 3) postprocessing: the correction of potential errors or deficiencies in the soil moisture simulation outputs. It should be noted that because the temporal coverage of SMAP does not overlap with that of TMI, FY-3B, WindSat or AMSRE, to make full use of the satellite-based soil moisture data, several rounds of simulations are performed. Hence, the simulated soil moisture will also be converted to part of the training target of the next round's neural network, meaning that some postprocessing steps are also preprocessing steps.

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

2.4 Neural network design (1): localized neural networks

In this study, instead of a universal network, we devised localized neural networks. The data within each individual zone are used to train a zonal neural network (hereinafter a subnetwork), which is used for soil moisture simulation at that zone. By comparison, localized neural networks help improve the training efficiency; yet, a smaller zonal size does not indicate a better simulation accuracy. We noticed that the LPRM algorithm-based products (AMSR2/TMI/WindSat/AMSRE-LPRM-X) are patchy, with clear boundaries between adjacent square-shaped zones over arid places, while the patch size is exactly 1°×1°, which is probably due to the spatial distribution of parameters. This finding indicates that subnetworks should be built at 1°×1° scale. Therefore, we divided the global extent except the polar areas (80°N~60°S) into 140×360 zones. For each subnetwork, if the number of valid pixels (for a 0.1° pixel in a given 10-day period, if all the subnetwork inputs have valid values, it is seen as a valid pixel; the maximal number of valid pixels=100×the number of 10-day periods) exceeds 100, the subnetwork is considered valid. There are usually more than 15,000 valid subnetworks within an individual neural network. The training is performed in the MATLAB software, and the number of nodes in the hidden layer of each subnetwork is set to 7.

2.5 Preprocessing and postprocessing steps

After standardization of the original soil moisture data, to improve the neural network training efficiency, the potential salt and pepper noises are removed. For each map, soil moisture values are filteredto the level of three standard deviations relative to themean in each zone. This preprocessing step is thus called ‘3σ denoise’.

After neural network operation, boundary fuzzification is first applied, which is a step in both preprocessing and postprocessing. Because the localized 1°×1° network is applied instead of the global network, the boundary between nearby zones may be too obvious over some areas. To blur the boundary, a simple algorithm represented as Figure S1 is applied. The boundary fuzzified soil moisture data are both transformed into the final product and the next round’s training target. To produce the final product, two postprocessing steps are essential: filling of missing values and masking. Because ‘3σ denoise’ deleted suspicious soil moisture retrievals, the simulation outputs contain few missing values, which can be simply

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

filled by sequentially searching and averaging nearby valid values (Chen et al., 2019). While the snow/ice mask of the ASCAT-SWI product can be transferred to the simulation output, the potential snow or ice cover before 2007 should be identified. For a grid in a specific ten-day period, if ice cover is reported by ASCAT-SWI in most years, it is also supposed to be potentially covered by snow/ice. However, if the thaw state is observed according to the MEaSUREs Global Record of Daily Landscape Freeze/Thaw Status V4 dataset, the grid is not masked. The simulated soil moisture in the rainforests identified in the 'ASCAT Constant' are retained but not recommended due to large uncertainty. On the other hand, to avoid error propagation with training times by ensuring a high-quality training target for the next round's simulation, for every simulated result, we performed a preprocessing step called suspicious value removal. We obtained the maximal and minimum values of SMAP_E soil moisture in each pixel. If the simulated value is out of the range of the SMAP data during 2015~2018, the value is suspicious and thus not included in the training target. Subsequently, '3σ denoise' is performed again before the simulated soil moisture becomes secondary training targets, which are referred to as SIM-1T, SIM-2T, and so on ('SIM' stands for the simulated soil moisture, the number after the hyphen indicates the round of simulation, and 'T' means it is applied as training target).

2.6 Neural network design (2)- five rounds of simulations

A key characteristic of this study is that not only the 11 available microwave soil moisture data with different temporal spans are all incorporated, but they are also utilized as fully as possible through up to 5 rounds of neural network-based simulations, with at least four different soil moisture products retrieved from three sensors applied as predictors in each round (see details below). While increasing the sources of soil moisture data inputs can be beneficial to training efficiency, the spatial coverage of the simulation output is sacrificed because the overlapping area decreases with the increase in soil moisture data varieties. After all, most products have missing data in specific regions (e.g., mountains, wetlands and urban settlements), and some sensors are even unable to produce data at global scale (TMI is limited to [N40°, S40°]; SMOS lacks data in Asia). To solve that dilemma, we classified all pixels according to the available soil moisture products with valid data during a 10-day period. However, to avoid soil moisture simulation under snow or ice cover, not all combinations are

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

considered (see Section 2.5) while some combinations are optional. Then, corresponding to the selected combinations, different independent neural networks are trained. Moreover, for a pixel, if the neural network driven by multiple predictor soil moisture has no subnetwork in the zone where it is located due to limited valid pixels, an alternative subnetwork driven by the combination of fewer soil moisture data inputs should be applied instead. Hence, it is a question that which neural network is the best choice for a specific pixel. Apart from applicability, the relative priority of different neural networks should be determined by comprehensively considering the number and quality of input soil moisture products, the variety of sensors, the quantity of training samples indicated by the number of 10-day periods, and the accuracy of training targets (the training target quality declines monotonously: SMAP>SIM-1T>SIM-2T>SIM-3T>SIM-4T). Sometimes, the priority order of independent networks cannot be determined easily, then various orders are provided, leading to various substeps, with the respective simulation results integrated later. Specifically, when the LAI data source changes, the division of a single round is also essential. Based on these principles, five rounds of neural network operations containing a total of 74 independent neural networks could be designed as follows.

The first round of simulation has two substeps. In substep 1, GEOV2 PROBA-V LAI data are used for soil moisture simulation during 2014~2018 (note: the other 8 quality impact factors are applied as well) whereas in substep 2, the GLASS LAI are applied, and the simulation period is 2012D19~2013D36 (i.e., September 2012 to 2013; D is the ordinal of the ten-day period). SMAP soil moisture is the training target, while ASCAT-SWI10 (abbreviated as ASCAT), SMOS-IC (SMOS), AMSR2-JAXA and AMSR2-LPRM-X (AMSR2-LPRM) are the four predictor soil moisture products (details are in Table S1~S2). The simulation results of the two substeps are combined and then transformed into the training target (SIM-1T). In the second round of simulation, the training target can either be SMAP or SIM-1T, while the input soil moisture data are ASCAT, SMOS, TMI-LPRM-X (TMI) and FY-3B-NSMC (FY). The simulation output, SIM-2, spans from 2011D20 to 2012D18 (Table S3~S4). In the third round (2010D16~2011D19), SMAP, SIM-1T and SIM-2T are all used as training targets, while the predictor soil moisture data are ASCAT, SMOS, TMI and WindSat-LPRM-X (WINDSAT). There are two substeps in round 3, distinguished by whether the priority of the neural networks is determined by the training sample quantity and the training target quality first (SIM-3-1) or the number of input soil moisture products is considered

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020


Open Access Earth System Science Data Discussions

first (SIM-3-2, Table S5~S8). Because these two methods emphasize different aspects of network quality, in some pixels. SIM-3-1 will be advantageous, but in others, SIM-3-2 is better. Hence, an algorithm is proposed to combine the advantages of both simulations, which is described in Table S9. Next, the 4th round is for simulation during 2007D1~2010D15. With SIM-2T and SIM-3T being the training targets, ASCAT, WINDSAT, TMI, AMSRE-JAXA, AMSRE-LPRM-X (AMSRE-LPRM) and AMSRE-NSIDC are all applied as predictors (LAI now comes from SPOT-VGT). Two substeps are also included. In the first substep, neural networks are sorted by paying more attention to the number of soil moisture inputs and the sensors they are derived from, while the training sample size and training target quality are prioritized to make an alternative estimation (Table S10~S13). Afterwards, SIM-4 is obtained by reasonably fusing these two results. In the final round, the soil moisture simulation is extended to as early as 2003. SIM-2T, SIM-3T and SIM-4T are combined to be the training target, while the soil moisture data entering the neural networks consist of WINDSAT, TMI, AMSRE-JAXA, AMSRE-LPRM and AMSRE-NSIDC (Table S14~S15).

2.7 Validation of surface soil moisture products

For the evaluation of global soil moisture data, the International Soil Moisture Network (ISMN) dataset (Dorigo et al., 2011; Dorigo et al., 2013) is the most frequently used (Al-Yaari et al., 2019; Karthikeyan et al., 2017b). Because SMAP, the training target, is the soil moisture within 0~5 cm, the simulated soil moisture is supposed for that soil layer as well. Accordingly, the measurements used for validation are limited to ≤ 5 cm in depth. The quality flags of ISMN(Dorigo et al., 2013) are also checked to retain only the 'good quality' data. After data screening and processing (see Text S2), more than 100,000 10-day averaged soil moisture records acquired from 728 stations of 29 networks are selected for validation of the soil moisture product whose temporal resolution is 10 days. The detailed information of these stations and the periods of the data used are listed in Table S16, while the spatial distribution of these stations is shown in Figure S2. The major climate types of the sites are determined from the updated map of the Köppen-Geiger climate classification (see Table 2 for the descriptions of each type (Kottek et al., 2006)).

Most ISMN networks are dense networks, as the stations are very close to each other, probably within the same 0.1° grid,

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

whereas some others are sparse networks (Figure S2). In addition, various sensors are simultaneously operated at some stations. Hence, to make full use of all available records, the site-scale 10-day averaged soil moisture data are aggregated to 0.1° grid-scale by averaging the same period's values within the same grid. Specifically, if soil moisture is not simulated due to snow or ice cover, the corresponding measurement is excluded. This resulted in a final collection of ~40,000 grid-scale 10-day period soil moisture records within the validation dataset.

The soil moisture datasets to be evaluated include the simulated (satellite-based) soil moisture product in this study (hereinafter 'SIM', covering 2003~2018), SMAP_E (the training target, covering April 2015~2018), the longest existing satellite-derived soil moisture record: ASCAT-SWI (converted to volumetric fraction; data period is 2007~2018), the reanalysis-based soil moisture: GLDAS Noah V2.1 and ERA5-Land (data were resampled to 0.1° by bilinear interpolation, averaged to 10-day scale and then evaluated during 2003~2018) as well as the soil moisture data developed by combining both satellite observation and model simulation: CCI and GLEAM soil moisture products (note that the difference between GLEAM v3.3a and v3.3b is that for v3.3a, the radiation and air temperature forcing data comes from ERA5, whereas for v3.3b, all meteorological data are satellite-based, yet the time series is only updated to September 2018). The overall performance of any soil moisture product is first evaluated using all of the validation dataset, with R square (R^2) and RMSE values (unit: $m^3\ m^{-3}$) adopted as the main indicators. The next step is the temporal pattern validation. For grids with enough (>20) 10-day average records, we compared the estimated soil moisture during all periods against the corresponding measurements, with the correlation coefficient (R) and RMSE calculated. Several supplementary indexes are also added, including bias, unbiased RMSE (ubRMSE) and the correlation between the anomalies (anomalies R, abbreviated here as 'A.R', can better indicate the prediction accuracy of interannual changes). Then, we compared the mean/median of the above evaluation indexes for different soil moisture products and tested whether the differences are significant. Moreover, the relative performances of various products in different climate zones are analyzed. Last, we performed spatial pattern validation. For every 10-day period, the soil moisture data in all grids with stations are evaluated by the R, RMSE, bias and ubRMSE values. The relative superiority of all products during different time periods in a year, and the changes of data coverage as well as data quality with time were also investigated.

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020


Open Access Earth System Science Data Discussions

2.8 Intra-annual variation analysisof surface soil moisture

It should be noted that although the nominal spatial resolution of the surface soil moisture product (SIM) is 0.1°, the intra-annual variation analysis would be more robust if we aggregate SIM to 0.5° resolution because the original resolution of SMAP soil moisture is ~0.4° while the resolution of most soil moisture products to be calibrated and fused is 0.25°. Then, we excluded the high latitude areas (60°N~90°N) where the available data are limited due to frequent ice cover. For the remaining areas (60°S~60°N), based on a total of 36×16 (years)=576 data points, we fitted the intra-annual cycle of soil moisture by using the Fourier function (probably the best choice for periodic function fitting) with the period fixed to 1 year (36 10-day periods). The number of terms is set to 1 unless the intra-annual cycle is obviously asymmetrical and can be much better characterized by a two-term Fourier function. Subsequently, the highest peak and lowest trough values of surface soil moisture as well as the corresponding locations in time (the ordinal of 10 days) are exported.

The direct driving factor of the variation in surface soil moisture is precipitation, for which we adopted the GPM IMERG Precipitation V06- final run data (Huffman et al., 2019). Apart from a direct correlation analysis, we also used the Fourier function to characterize the intra-annual cycle of precipitation and then explored the relationship between the cycles of precipitation and surface moisture (the derived fitting function is dropped if the adjusted R^2 is lower than 0.1), with the peak time difference in each 0.5° grid calculated (if both cycles have two peaks, the average locations of the two peaks are calculated). Because the simulated surface soil moisture only represents the average condition during each 10-day period while the accuracy is not high enough, we only studied the surface soil moisture decline after ten consecutive rainless days. Effective precipitation is calculated by precipitation minus canopy interception that is estimated by combining the modified Merriam canopy interception model (Kozak et al., 2010; Merriam, 1960) (the vegetation canopy cover and LAI is acquired from the GEOV2 dataset). If the total effective precipitation within two consecutive 10-day periods (20 days) is less than a given threshold (initially set to 10 mm), we suppose that the soil moisture change in the latter period compared to the previous period is mostly due to surface evaporation and percolation (capillary rise can be negligible (McColl et al., 2017)) and thus should be negative. Hence, for a 0.5° grid, if each time when the number of negative values does not exceed two times that of the positive ones, the precipitation threshold was reduced by 1 mm until that condition is satisfied. This loop is

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

terminated when the available data points in dry periods are less than 36 (the maximal number of data points is 576), with the grid excluded from the analysis. In desert areas, the random noise of the surface soil moisture product can hide the signal of moisture changes, while in rainy areas, 20 days without effective precipitation seldom occurs, leading to no results over most regions. In the remaining places, the intra-annual variation in the surface moisture loss during dry days can be fitted by the Fourier function as well, which is then analyzed using the above methods.

3 Results

3.1 The neural network efficiency: a comparison between SIM and SMAP

To examine the training and simulation efficiency of the neural network, we compared the simulated surface soil moisture (SIM) with the training target- SMAP during April 2015~2018. The R^2 reaches up to 0.95, while the RMSE is 0.031 m^3/m^3 (Figure 2a). If only the grids with measured data are considered, the consistency between SIM and SMAP is even stronger, with an R^2 of 0.97 and an RMSE of 0.016 (Figure 2b). When validated against site measurements, for both SIM and SMAP, the R^2 and RMSE values are 0.46 and 0.083, respectively, indicating the same overall data accuracy (Figure 2c and 2d). All these findings justify the high predictive efficiency of the neural network set designed in this study.

As for temporal accuracy, according to Table 3, SIM is just slightly lower than SMAP (the differences in the five indicators, R, RMSE, bias, ubRMSE and A.R, are all nonsignificant). Figure 3 indicates the generally same level of temporal accuracy for SIM and SMAP under all climates. SIM cannot well characterize the temporal variation in soil moisture in the 'Dfc' (snow climate, fully humid, see Table 2) region because SMAP does not have a high temporal accuracy there, probably due to frequent freezing and melting processes.

Next, we compared the spatial accuracy of SIM and SMAP. The spatial correlation of SIM is somewhat reduced compared to the training target, while the RMSE is slightly increased (Table 4), indicating a subtle loss of spatial information through neural network training and simulation. Because ISMN stations mostly locate in the mid to high latitudes of the Northern Hemisphere, Figure 4 shows that: 1) the accuracy of SIM is highest in summers (growing seasons) and lowest in winters, which is the same as its origin- SMAP, probably due to the impact of freezing on soil moisture retrieval from microwave

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

sensors; and 2) SIM shares a similar spatial accuracy as SMAP in most periods, but the inferiority of SIM occurs during May to June and November to December.

3.2 The accuracy comparisonbetween SIM andpopular global long-term soil moisture products

3.2.1 The data quality comparison between SIM and the satellite-derived product

The satellite-derived global surface soil moisture product, ASCAT-SWI, now covers 12 years, 2007~2018. During that period, the overall R^2 and RMSE for SIM are 0.44 and 0.086 (Figure 5), respectively, which appears much better than those for ASCAT-SWI (R^2=0.33, RMSE=0.1).

According to the temporal validation results (Table 5), the evaluation indexes including R, RMSE, bias and ubRMSE for SIM are all significantly ($p<0.05$) better than those for ASCAT-SWI (anomalies R for SIM is also higher, but not significant). The temporal accuracy of SIM appears obviously higher in all climatic zones except for polar areas (Dsb, Dwc and ET). Specifically, in arid areas (BWh and BWk), the temporal correlation coefficients for ASCAT-SWI are low and even negative, but high for SIM (Figure 6).

The spatial accuracy of SIM is considered to be significantly higher than ASCAT-SWI as well when any one evaluation index is considered (Table 6). Moreover, it shows that SIM is somewhat superior to ASCAT-SWI all year round, especially during the growing seasons (Figure 7).

3.2.2 The data quality comparison between SIM and land surface model products

First, the overall accuracies of SIM and GLDAS Noah V2.1 surface soil moisture data during 2003~2018 are compared. The relationship between GLDAS soil moisture and the measured values is obviously nonlinear, resulting in a smaller R^2 and higher RMSE for GLDAS product compared to SIM (Figure 8).

The supposed higher temporal accuracy of SIM than GLDAS can be justified by comparing the indicators including R, RMSE and ubRMSE (Table 7). The advantage of SIM over GLDAS could be identified in almost all climatic regions, especially cold areas such as BWk, Dfa, Dfc, Dwc and ET, perhaps because the soil thawing and freezing process are not simulated well in those regions.

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

The spatial accuracy of SIM, which is indicated by R, RMSE, bias and ubRMSE, is expected to be significantly higher than GLDAS as well (Table 8). The spatial correlation of SIM is somewhat higher than that of GLDAS during March to May and September to November, and the spatial RMSE is lower during all year round except for January and February (Figure 10). ERA5-Land is a newly published reanalysis-based model product with 0.1° resolution. The overall quality validation (Figure S3) reveals a frequent overestimation of soil moisture by ERA5-Land as well as a nonlinear relationship between the predicted and measured values. Accordingly, although the R^2 for ERA5-Land is 0.41, just slightly lower than that of SIM, the RMSE is as high as 0.123. The temporal correlation indicators (R and A.R) for ERA5-Land are somewhat higher than SIM in general (Table S17), but in most cold areas (Dfa, Dwc and ET), the opposite condition occurs (Figure S4a and S4d). The temporal ubRMSE values for SIM and ERA5-Land do not differ significantly, but SIM usually performs better in relatively arid places (Figure S4c). While the relative temporal accuracy of SIM and ERA5-Land is unclear, the spatial pattern of SIM ought to be more accurate than that of ERA5-Land considering the significantly better spatial correlation, RMSE, bias and ubRMSE (Table S18). The observed advantage of SIM over ERA5-Land exists all year round, especially during the growing seasons from March to November (Figure S5).

3.2.3 The data quality comparison between SIM and the soil moisture products derived from both satellite data and model simulation

CCI is a typical surface soil moisture dataset developed by combining satellite observations and model simulations. However, validation against measurements indicates that the CCI product is not of very good quality; the R^2 is only 0.31 with an RMSE value of up to 0.095 (Figure S6), which justifies our inference that the CCI algorithm may undesirably combine the disadvantages of both methods. The temporal pattern of SIM, indicated by R and RMSE, is supposed to be significantly better than CCI (Table S19), under all climate conditions (Figure S7). Our result indicates that SIM also occupies a consistently higher spatial accuracy, especially during the growing seasons (Table S20 and Figure S8).

Next, we focus on the interannual change in data quality. According to Figure 11a~c, while the correlation coefficient for SIM does not vary significantly among different years, the RMSE and ubRMSE values in earlier periods are somewhat raised compared to those after 2012. The main reasons are as follows: 1) five rounds of simulations were performed, with

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

the output converted into the training target of the next round's neural networks. Hence, as the simulation period extends into the past, the error also propagates; and 2) the quality of microwave soil moisture data is generally lower in earlier times due to the relatively unadvanced microwave sensors with low signal-to-noise ratio (SNR). It is known that the valid information amount contained in each predictor will influence the training efficiency. Though the data quality of SIM can hardly be maintained, through various steps of data correction, the degradation degree is much slighter than CCI, which is probably because the simple CDF matching algorithm cannot efficiently calibrate the low-quality soil moisture data retrieved from earlier sensors.

Another potential problem in CCI is its limited spatial coverage before 2007, when ASCAT soil moisture is unavailable. By comparing the coverage of the 10-day scale SIM and CCI data (rainforests are excluded), it isshown that SIM covers all land surfaces except for permafrost, while the interannual variation in coverage is also negligible throughout the entire period (the intra-annual cycles of data coverages result from the changes in frozen areas), which are preferable to CCI (Figure 11d) and can thus ensure a supposed stronger applicability of SIM data product in long-term studies or modelling.

GLEAM products also contain satellite information due to the assimilation of CCI data, but model simulation plays a much more important role. By validation, the overall R^2 and RMSE values for the GLEAM v3.3a product (2003~2018) are 0.38 and 0.142 whereas those for the v3.3b product are 0.36 and 0.13, respectively. Both estimates are nonlinearly correlated with and are generally higher than measured values (Figure S9). Therefore, with an R^2 of 0.42 and an RMSE of 0.087, SIM is supposed superior to GLEAM v3.3a/b in general. The temporal and spatial accuracies of GLEAM products and SIM are compared in Tables S21~S24. The advantage of GLEAM is its ability to well characterize the temporal variations of soil moisture, with higher temporal R achieved in most climatic regions (Figure S10a and S10d). However, the main potential disadvantage is the obvious overestimation, which leads to significantly higher RMSE values than SIM in all regions and all periods (Figure S10b and Figure S11b). Moreover, the spatial patterns of GLEAM products are expected to be less convincing than SIM, considering the lower spatial correlation coefficients, especially in Spring (March to May) and Autumn (September to November) (Figure S11a). Therefore, the potential advantages of SIM may exceed those of GLEAM.

In conclusion, surface soil moisture developed mainly based on land surface models (GLEAM and ERA5-Land) have high

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

temporal accuracy, but the absolute values and spatial patterns are relatively unreliable, whereas SIM shows good performances in all aspects. Generally, this study indicates that the expected order of data applicability among various global long-term surface soil moisture products is SIM (applicable to all studies)> GLEAM (suitable for studies of temporal variation)> ERA5-Land (applicable to temporal pattern studies)> GLDAS Noah V2.1 (somewhat applicable to all studies)> ASCAT-SWI> CCI. The training R^2 of the previous neural networks designed for global surface soil moisture mapping is 0.45~0.55, while the temporal R and RMSE values against measurements are 0.52 and 0.084 (Yao et al., 2017), and the overall R and RMSE are 0.44 and 0.113 (Yao et al., 2019). In this study, by elaborating the neural network, the training R^2 is elevated to 0.95, with the temporal R and RMSE (0.69 and 0.08) as well as overall R and RMSE (0.65 and 0.087) values also improved. In addition, our 10-day period average product is both spatially and temporally continuous during 16 years, with a high spatial resolution, and it covers all land except for frozen ground (data are available in rainforests, but not recommended). Hence, our product is probably more useful than previous machine-learned products.

3.3 The spatial and temporal patterns of the calculated surface soil moisture

For the calculated global surface soil moisture, the spatial pattern averaged during 2003~2018 is shown in Figure 12a (the maps for separate months are shown in Figure S12a). Validation shows that apart from our calculations, GLDAS has the highest spatial accuracy, so the spatial map of GLDAS surface moisture is attached below (Figure S12b). By comparison, the spatial patterns of SIM and GLDAS are similar, but there also exist some differences (see the regions circled in red). Obviously, SIM has a higher spatial heterogeneity, probably more reflections on wetlands and irrigated fields (e.g., the Hetao Irrigation Area in China), whereas GLDAS appears patchy in arid places. The latitudinal pattern comparison in Figure S13a also implies that SIM contains more detailed spatial information.

As for the interannual variation, because the GLEAM v3.3a product proves to have the best accuracy in characterizing the temporal anomalies of soil moisture, which also covers the globe, this product is selected as the reference to justify our calculation. According to Figure S13b, both GLEAM and SIM support a significant rising trend in global mean surface soil moisture during 2003~2018, while the average rates are both ~0.03 $m^3\ m^{-3}\ yr^{-1}$ (Figure S13b). The spatial patterns of the

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

interannual trends in SIM and GLEAM are shown in Figure 12c~d, which are generally consistent. Soil moisture gains are both found over the border area between USA and Canada, as well as over Paraguay, Kazakhstan, Northeastern and Southern China (the regions circled in blue), while soil moisture declines took place in North Asia and eastern Brazil (the regions with red circles). The main discrepancy between the soil moisture trends predicted by the two products lie in Central Africa, the Arabian Peninsula and northwestern Canada.

Because validation against measurements proves that the intra-annual soil moisture variation in the 'Dfc' climate region cannot be captured by SMAP or SIM, the acquired intra-annual analysis results in this region are not considered. Over low-latitude areas (30°S~30°N), surface soil moisture peaks in summers (seasons are opposite in the Northern and Southern Hemispheres); however, in most of the mid-latitude areas (30°S~60°S; 30°N~60°N) except for eastern Asia (i.e., east of the Yenisei River), the soil moisture is high in winters (non-growing seasons) and low in summers (Figure 13a and Figure S13a). The intra-annual range of surface moisture is largest in the tropical monsoon climate regions, including the African savannas, the Orinoco Plain, the Ganges plain and the plains in the Indochina Peninsula, as well as some seasonal frozen areas, whereas lowest in arid places (Figure 13b; Figure S14b~c). Precipitation is a direct driver of surface soil moisture changes (Figure S15a~b), and the intra-annual cycle of soil moisture often strictly follows that of precipitation as long as it exists (Figure 13c and Figure S15c). Therefore, considering that precipitation is highest in summer in the low latitudes, where plants often grow in all seasons, whereas in the westerlies, rainfall is temporally even (eastern Asia is an exception perhaps due to monsoon and topographic conditions) yet with much higher evapotranspiration in summer, the global intra-annual patterns of soil moisture can be explained. The peak time difference between surface moisture and precipitation is approximately one 10-day period, or six days on average at global scale (Figure 13d), which is expected to be related to the time lag effect. On dry days, one may expect the fastest surface moisture decline in summers when evapotranspiration is high. However, this study reveals that in midlatitudes, the opposite condition occurs: the surface water loss without rain is lowest in summer (Figure 13e and Figure S16a). Further analysis proves a positive correlation between surface moisture and its rate of decline, with $R>0.8$ over 85% of places (Figure S16b~c), indicating that because soil moisture in the westerlies is often higher in winters, the available surface water for evaporation and percolation loss is limited in summer, and plants tend

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

to utilize water in deeper soil layers. When droughts occurred during a random period, the mean surface moisture decline is highest in the tropical monsoon climate regions (Figure 13f). Therefore, if it lacks enough water during rainy seasons there, the significant water loss (Figure S16d) may ruin the local ecosystem.

4 Discussion and conclusions

In this study, a new global long-term satellite-based surface soil moisture dataset was developed by fusing 11 microwave soil moisture products. Through validation against measurements at ISMN stations, our product proved to be superior to all the existing popular datasets. However, the achieved accuracy (R^2=0.42; RMSE=0.087) is still unsatisfactory because the acquired R^2 for soil moisture is lower than that for many other terrestrial essential climate variables. The target accuracy (RMSE) for surface soil moisture set by GCOS is 0.04 m^3 m^{-3}, which is much lower than what we have met, indicating the need to further improve the global soil moisture data quality.

Fortunately, this study provided a novel approach that has the potential to create increasingly better soil moisture products in future. Due to the elaborate design of the neural network set (localized networks, multiple rounds of simulation, various sources of soil moisture inputs, the determination of quality impact factors and the organization of multiple independent neural networks), high training efficiency is achieved, resulting in little amplification of noises and the maintenance of valid information during up to 16 years of simulation (the overall data accuracy is only slightly lower than SMAP, which is the training target). Therefore, if microwave sensors with higher SNR than SMAP are launched in future, by using the soil moisture retrieved from the new sensors as the reference, we can develop a temporally continuous soil moisture dataset since 2003, which is expected to have even higher accuracy than the SMAP product. In that sense, the data fusion algorithm proposed here will be very meaningful in future.

Another way to improve global surface soil moisture data accuracy is to combine satellite-based products with land surface models such as GLEAM. Remote sensing inversion can delineate more detailed spatial information on soil moisture, whereas the reanalysis-based models have advantages in characterizing temporal variations, except for irrigated croplands. Furthermore, because root-zone soil moisture is the direct influential factor of vegetation growth, it often plays a more

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

important role than surface moisture in ecosystems, which cannot be obtained from microwave remote sensing. Hence, combining the advantages of observation and model simulation helps to improve the data accuracies of both surface and root-zone soil moisture. Unfortunately, while the CCI algorithm integrates the disadvantages of both methods, GLEAM only incorporated very limited observed information. We propose that one possible approach is to use the grid-specific confidence range and the spatial pattern of satellite-based soil moisture (e.g., our product: SIM) to constrain the model parameters or add supplementary modules if necessary. In detail, SIM can be used as the initial surface soil moisture map. Then, after each time of soil moisture simulation in multiple layers (both root-zone and surface), the model efficiency is examined through a spatial correlation test between the simulated surface moisture and SIM. In addition, whether the simulated value falls within the confidence range (e.g., ±20%) of that reported by SIM should also be tested. By recurrent adjustments, the model parameters in each grid can be optimized. For irrigated croplands, if irrigation is not considered in models, the simulated surface soil moisture will soon fall below the confidence range, and the spatial correlation will also decline no matter which parameters are given. Therefore, a well-designed irrigation module (Chen et al., 2019) should be introduced over there. Finally, for regions with massive human-induced land cover changes (e.g., afforestation), optical remote sensing should be applied for better estimation of evapotranspiration.

5 Data availability

The global surface soil moisture dataset is available on: https://doi.pangaea.de/10.1594/PANGAEA.912597 (Chen, 2020).

Author contributions

Yongzhe Chen conducted the research, completed the original draft and revised it. The correspondence author, Xiaoming Feng, supervised the research and revised the draft. Bojie Fu administrated the project and funded the research. All co-authors reviewed the manuscript and contributed to the writing process.

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

Competing interests

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

Acknowledgements

This work was supported by the National Key Research and Development Program of China (NO. 2017YFA0604700), the National Science Foundation of China (41722104) and the Chinese Academy of Sciences (QYZDY-SSW-DQC025). We are also grateful to all the data contributors who made it possible to complete this work.

References

Al-Yaari, A., Wigneron, J. P., Dorigo, W., Colliander, A., Pellarin, T., Hahn, S., Mialon, A., Richaume, P., Fernandez-Moran, R., Fan, L., Kerr, Y. H., and De Lannoy, G.: Assessment and inter-comparison of recently developed/reprocessed microwave satellite soil moisture products using ISMN ground-based measurements, Remote Sensing of Environment, 224, 289-303, 2019.

Al-Yaari, A., Wigneron, J. P., Kerr, Y., de Jeu, R., Rodriguez-Fernandez, N., van der Schalie, R., Al Bitar, A., Mialon, A., Richaume, P., Dolman, A., and Ducharne, A.: Testing regression equations to derive long-term global soil moisture datasets from passive microwave observations, Remote Sensing of Environment, 180, 453-464, 2016.

Albergel, C., de Rosnay, P., Gruhier, C., Muñoz-Sabater, J., Hasenauer, S., Isaksen, L., Kerr, Y., and Wagner, W.: Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations, Remote Sensing of Environment, 118, 215-226, 2012.

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J. C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Piguet, B., and Martin, E.: From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations, Hydrology and Earth System Sciences, 12, 1323-1337, 2008.

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Muñoz-Sabater, J., Pappenberger, F., de Rosnay, P., Stockdale, T., and Vitart, F.: ERA-Interim/Land: a global land surface reanalysis data set, Hydrology and Earth System Sciences, 19, 389-407, 2015.

Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmara, H., Pacholcyzk, P., and Smets, B.: GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part1: Principles of development and production, Remote Sensing of Environment, 137, 299-309, 2013.

Bartalis, Z., Wagner, W., Naeimi, V., Hasenauer, S., Scipal, K., Bonekamp, H., Figa, J., and Anderson, C.: Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT), Geophysical Research Letters, 34, 2007.

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

565 Bi, H., Ma, J., Zheng, W., and Zeng, J.: Comparison of soil moisture in GLDAS model simulations and in situ observations over the Tibetan Plateau, Journal of Geophysical Research: Atmospheres, 121, 2658-2678, 2016.

Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M.: The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy, Bulletin of the American Meteorological Society, 95, 1431-1443, 2014.

Burgin, M., Colliander, A., Njoku, E., Chan, S., Cabot, F., Kerr, H. Y., Bindlish, R., Jackson, T., Entekhabi, D., and Yueh, S.: A
570 Comparative Study of the SMAP Passive Soil Moisture Product With Existing Satellite-Based Soil Moisture Products, IEEE Transactions on Geoscience and Remote Sensing, 55, 2959-2971, 2017.

Camacho, F., Cernicharo, J., Lacaze, R., Baret, F., and Weiss, M.: GEOV1: LAI, FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part 2: Validation and intercomparison with reference products, Remote Sensing of Environment, 137, 310-329, 2013.

575 Cammalleri, C., Verger, A., Lacaze, R., and Vogt, J. V.: Harmonization of GEOV2 fAPAR time series through MODIS data for global drought monitoring, International Journal of Applied Earth Observation and Geoinformation, 80, 1-12, 2019.

Chan, S. K., Bindlish, R., O'Neill, P., Jackson, T., Njoku, E., Dunbar, S., Chaubell, J., Piepmeier, J., Yueh, S., Entekhabi, D., Colliander, A., Chen, F., Cosh, M. H., Caldwell, T., Walker, J., Berg, A., McNairn, H., Thibeault, M., Martínez-Fernández, J., Uldall, F., Seyfried, M., Bosch, D., Starks, P., Holifield Collins, C., Prueger, J., van der Velde, R., Asanuma, J., Palecki, M., Small, E. E., Zreda, M., Calvet, J.,
580 Crow, W. T., and Kerr, Y.: Development and assessment of the SMAP enhanced passive soil moisture product, Remote Sensing of Environment, 204, 931-941, 2018.

Chen, F., Crow, W. T., Bindlish, R., Colliander, A., Burgin, M. S., Asanuma, J., and Aida, K.: Global-scale evaluation of SMAP, SMOS and ASCAT soil moisture products using triple collocation, Remote Sensing of Environment, 214, 1-13, 2018.

Chen, Y.: A new dataset of satellite observation-based global surface soil moisture covering 2003-2018, doi:
585 https://doi.org/10.1594/PANGAEA.912597, 2020.

Chen, Y., Feng, X., Fu, B., Shi, W., Yin, L., and Lv, Y.: Recent Global Cropland Water Consumption Constrained by Observations, Water Resources Research, 55, 3708-3738, 2019.

Chen, Y., Yang, K., Qin, J., Zhao, L., Tang, W., and Han, M.: Evaluation of AMSR-E retrievals and GLDAS simulations against observations of a soil moisture network on the central Tibetan Plateau, Journal of Geophysical Research: Atmospheres, 118, 4466-4475,
590 2013.

Colliander, A., Jackson, T. J., Bindlish, R., Chan, S., Das, N., Kim, S. B., Cosh, M. H., Dunbar, R. S., Dang, L., Pashaian, L., Asanuma, J., Aida, K., Berg, A., Rowlandson, T., Bosch, D., Caldwell, T., Caylor, K., Goodrich, D., al Jassar, H., Lopez-Baeza, E., Martínez-Fernández, J., González-Zamora, A., Livingston, S., McNairn, H., Pacheco, A., Moghaddam, M., Montzka, C., Notarnicola, C., Niedrist, G., Pellarin, T., Prueger, J., Pulliainen, J., Rautiainen, K., Ramos, J., Seyfried, M., Starks, P., Su, Z., Zeng, Y., van der Velde, R., Thibeault, M., Dorigo,
595 W., Vreugdenhil, M., Walker, J. P., Wu, X., Monerris, A., O'Neill, P. E., Entekhabi, D., Njoku, E. G., and Yueh, S.: Validation of SMAP surface soil moisture products with core validation sites, Remote Sensing of Environment, 191, 215-231, 2017.

Crow, W. T., Wagner, W., and Naeimi, V.: The Impact of Radar Incidence Angle on Soil-Moisture-Retrieval Skill, IEEE Geoscience and Remote Sensing Letters, 7, 501-505, 2010.

Das, N. and O'neill, P.: Selection of Soil Attributes Datasets for the SMAP Mission, 2010.

600 Dobson, M. C., Ulaby, F. T., Hallikainen, M. T., and El-rayes, M. A.: Microwave Dielectric Behavior of Wet Soil-Part II: Dielectric

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

Mixing Models, IEEE Transactions on Geoscience and Remote Sensing, GE-23, 35-46, 1985.

Dorigo, W., de Jeu, R., Chung, D., Parinussa, R., Liu, Y., Wagner, W., and Fernández-Prieto, D.: Evaluating global trends (1988–2010) in harmonized multi-satellite surface soil moisture, Geophysical Research Letters, 39, 2012.

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P. D., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y. Y., Miralles, D., Mistelbauer, T., Nicolai-Shaw, N., Parinussa, R., Pratola, C., Reimer, C., van der Schalie, R., Seneviratne, S. I., Smolander, T., and Lecomte, P.: ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions, Remote Sensing of Environment, 203, 185-215, 2017.

Dorigo, W. A., Scipal, K., Parinussa, R. M., Liu, Y. Y., Wagner, W., de Jeu, R. A. M., and Naeimi, V.: Error characterisation of global active and passive microwave soil moisture datasets, Hydrology and Earth System Sciences, 14, 2605-2616, 2010.

Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., van Oevelen, P., Robock, A., and Jackson, T.: The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements, Hydrology and Earth System Sciences, 15, 1675-1698, 2011.

Dorigo, W. A., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiová, A., Sanchis-Dufau, A. D., Zamojski, D., Cordes, C., Wagner, W., and Drusch, M.: Global Automated Quality Control of In Situ Soil Moisture Data from the International Soil Moisture Network, Vadose Zone Journal, 12, 2013.

Du, Y., Ulaby, F. T., and Dobson, M. C.: Sensitivity to soil moisture by active and passive microwave sensors, IEEE Transactions on Geoscience and Remote Sensing, 38, 105-114, 2000.

Entekhabi, D., Njoku, E., O'neill, P., Kellogg, K. H., Crow, W., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T., Johnson, F. M., Kimball, J., Piepmeier, J., Koster, R. D., Martin, E., McDonald, C. K., Moghaddam, M., Moran, M. S., Reichle, R., Shi, J. C., Spencer, D., Thurman, S. W., Tsang, L., and Zyl, J. V.: The Soil Moisture Active Passive (SMAP) Mission, Proceedings of the IEEE, 98, 704-716, 2010.

Feng, X., Li, J., Cheng, W., Fu, B., Wang, Y., Lü, Y., and Shao, M. a.: Evaluation of AMSR-E retrieval by detecting soil moisture decrease following massive dryland re-vegetation in the Loess Plateau, China, Remote Sensing of Environment, 196, 253-264, 2017.

Fernandez-Moran, R., Al-Yaari, A., Mialon, A., Mahmoodi, A., Al Bitar, A., De Lannoy, G., Rodriguez-Fernandez, N., Lopez-Baeza, E., Kerr, Y., and Wigneron, J.-P.: SMOS-IC: An Alternative SMOS Soil Moisture and Vegetation Optical Depth Product, Remote Sensing, 9, 2017a.

Fernandez-Moran, R., Wigneron, J. P., De Lannoy, G., Lopez-Baeza, E., Parrens, M., Mialon, A., Mahmoodi, A., Al-Yaari, A., Bircher, S., Al Bitar, A., Richaume, P., and Kerr, Y.: A new calibration of the effective scattering albedo and soil roughness parameters in the SMOS SM retrieval algorithm, International Journal of Applied Earth Observation and Geoinformation, 62, 27-38, 2017b.

Fujii, H., Koike, T., and Imaoka, K.: Improvement of the AMSR-E Algorithm for Soil Moisture Estimation by Introducing a Fractional Vegetation Coverage Dataset Derived from MODIS Data, Journal of The Remote Sensing Society of Japan, 29, 282-292, 2009.

Gaiser, P. W., Germain, K. M. S., Twarog, E. M., Poe, G. A., Purdy, W., Richardson, D., Grossman, W., Jones, W. L., Spencer, D., Golba, G., Cleveland, J., Choy, L., Bevilacqua, R. M., and Chang, P. S.: The WindSat spaceborne polarimetric microwave radiometer: sensor description and early orbit performance, IEEE Transactions on Geoscience and Remote Sensing, 42, 2347-2361, 2004.

Gao, H., Wood, E. F., Jackson, T. J., Drusch, M., and Bindlish, R.: Using TRMM/TMI to Retrieve Surface Soil Moisture over the Southern United States from 1998 to 2002, Journal of Hydrometeorology, 7, 23-38, 2006.

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

Griend, A. A. V. d. and Wigneron, J.: On the measurement of microwave vegetation properties: some guidelines for a protocol, IEEE Transactions on Geoscience and Remote Sensing, 42, 2277-2289, 2004.

Gruber, A., Dorigo, W. A., Crow, W., and Wagner, W.: Triple Collocation-Based Merging of Satellite Soil Moisture Retrievals, IEEE Transactions on Geoscience and Remote Sensing, 55, 6780-6792, 2017.

Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W., and Dorigo, W.: Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology, Earth System Science Data, 11, 717-739, 2019.

Gu, X., Li, J., Chen, Y. D., Kong, D., and Liu, J.: Consistency and Discrepancy of Global Surface Soil Moisture Changes From Multiple Model-Based Data Sets Against Satellite Observations, Journal of Geophysical Research: Atmospheres, 124, 1474-1495, 2019.

Hansen, M. and Song, X. P.: Vegetation Continuous Fields (VCF) Yearly Global 0.05 Deg. NASA EOSDIS Land Processes DAAC, 2018.

Hashimoto, S., Carvalhais, N., Ito, A., Migliavacca, M., Nishina, K., and Reichstein, M.: Global spatiotemporal distribution of soil respiration modeled using a global database, Biogeosciences, 12, 4121-4132, 2015.

Hoffmann, L., Günther, G., Li, D., Stein, O., Wu, X., Griessbach, S., Heng, Y., Konopka, P., Müller, R., Vogel, B., and Wright, J. S.: From ERA-Interim to ERA5: the considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations, Atmospheric Chemistry and Physics, 19, 3097-3124, 2019.

Holmes, T. R. H., De Jeu, R. A. M., Owe, M., and Dolman, A. J.: Land surface temperature from Ka band (37 GHz) passive microwave observations, Journal of Geophysical Research: Atmospheres, 114, 2009.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and Ferreira, L. G.: Overview of the radiometric and biophysical performance of the MODIS vegetation indices, Remote Sensing of Environment, 83, 195-213, 2002.

Huffman, G., Bolvin, D., Braithwaite, D., Hsu, K., Joyce, R., and Xie, P.: Integrated Multi-satellitE Retrievals for GPM (IMERG), version 6. NASA's Precipitation Processing Center, 2019.

Jackson, T. J. and Schmugge, T. J.: Vegetation effects on the microwave emission of soils, Remote Sensing of Environment, 36, 203-212, 1991.

Jackson, T. J., Schmugge, T. J., and Wang, J. R.: Passive microwave sensing of soil moisture under vegetation canopies, Water Resources Research, 18, 1137-1142, 1982.

Jiao, Q., Li, R., Wang, F., Mu, X., Li, P., and An, C.: Impacts of Re-Vegetation on Surface Soil Moisture over the Chinese Loess Plateau Based on Remote Sensing Datasets, Remote Sensing, 8, 2016.

Karthikeyan, L., Pan, M., Wanders, N., Kumar, D. N., and Wood, E. F.: Four decades of microwave satellite soil moisture observations: Part 1. A review of retrieval algorithms, Advances in Water Resources, 109, 106-120, 2017a.

Karthikeyan, L., Pan, M., Wanders, N., Kumar, D. N., and 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, 2017b.

Kawanishi, T., Sezai, T., Ito, Y., Imaoka, K., Takeshima, T., Ishido, Y., Shibata, A., Miura, M., Inahata, H., and Spencer, R. W.: The Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), NASDA's contribution to the EOS for global energy and water cycle studies, IEEE Transactions on Geoscience and Remote Sensing, 41, 184-194, 2003.

Kerr, Y. H., Al-Yaari, A., Rodriguez-Fernandez, N., Parrens, M., Molero, B., Leroux, D., Bircher, S., Mahmoodi, A., Mialon, A., Richaume, P., Delwart, S., Al Bitar, A., Pellarin, T., Bindlish, R., Jackson, T. J., Rüdiger, C., Waldteufel, P., Mecklenburg, S., and Wigneron, J. P.: Overview of SMOS performance in terms of global soil moisture monitoring after six years in operation, Remote Sensing

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

of Environment, 180, 40-63, 2016.

Kerr, Y. H., Waldteufel, P., Wigneron, J., Martinuzzi, J., Font, J., and Berger, M.: Soil moisture retrieval from space: the Soil Moisture and Ocean Salinity (SMOS) mission, IEEE Transactions on Geoscience and Remote Sensing, 39, 1729-1735, 2001.

Kim, H., Parinussa, R., Konings, A. G., Wagner, W., Cosh, M. H., Lakshmi, V., Zohaib, M., and Choi, M.: Global-scale assessment and combination of SMAP with ASCAT (active) and AMSR2 (passive) soil moisture products, Remote Sensing of Environment, 204, 260-275, 2018.

Kim, S., Liu, Y. Y., Johnson, F. M., Parinussa, R. M., and Sharma, A.: A global comparison of alternate AMSR2 soil moisture products: Why do they differ?, Remote Sensing of Environment, 161, 43-62, 2015.

Koike, T.: Soil moisture algorithm descriptions of GCOM-W1 AMSR2 (Rev. A), Earth Observation Research Center, Japan Aerospace Exploration Agency, 2013.

Koike, T., Nakamura, Y., Kaihotsu, I., Davaa, G., Matsuura, N., Tamagawa, K., and Fujii, H.: Development of an advanced microwave scanning radiometer (AMSR-E) algorithm for soil moisture and vegetation water content, Proceedings of Hydraulic Engineering, 48, 217-222, 2004.

Konings, A. G., Piles, M., Das, N., and Entekhabi, D.: L-band vegetation optical depth and effective scattering albedo estimation from SMAP, Remote Sensing of Environment, 198, 460-470, 2017.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World Map of the Köppen-Geiger climate classification updated, Meteorologische Zeitschrift, 15, 259-263, 2006.

Kozak, J. A., Ahuja, L. R., Green, T. R., and Ma, L.: Modelling crop canopy and residue rainfall interception effects on soil hydrological components for semi-arid agriculture, Hydrological Processes, 21, 229-241, 2010.

Kumar, S. V., Peters-Lidard, C. D., Santanello, J. A., Reichle, R. H., Draper, C. S., Koster, R. D., Nearing, G., and Jasinski, M. F.: Evaluating the utility of satellite soil moisture retrievals over irrigated areas and the ability of land data assimilation methods to correct for unmodeled processes, Hydrology and Earth System Sciences, 19, 4463-4478, 2015.

Lakhankar, T., Ghedira, H., Temimi, M., Azar, E. A., and Khanbilvardi, R.: Effect of Land Cover Heterogeneity on Soil Moisture Retrieval Using Active Microwave Remote Sensing Data, Remote Sensing, 1, 2009.

Lei, F., Crow, W. T., Shen, H., Su, C.-H., Holmes, T. R. H., Parinussa, R. M., and Wang, G.: Assessment of the impact of spatial heterogeneity on microwave satellite soil moisture periodic error, Remote Sensing of Environment, 205, 85-99, 2018.

Leroux, D. J., Kerr, Y. H., Bitar, A. A., Bindlish, R., Jackson, T. J., Berthelot, B., and Portet, G.: Comparison Between SMOS, VUA, ASCAT, and ECMWF Soil Moisture Products Over Four Watersheds in U.S, IEEE Transactions on Geoscience and Remote Sensing, 52, 1562-1571, 2014.

Lievens, H., Martens, B., Verhoest, N. E. C., Hahn, S., Reichle, R. H., and Miralles, D. G.: Assimilation of global radar backscatter and radiometer brightness temperature observations to improve soil moisture and land evaporation estimates, Remote Sensing of Environment, 189, 194-210, 2017.

Liu, Y. Y., de Jeu, R. A. M., McCabe, M. F., Evans, J. P., and van Dijk, A. I. J. M.: Global long-term passive microwave satellite-based retrievals of vegetation optical depth, Geophysical Research Letters, 38, 2011.

Liu, Y. Y., Dorigo, W. A., Parinussa, R. M., de Jeu, R. A. M., Wagner, W., McCabe, M. F., Evans, J. P., and van Dijk, A. I. J. M.: Trend-preserving blending of passive and active microwave soil moisture retrievals, Remote Sensing of Environment, 123, 280-297,

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

2012.

Lu, Z., Chai, L., Ye, Q., and Zhang, T.: Reconstruction of time-series soil moisture from AMSR2 and SMOS data by using recurrent nonlinear autoregressive neural networks, 26-31 July 2015 2015, 980-983.

Ma, H., Zeng, J., Chen, N., Zhang, X., Cosh, M. H., and Wang, W.: Satellite surface soil moisture from SMAP, SMOS, AMSR2 and ESA CCI: A comprehensive assessment using global ground-based observations, Remote Sensing of Environment, 231, 111215, 2019.

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, Geoscientific Model Development, 10, 1903-1925, 2017.

Martínez-Fernández, J., González-Zamora, A., Sánchez, N., Gumuzzio, A., and Herrero-Jiménez, C. M.: Satellite soil moisture for agricultural drought monitoring: Assessment of the SMOS derived Soil Water Deficit Index, Remote Sensing of Environment, 177, 277-286, 2016.

McColl, K. A., Alemohammad, S. H., Akbar, R., Konings, A. G., Yueh, S., and Entekhabi, D.: The global distribution and dynamics of surface soil moisture, Nature Geoscience, 10, 100, 2017.

Merriam, R. A.: A note on the interception loss equation, Journal of Geophysical Research, 65, 3850-3851, 1960.

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, Hydrology and Earth System Sciences, 15, 453-469, 2011.

Mladenova, I. E., Jackson, T. J., Njoku, E., Bindlish, R., Chan, S., Cosh, M. H., Holmes, T. R. H., de Jeu, R. A. M., Jones, L., Kimball, J., Paloscia, S., and Santi, E.: Remote monitoring of soil moisture using passive microwave-based techniques — Theoretical basis and overview of selected algorithms for AMSR-E, Remote Sensing of Environment, 144, 197-213, 2014.

Neill, P. E. O., Podest, E., and Njoku, E. G.: Utilization of ancillary data sets for SMAP algorithm development and product generation, 24-29 July 2011 2011, 2436-2439.

Njoku, E. G., Ashcroft, P., Chan, T. K., and Li, L.: Global survey and statistics of radio-frequency interference in AMSR-E land observations, IEEE Transactions on Geoscience and Remote Sensing, 43, 938-947, 2005.

Njoku, E. G. and Chan, T. K.: Vegetation and surface roughness effects on AMSR-E land observations, Remote Sensing of Environment, 100, 190-199, 2006.

Njoku, E. G., Jackson, T. J., Lakshmi, V., Chan, T. K., and Nghiem, S. V.: Soil moisture retrieval from AMSR-E, IEEE Transactions on Geoscience and Remote Sensing, 41, 215-229, 2003.

Oliva, R., Daganzo, E., Kerr, Y. H., Mecklenburg, S., Nieto, S., Richaume, P., and Gruhier, C.: SMOS Radio Frequency Interference Scenario: Status and Actions Taken to Improve the RFI Environment in the 1400–1427-MHz Passive Band, IEEE Transactions on Geoscience and Remote Sensing, 50, 1427-1439, 2012.

Owe, M., de Jeu, R., and Holmes, T.: Multisensor historical climatology of satellite-derived global land surface moisture, Journal of Geophysical Research: Earth Surface, 113, 2008.

Owe, M., Jeu, R. d., and 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, 1643-1654, 2001.

Panciera, R., Walker, J. P., Kalma, J. D., Kim, E. J., Saleh, K., and Wigneron, J.-P.: Evaluation of the SMOS L-MEB passive microwave soil moisture retrieval algorithm, Remote Sensing of Environment, 113, 435-444, 2009.

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

745 Parinussa, R. M., Holmes, T. R. H., and Jeu, R. A. M. d.: Soil Moisture Retrievals From the WindSat Spaceborne Polarimetric Microwave Radiometer, IEEE Transactions on Geoscience and Remote Sensing, 50, 2683-2694, 2012.

Parinussa, R. M., Holmes, T. R. H., Wanders, N., Dorigo, W. A., and de Jeu, R. A. M.: A Preliminary Study toward Consistent Soil Moisture from AMSR2, Journal of Hydrometeorology, 16, 932-947, 2014.

Parinussa, R. M., Holmes, T. R. H., Yilmaz, M. T., and Crow, W. T.: The impact of land surface temperature on soil moisture anomaly
750 detection from passive microwave observations, Hydrology and Earth System Sciences, 15, 3135-3151, 2011.

Qiu, J., Gao, Q., Wang, S., and Su, Z.: Comparison of temporal trends from multiple soil moisture data sets and precipitation: The implication of irrigation on regional soil moisture trend, International Journal of Applied Earth Observation and Geoinformation, 48, 17-27, 2016.

Qu, Y., Zhu, Z., Chai, L., Liu, S., Montzka, C., Liu, J., Yang, X., Lu, Z., Jin, R., Li, X., Guo, Z., and Zheng, J.: Rebuilding a Microwave
755 Soil Moisture Product Using Random Forest Adopting AMSR-E/AMSR2 Brightness Temperature and SMAP over the Qinghai–Tibet Plateau, China, Remote Sensing, 11, 2019.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The Global Land Data Assimilation System, Bulletin of the American Meteorological Society, 85, 381-394, 2004.

760 Rodríguez-Fernández, J. N., Kerr, H. Y., Van der Schalie, R., Al-Yaari, A., Wigneron, J.-P., De Jeu, R., Richaume, P., Dutra, E., Mialon, A., and Drusch, M.: Long Term Global Surface Soil Moisture Fields Using an SMOS-Trained Neural Network Applied to AMSR-E Data, Remote Sensing, 8, 2016.

Rodriguez-Fernandez, N., Aires, F., Richaume, P., Kerr, Y. H., Prigent, C., Kolassa, J., Cabot, F., Jiménez, C., Mahmoodi, A., and Drusch, M.: Soil Moisture Retrieval Using Neural Networks: Application to SMOS, IEEE Transactions on Geoscience and Remote Sensing, 53,
765 5991-6007, 2015.

Samaniego, L., Thober, S., Kumar, R., Wanders, N., Rakovec, O., Pan, M., Zink, M., Sheffield, J., Wood, E. F., and Marx, A.: Anthropogenic warming exacerbates European soil moisture droughts, Nature Climate Change, 8, 421-426, 2018.

Schroeder, R., McDonald, C. K., Chapman, D. B., Jensen, K., Podest, E., Tessler, D. Z., Bohn, J. T., and Zimmermann, R.: Development and Evaluation of a Multi-Year Fractional Surface Water Data Set Derived from Active/Passive Microwave Remote Sensing Data, Remote
770 Sensing, 7, 2015.

Shi, J., Jackson, T., Tao, J., Du, J., Bindlish, R., Lu, L., and Chen, K. S.: Microwave vegetation indices for short vegetation covers from satellite passive microwave sensor AMSR-E, Remote Sensing of Environment, 112, 4285-4300, 2008.

Stillman, S. and Zeng, X.: Evaluation of SMAP Soil Moisture Relative to Five Other Satellite Products Using the Climate Reference Network Measurements Over USA, IEEE Transactions on Geoscience and Remote Sensing, 56, 6296-6305, 2018.

775 Stocker, B. D., Zscheischler, J., Keenan, T. F., Prentice, I. C., Seneviratne, S. I., and Peñuelas, J.: Drought impacts on terrestrial primary production underestimated by satellite monitoring, Nature Geoscience, 12, 264-270, 2019.

Ulaby, F. T., Batlivala, P. P., and Dobson, M. C.: Microwave Backscatter Dependence on Surface Roughness, Soil Moisture, and Soil Texture: Part I-Bare Soil, IEEE Transactions on Geoscience Electronics, 16, 286-295, 1978.

Van der Schalie, R., De Jeu, R., Parinussa, R., Rodríguez-Fernández, N., Kerr, Y., Al-Yaari, A., Wigneron, J.-P., and Drusch, M.: The
780 Effect of Three Different Data Fusion Approaches on the Quality of Soil Moisture Retrievals from Multiple Passive Microwave Sensors,

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

Remote Sensing, 10, 2018.

van der Schalie, R., de Jeu, R. A. M., Kerr, Y. H., Wigneron, J. P., Rodríguez-Fernández, N. J., Al-Yaari, A., Parinussa, R. M., Mecklenburg, S., and Drusch, M.: The merging of radiative transfer based surface soil moisture data from SMOS and AMSR-E, Remote Sensing of Environment, 189, 180-193, 2017.

Verger, A., Baret, F., and Weiss, M.: Near Real-Time Vegetation Monitoring at Global Scale, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7, 3473-3481, 2014.

Wagner, W., Lemoine, G., and Rott, H.: A Method for Estimating Soil Moisture from ERS Scatterometer and Soil Data, Remote Sensing of Environment, 70, 191-207, 1999.

Wigneron, J., Calvet, J., Rosnay, P. d., Kerr, Y., Waldteufel, P., Saleh, K., Escorihuela, M. J., and Kruszewski, A.: Soil moisture retrievals from biangular L-band passive microwave observations, IEEE Geoscience and Remote Sensing Letters, 1, 277-281, 2004.

Wigneron, J. P., Kerr, Y., Waldteufel, P., Saleh, K., Escorihuela, M. J., Richaume, P., Ferrazzoli, P., de Rosnay, P., Gurney, R., Calvet, J. C., Grant, J. P., Guglielmetti, M., Hornbuckle, B., Mätzler, C., Pellarin, T., and Schwank, M.: L-band Microwave Emission of the Biosphere (L-MEB) Model: Description and calibration against experimental data sets over crop fields, Remote Sensing of Environment, 107, 639-655, 2007.

Xiao, Z., Liang, S., Wang, J., Chen, P., Yin, X., Zhang, L., and Song, J.: Use of General Regression Neural Networks for Generating the GLASS Leaf Area Index Product From Time-Series MODIS Surface Reflectance, IEEE Transactions on Geoscience and Remote Sensing, 52, 209-223, 2014.

Xiao, Z., Liang, S., Wang, J., Xiang, Y., Zhao, X., and Song, J.: Long-Time-Series Global Land Surface Satellite Leaf Area Index Product Derived From MODIS and AVHRR Surface Reflectance, IEEE Transactions on Geoscience and Remote Sensing, 54, 5301-5318, 2016.

Yang, H., Weng, F., Lv, L., Lu, N., Liu, G., Bai, M., Qian, Q., He, J., and Xu, H.: The FengYun-3 Microwave Radiation Imager On-Orbit Verification, IEEE Transactions on Geoscience and Remote Sensing, 49, 4552-4560, 2011.

Yang, J., Zhang, P., Lu, N., Yang, Z., Shi, J., and Dong, C.: Improvements on global meteorological observations from the current Fengyun 3 satellites and beyond, International Journal of Digital Earth, 5, 251-265, 2012.

Yao, P., Lu, H., Yue, S., Yang, F., Lyu, H., Yang, K., McColl, K. A., Gianotti, D., and ENTekhabi, D.: Estimating Surface Soil Moisture from AMSR2 Tb with Artificial Neural Network Method and SMAP Products, IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 6998-7001, 2019.

Yao, P., Shi, J., Zhao, T., Lu, H., and Al-Yaari, A.: Rebuilding Long Time Series Global Soil Moisture Products Using the Neural Network Adopting the Microwave Vegetation Index, Remote Sensing, 9, 2017.

Yilmaz, M. T., Hunt, E. R., and Jackson, T. J.: Remote sensing of vegetation water content from equivalent water thickness using satellite imagery, Remote Sensing of Environment, 112, 2514-2522, 2008.

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

Tables

Table 1: The global surface soil moisture products used in this study.

Product name	Spatial coverage	Approximate Spatial resolution	Used time period [a]
SMAP_E	Global	0.1°	2015D10~2018D36
ASCAT-SWI	Global	0.1°	2007D1~2018D36
AMSR2-JAXA	Global	0.1°	2012D19~2018D36
AMSR2-LPRM-X	Global	0.1°	2012D19~2018D36
SMOS-IC	Global [b]	0.25°	2010D16~2018D36
TMI-LPRM-X	[N40°, S40°]	0.25°	2003D1~2015D10
FY-3B-NSMC	Global	0.25°	2011D20~2017D36
WindSat-LPRM-X	Global	0.25°	2003D4~2012D21
AMSRE-JAXA	Global	0.25°	2003D1~2011D27
AMSRE-LPRM-X	Global	0.25°	2003D1~2011D27
AMSRE-NSIDC	Global	0.25°	2003D1~2011D27

[a] D represents the ordinal of the 10-day period in a year. For example, 2015D10 stands for April 1st to April 10th in 2015, while 2018D36 is December 21st to December 31st in 2018. This column lists the period of data used in this study but not the full temporal coverage of each product.

[b] SMOS has missing values in some areas in Europe and Asia due to RFI.

Table 2: The description of the Köppen-Geiger climate classification types at all the selected ISMN stations.

Climate_Köppen	General description
Aw	Equatorial savannah with dry winter
BSk	Steppe climate, cold and arid
BWh	Desert climate, hot and arid
BWk	Desert climate, cold and arid

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

Cfa	Warm temperate climate, fully humid, hot summer
Cfb	Warm temperate climate, fully humid, warm summer
Csa	Warm temperate climate with dry, hot summer
Csb	Warm temperate climate with dry, warm summer
Dfa	Snow climate, fully humid, hot summer
Dfb	Snow climate, fully humid, warm summer
Dfc	Snow climate, fully humid, cool summer and cold winter
Dsb	Snow climate with dry, warm summer
Dwc	Snow climate with cool summer and cold, dry winter
ET	Tundra climate

820 **Table 3: The values of five evaluation indexes (correlation coefficient: R, RMSE, bias, unbiased RMSE (ubRMSE) with the unit of m^3 m^{-3}, and the anomalies R (A.R)) on the temporal accuracy of the simulated soil moisture (SIM) and SMAP during April 2015~2018, when compared with site measurements.**

Index	R		RMSE		bias		ubRMSE		A.R	
Product	SIM	SMAP	SIM	SMAP	SIM	SMAP	SIM	SMAP	SIM	SMAP
Mean	0.756	0.762	0.075	0.074	0.015	0.016	0.043	0.043	0.700	0.707
Median	0.795	0.798	0.067	0.066	0.009	0.013	0.043	0.043	0.720	0.744

Table 4:The statistics of the five evaluation indexes (R, RMSE, bias, ubRMSE and A.R) on the spatial pattern accuracy of SIM and SMAP in each 10-day period during April 2015~2018.

Index	R		RMSE		bias		ubRMSE	
Product	SIM	SMAP	SIM	SMAP	SIM	SMAP	SIM	SMAP
Mean	0.652	0.659	0.084	0.084	0.016	0.016	0.082	0.081
Median	0.655	0.664	0.082	0.081	0.019	0.019	0.080	0.078

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

Open Access Earth System Science Data Discussions

825 **Table 5:The values of evaluation indexes (R, RMSE, bias, ubRMSE and A.R) on the temporal accuracy of SIM and ASCAT-SWI (abbreviated as ASC) during 2007~2018.**

Index	R		RMSE		bias		ubRMSE		A.R	
Product	SIM	ASC	SIM	ASC	SIM	ASC	SIM	ASC	SIM	ASC
Mean	0.687	0.561	0.079	0.095	0.002	-0.007	0.047	0.062	0.627	0.554
Median	0.735	0.627	0.074	0.088	-0.001	-0.010	0.048	0.062	0.654	0.595

Table 6: The statistics of the five evaluation indexes (R, RMSE, bias, ubRMSE and A.R) on the spatial pattern accuracy of SIM and ASCAT during 2007~2018.

Index	R		RMSE		bias		ubRMSE	
Product	SIM	ASCAT	SIM	ASCAT	SIM	ASCAT	SIM	ASCAT
Mean	0.636	0.561	0.087	0.102	0.005	-0.010	0.085	0.097
Median	0.650	0.572	0.086	0.100	0.007	-0.009	0.085	0.095

Table 7:The values of the five evaluation indexes (R, RMSE, bias, ubRMSE and A.R) on the temporal accuracy of
830 **SIM and GLDAS (GLD) soil moisture during 2003~2018.**

Index	R		RMSE		bias		ubRMSE		A.R	
Product	SIM	GLD	SIM	GLD	SIM	GLD	SIM	GLD	SIM	GLD
Mean	0.689	0.613	0.080	0.091	0.001	0.028	0.047	0.051	0.620	0.519
Median	0.737	0.661	0.075	0.082	-0.002	0.029	0.048	0.049	0.661	0.567

Table 8. The statistics of the five evaluation indexes (R, RMSE, bias, ubRMSE and A.R) on the spatial pattern accuracy of SIM and GLDAS (GLD) during 2003~2018.

Index	R		RMSE		bias		ubRMSE	
Product	SIM	GLD	SIM	GLD	SIM	GLD	SIM	GLD
Mean	0.617	0.593	0.090	0.097	-0.005	0.035	0.086	0.087
Median	0.643	0.630	0.089	0.096	0.001	0.041	0.086	0.086

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

Open Access Earth System Science Data Discussions

Figures

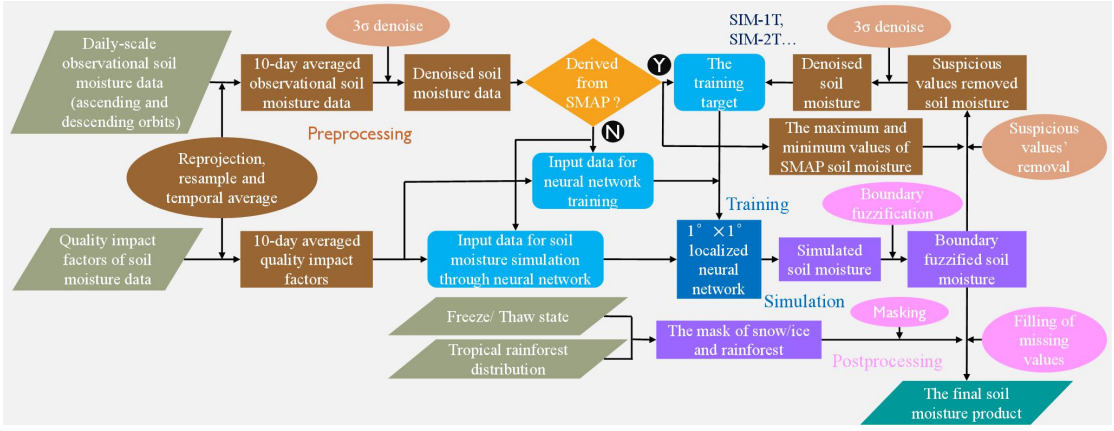


Figure 1: The flow chart of global surface soil moisture data production.

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

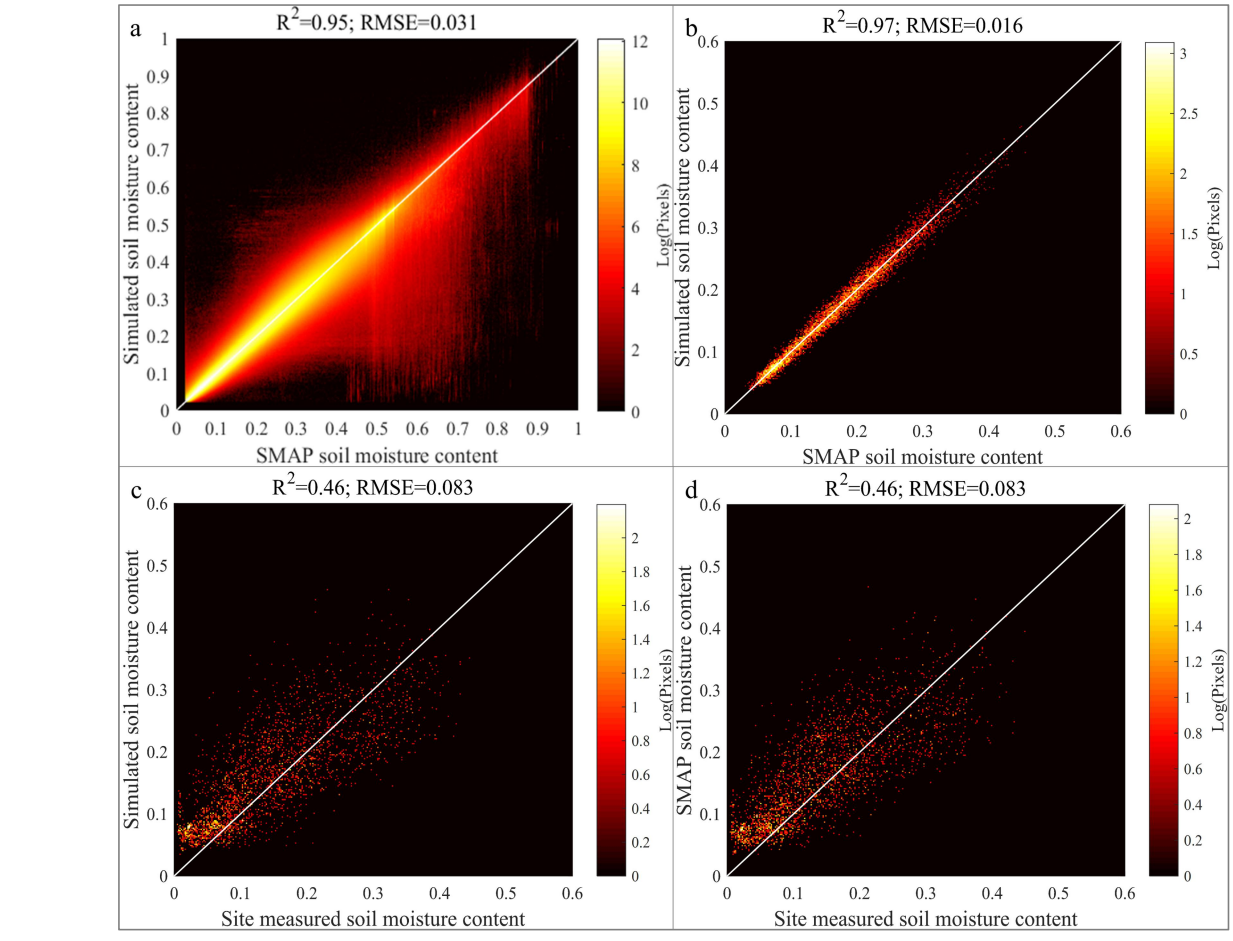


Figure 2: Comparison between the neural network simulated surface soil moisture (SIM) and SMAP data. The scatter plots are between: (a) SIM and SMAP values at all grids; b) SIM and SMAP values at only the grids with measurements; (c) SIM and the site measured soil moisture; and (d) SMAP and the site measurements during April 2015~2018. All plots are represented as the point density on a logarithmic scale, while the unit of soil moisture content and RMSE values is $m^3 m^{-3}$.

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

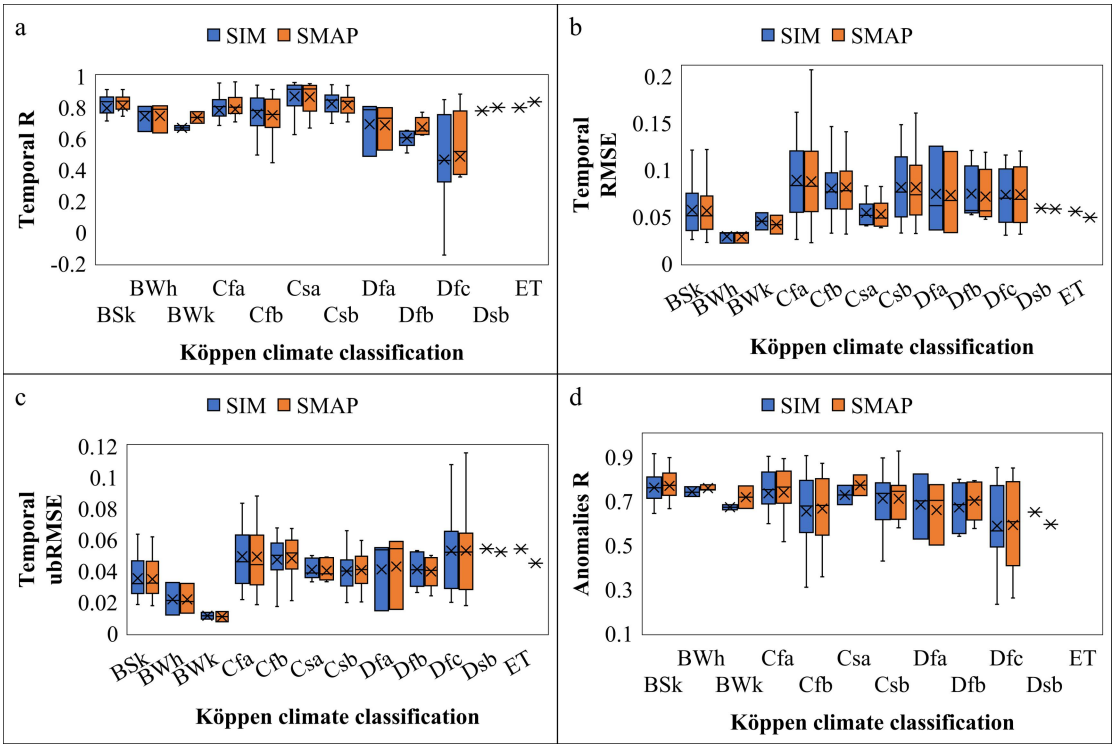


Figure 3: Comparison between the temporal accuracy of SIM and SMAP in regions with different Köppen-Geiger climate types. The four indexes are (a) R, (b) RMSE, (c) ubRMSE and (d) Anomalies R (A.R). The lengths of the error bars are 1.5 times that of the interquartile range, while the upper and lower boundaries and the central lines of the boxes indicate the 75th, 50th and 25th percentile values, with mean values marked by '×' (the forms of all the following boxplots are the same).

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

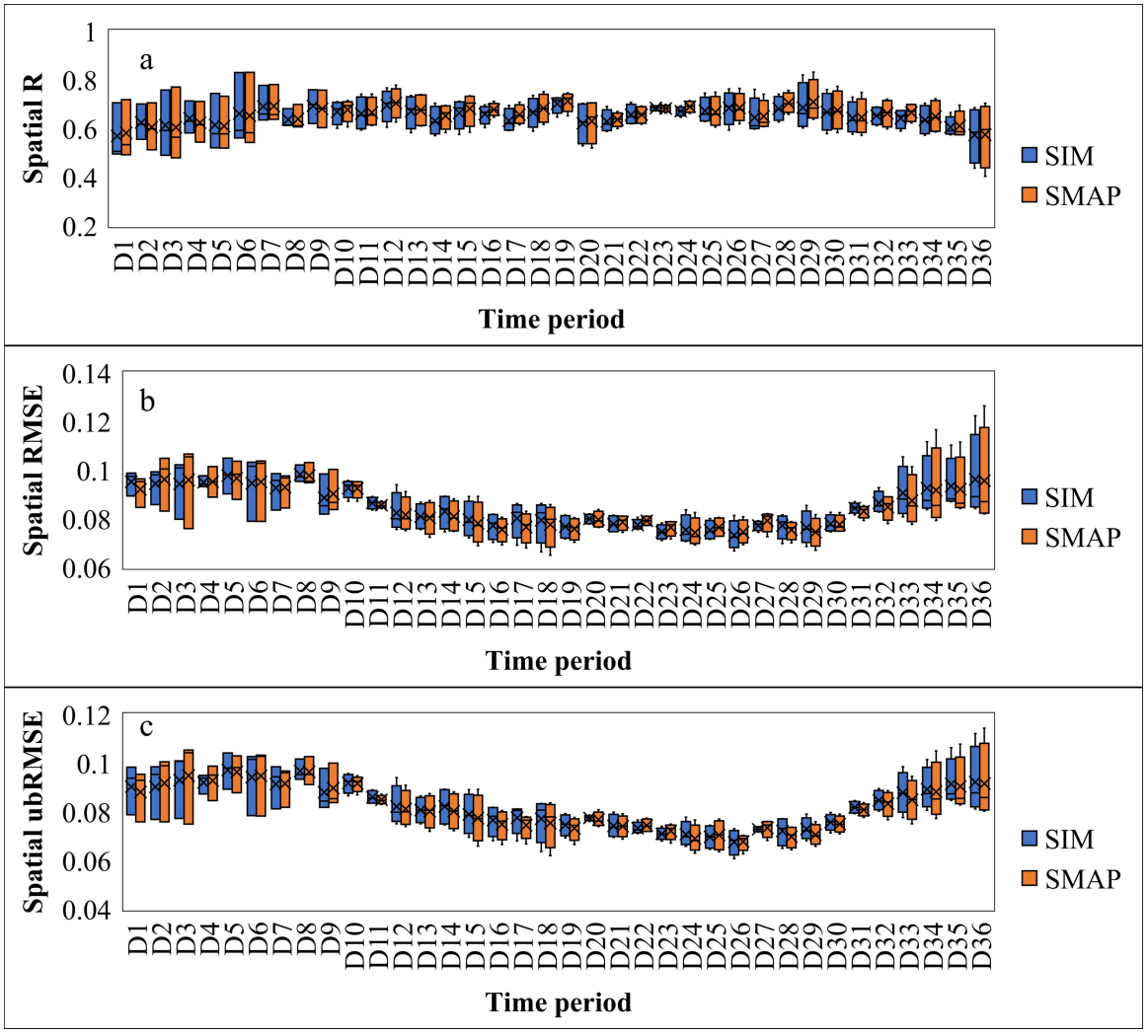


Figure 4: Comparison between the spatial pattern accuracy of SIM and SMAP in different 10-day periodsduring April 2015~2018. The three evaluation indexes are: (a) R, (b) RMSE and (c) ubRMSE. The length of each box/error bar is determined from the evaluation index values in three (January to March) or four (April to December) years.

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

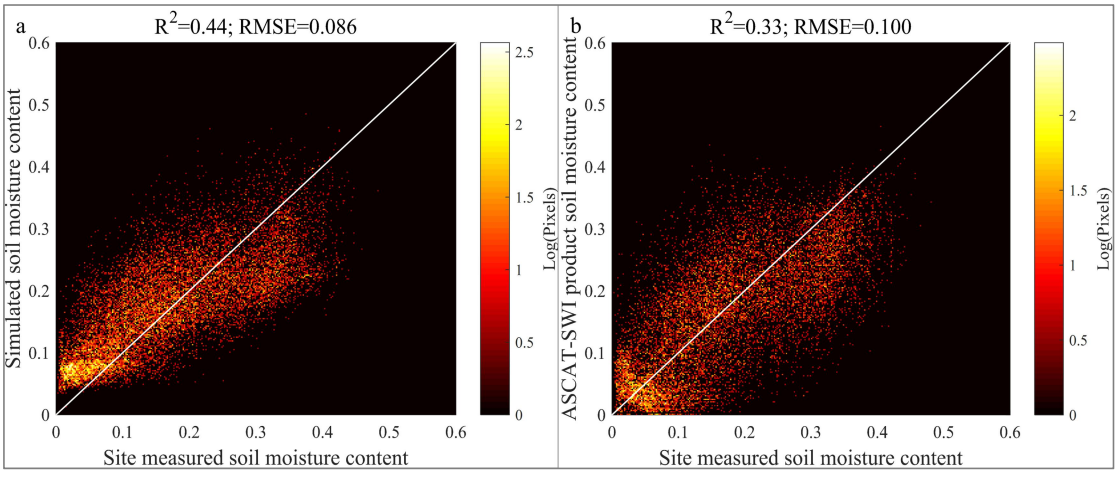


Figure 5: The overall data accuracy comparison between SIM and the ASCAT-SWI data product. The scatter plot is between: (a) SIM or (b) ASCAT-SWI soil moisture and the site measured values during 2007~2018. The unit of all plots is the density of points on a logarithmic scale.

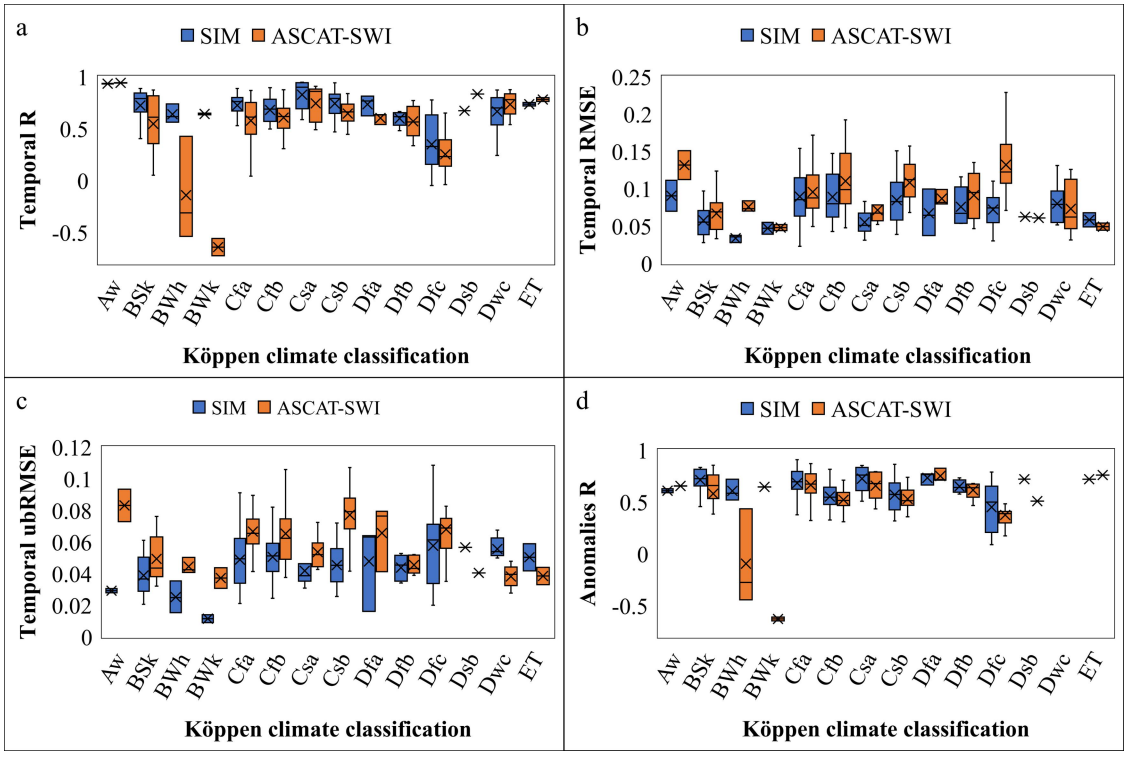


Figure 6: Comparison between the temporal variation accuracy of SIM and ASCAT-SWI in different Köppen-Geiger climatic regions. The four indexes are: (a) R, (b) RMSE, (c) ubRMSE, and (d) Anomalies R (A.R).

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

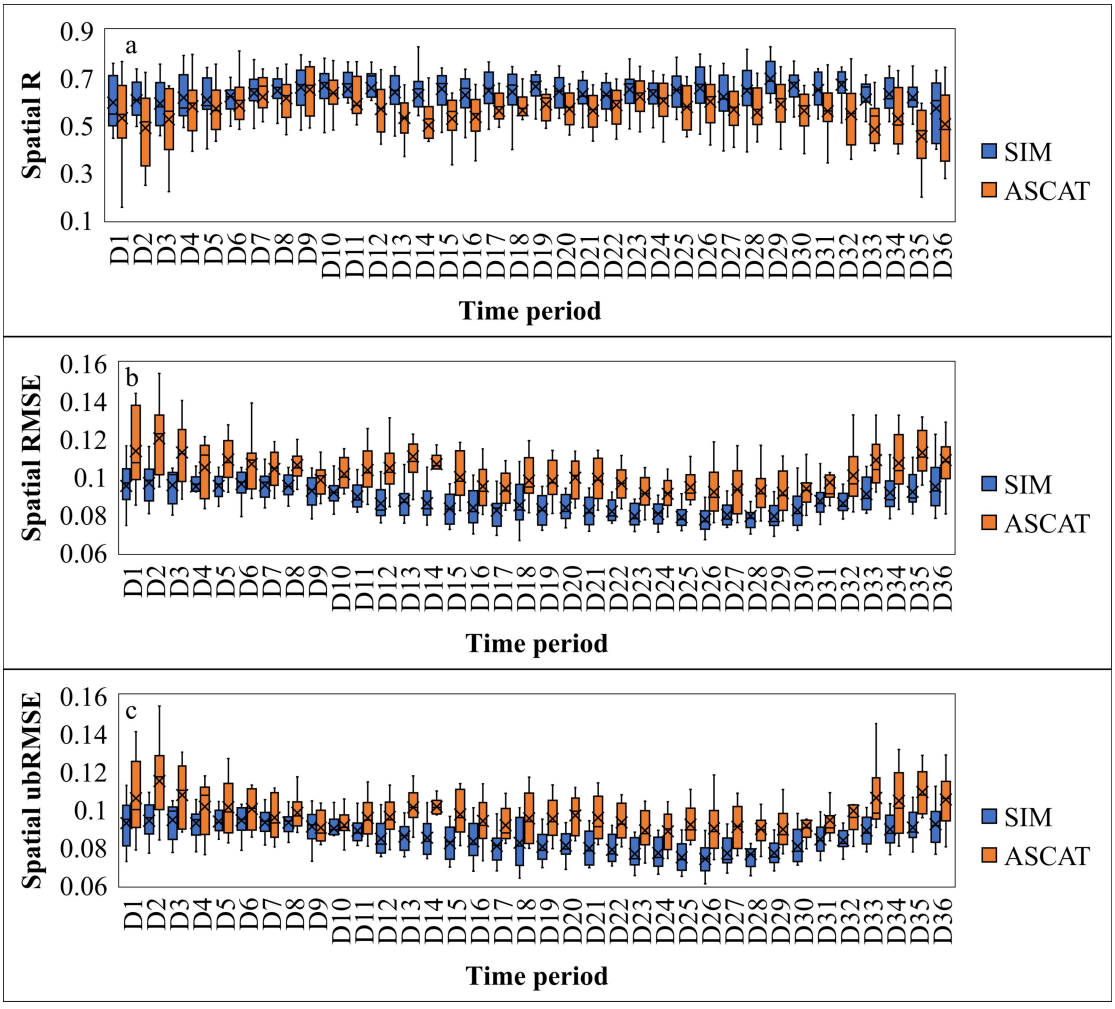


Figure 7: Comparison between the spatial accuracy of SIM and ASCAT-SWI during different 10-day periods. The evaluation indexes are: (a) R, (b) RMSE, and (c) ubRMSE.

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

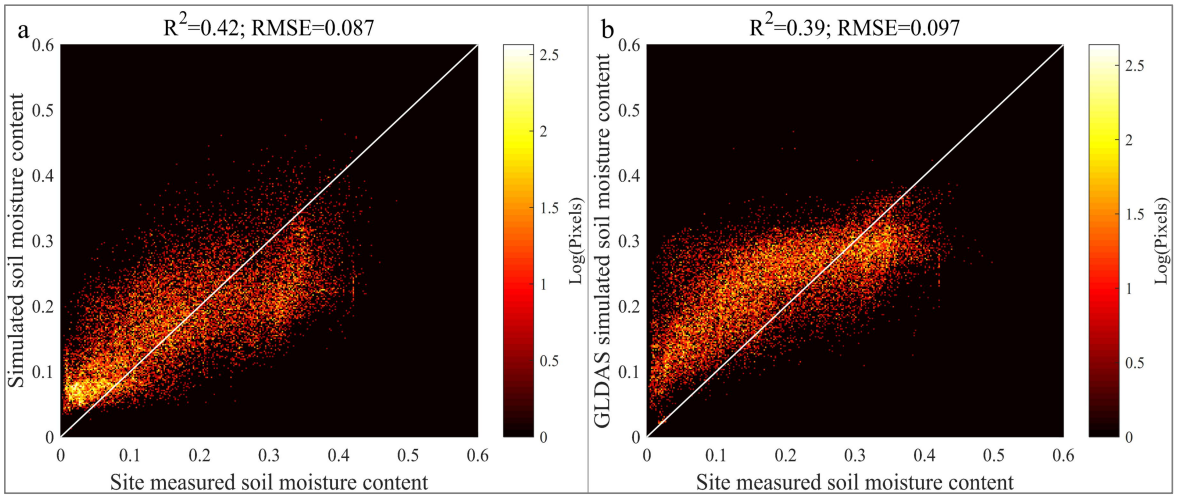


Figure 8: The overall data accuracy comparison between SIM and the surface soil moisture simulated by GLDAS Noah V2.1. The scatter plot is between the (a) SIM or (b) GLDAS soil moisture and the measured soil moisture during 2003~2018.

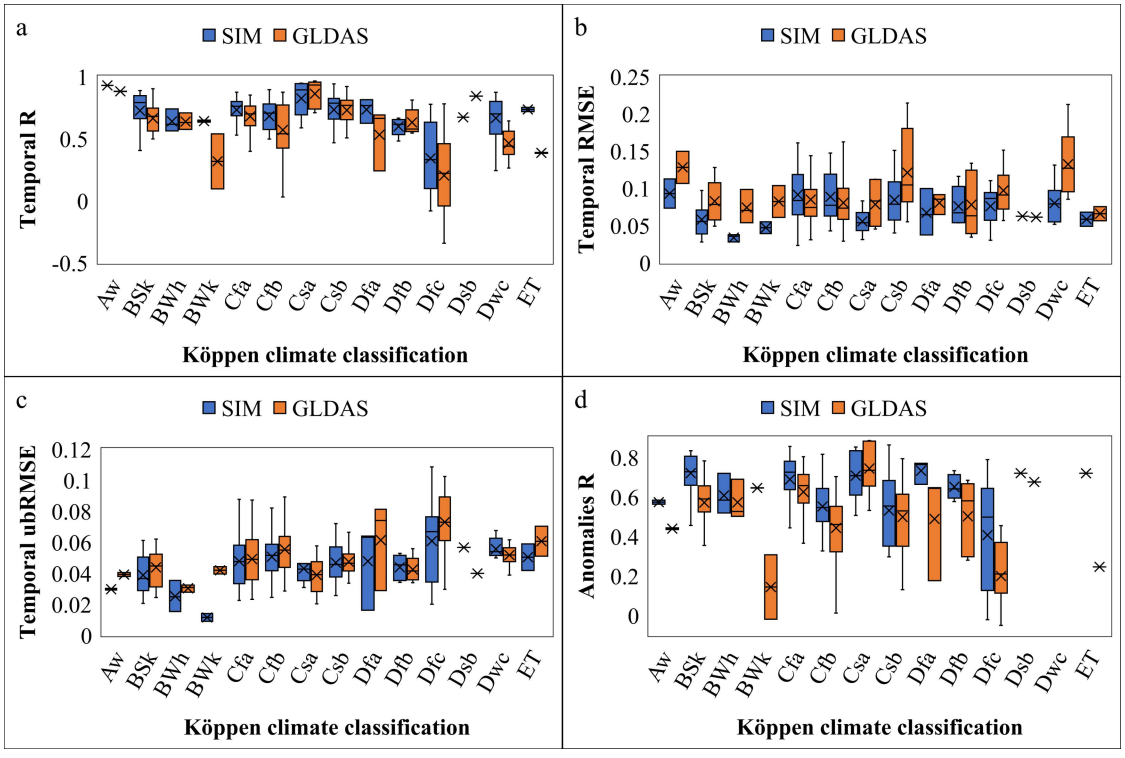
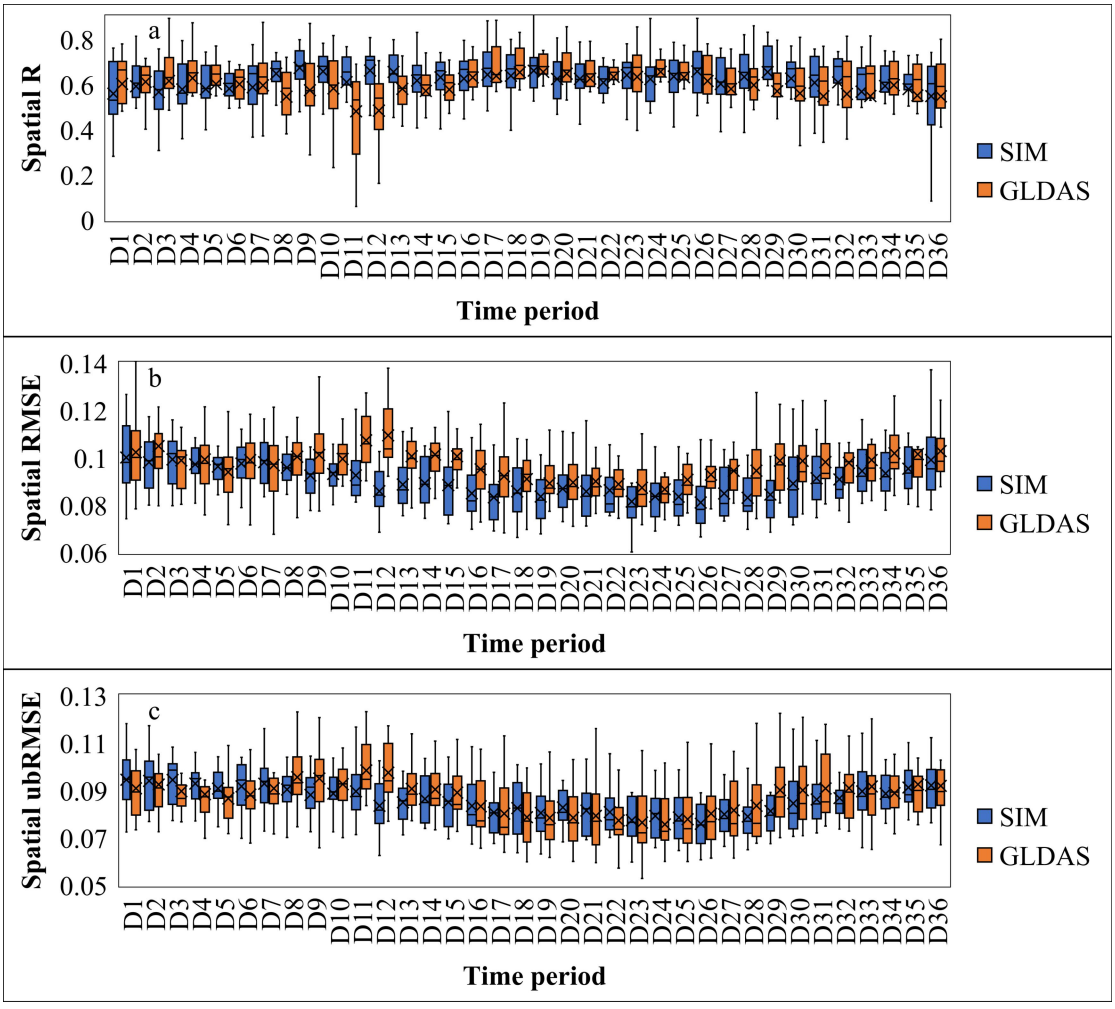


Figure 9:Comparison between the temporal accuracy of SIM and GLDAS surface soil moisture in regions with different Köppen-Geiger climate types. The four indexes are: (a) R, (b) RMSE, (c) ubRMSE, and (d) Anomalies R.

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020



875 **Figure 10: Comparison between the spatial accuracy of SIM and GLDAS during different 10-day periods. The evaluation indexes are the same as those in Figure 7.**

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

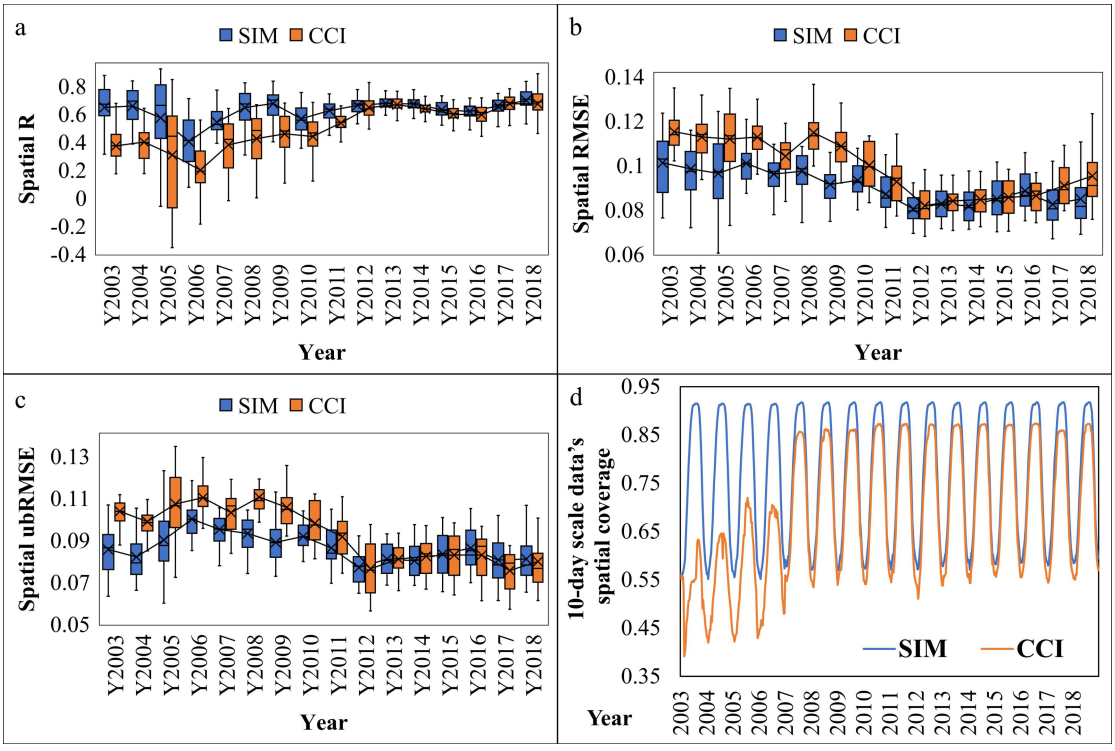


Figure 11: Changes in the data quality and data spatial coverages of SIM and CCI soil moisture with year. The interannual changes in (a) spatial correlation coefficients, (b) spatial RMSE, (c) spatial ubRMSE values, and (d) the spatial coverages of 10-day period data of SIM and CCI.

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

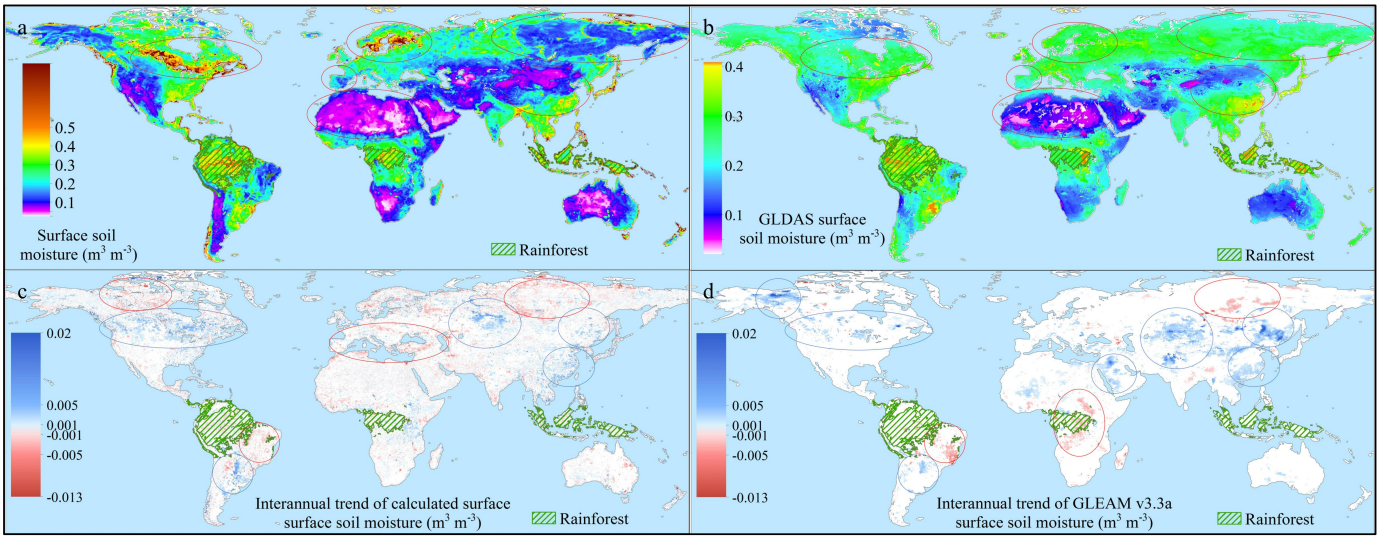
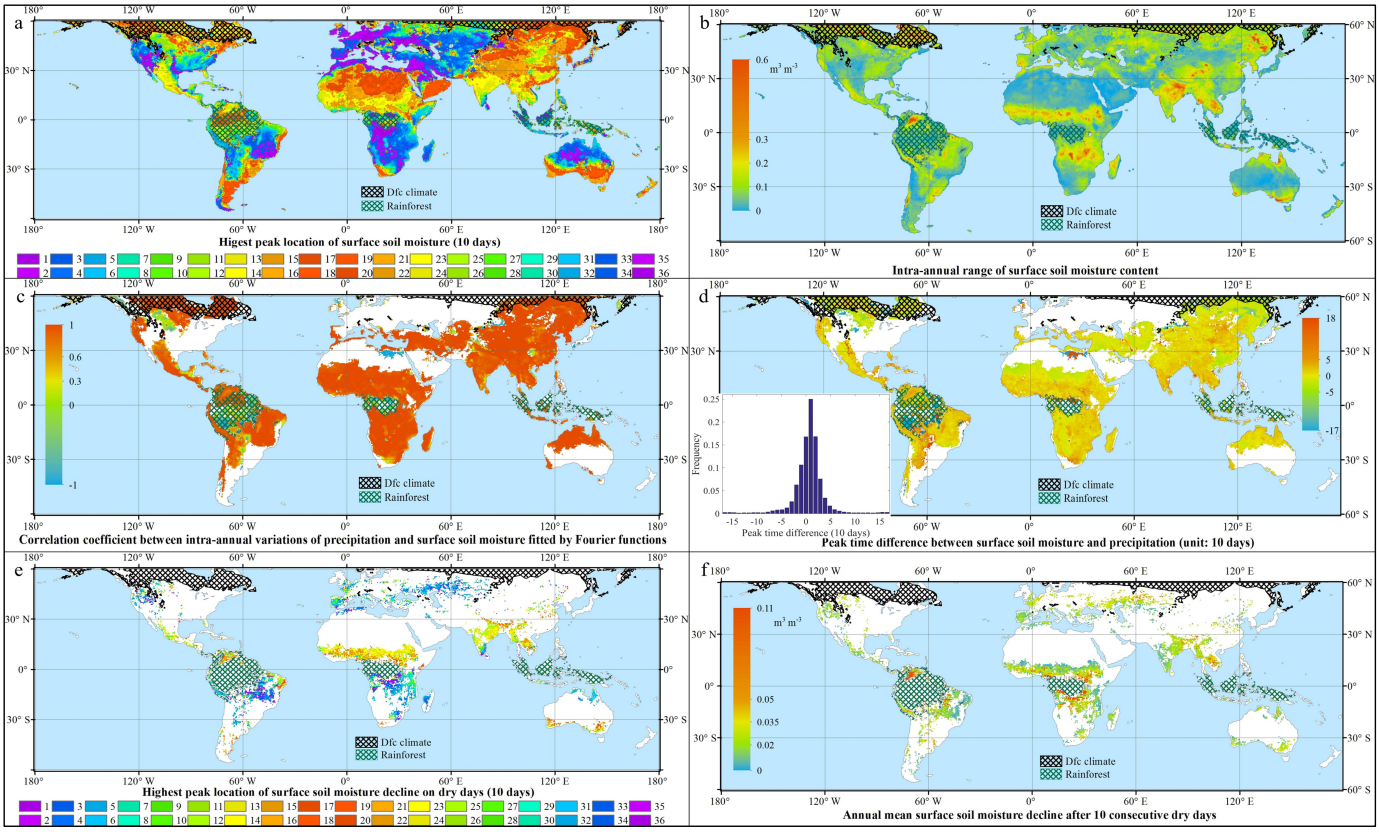


Figure 12: The spatial and temporal patterns of the neural network simulated surface soil moisture and comparison against other products: (a~b) the global map of (a) calculated soil moisture and (b) GLDAS Noah V2.1 soil moisture (averaged during 2003~2018); (c~d) the interannual trend map of (c) calculated soil moisture and (d) GLEAM v3.3a soil moisture from 2003 to 2018. The circled regions in (a~b) are the places with obvious differences, while the circled regions in (c~d) are those with significant trends.

https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020



https://doi.org/10.5194/essd-2020-59
Preprint. Discussion started: 8 May 2020

Open Access Earth System Science Data Discussions

Figure 13: The intra-annual variation in surface soil moisture and its relationship with precipitation at global scale. (a) The global distribution of the highest peak location in time of the calculated surface soil moisture (unit: 10 days, note that the seasons are opposite in the Northern and Southern Hemispheres); (b) the global spatial pattern of the intra-annual range of surface soil moisture; (c) the map of the correlation coefficient between the intra-annual variations in precipitation and surface soil moisture fitted by Fourier periodic functions; (d) the peak time difference between the intra-annual variations in surface soil moisture and precipitation (unit: 10 days), and the inset is the frequency histogram; (e) the spatial distribution of the highest peak location of surface soil moisture decline on dry days (unit: 10 days); and (f) map of the estimated annual mean decline in surface soil moisture after 10 consecutive dry days (assuming the dry period occurs randomly over a year).