An improved dataset of satellite observation-based global surface soil moisture covering 2003~2018 (RSSSM)

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 terrestrial ecosystems. However, long-term satellite monitoring of surface soil moisture is still lacking at the global scale. In this study, we conducted data calibration and data fusion of 11 well-acknowledged microwave remote sensing soil moisture products since 2003 through a neural network approach, with SMAP soil moisture data applied as the primary training target. The training efficiency was 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 neural networks (5 rounds of simulations; 8 substeps; 67 independent neural networks; and more than one million zonal subnetworks). We achieved global satellite monitoring of surface soil moisture during 2003~2018 at 0.1° resolution. The temporal resolution is approximately 10 days, or to be specific, there are 3 data records within a month, for days 1~10, 11~20 and from 21 to the last day of that month. This new dataset, named RSSSM, is proved comparable to the in-situ surface soil moisture measurements at the International Soil Moisture Network sites (overall R^2 and RMSE values of 0.42 and 0.087 m^3/m^3), while the overall R^2 and RMSE values for the existing products (ASCAT-SWI, GLDAS Noah, ERA5-Land, CCI/ECV and GLEAM) are within the range of 0.31~0.41 and 0.095~0.142 m^3/m^3, respectively. The advantage of RSSSM is especially obvious in arid or relatively cold areas, and during growing seasons. Moreover, the persistent high data quality as well as complete spatial coverage ensure the applicability of RSSSM to studies on both the spatial and temporal patterns. Our new data suggest an increase in the global mean surface soil moisture. These data also reveal that without considering the deserts and rainforests, the surface moisture decline on consecutive rainless days is highest in summers over the low latitudes (30° S~30°N) but highest in winters over most midlatitude areas (30°N~60°N; 30°S~60°S). Notably, the error propagation is well controlled with the extension of the simulation period to the past, indicating that the data fusion algorithm proposed in this study will be even more meaningful in the future when more advanced microwave sensors are become operational. The dataset can be accessed at https://doi.pangaea.de/10.1594/PANGAEA.912597 (Chen, 2020).

1 Introduction

Soil moisture plays an important role in modulating the exchange 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). Soil moisture has been endorsed by the Global Climate Observing System (GCOS) as an essential climate variable (Bojinski et al., 2014), as it is probably the best indicator of ecological droughts (Martínez-Fernández et al., 2016; Samaniego et al., 2018). However, due to the large uncertainty in global-scale soil moisture data, the applicability of these data in global ecosystem models is currently limited (Hashimoto et al., 2015; Stocker et al., 2019).

Reanalysis-based land surface model products are the most frequently used, mainly including the Global Land Data Assimilation System (GLDAS, with 0.25° resolution) (Rodell et al., 2004), European Reanalysis (ERA)-interim (0.75°) (Balsamo et al., 2015) and its successors- ERA5 (0.25°) and ERA5-Land (0.1°) (Hoffmann et al., 2019)). Although 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 are rarely considered (Kumar et al., 2015; Qiu et al., 2016). Apart from surface soil moisture that can be observed by satellites, the modeling method also provides information on the moisture in deeper soil layers.

With the advance of remote sensing technology, soil moisture products derived from microwave remote sensing, which have been proven to be superior to those derived from other electromagnetic wave bands (Karthikeyan et al., 2017a), have become an alternative to surface soil moisture monitoring (current satellite microwave sensors can detect only soil moisture within the top 5 cm of soil) (Feng et al., 2017; Jiao et al., 2016; Piles et al., 2018). 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 considered to be unsatisfactory untill the launch of AMSRE in June 2002 (Karthikeyan et al., 2017b; Kawanishi et al., 2003). Currently, the ASCAT product is the longest continuous record of global surface soil moisture that is derived only from microwave remote sensing (Bartalis et al., 2007), and the temporal span of this product is from 2007 until present. Apart from the relatively

short time series, satellite-based soil moisture products usually have lower accuracies than modeled products (Albergel et al., 2012; Chen et al., 2013) due to various disturbances, such as high vegetation cover, high open water fractions and complex topography (Draper et al., 2012; Fan et al., 2020; Ye et al., 2015). Moreover, large discrepancies exist among the soil moisture retrieved from various sensors and that by using different algorithms (Kim et al., 2015a; Mladenova et al., 2014). Although new sensors such as SMOS (Stillman and Zeng, 2018) and SMAP (Entekhabi et al., 2010), can produce significantly improved estimates because L-band microwaves (1~2 GHz) can penetrate the vegetation canopy better than the other bands (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 Eurasia 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 available since only March 2015.

Because both simulated and satellite-observed soil moisture products have advantages and disadvantages, the interest in data fusion is increasing. 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 the v4.5 product covers 1978~2018. Two steps contribute to the combined CCI product. The first step involves rescaling all microwave sensors' retrievals against the reference data (GLDAS Noah product) by cumulative distribution function (CDF) matching, while the second step merges 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 the CCI covers more than 40 years, the data before June 2002 have many missing values and are of low quality. Upon rescaling through CDF matching, the spatial patterns of the satellite products are generally replaced by those of GLDAS (Gruber et al., 2019; Liu et al., 2012; Liu et al., 2011b). Although the satellite-observed temporal patterns are retained, the merging algorithm is probably too simple (Liu et al., 2012) to harmonize the discrepancy among the temporal variations in various products (Feng et al., 2017). Another popular data product, Global Land Evaporation Amsterdam Model (GLEAM) soil moisture, is produced by assimilating CCI data (Burgin et al., 2017; Martens et al., 2017; Miralles et al., 2011). Currently, CCI soil moisture anomalies (the deviations to the

seasonal climatology, which indicate whether the soil moisture at a time point is more humid or drier than the multiyear average) are assimilated instead of the original CCI time series (Martens et al., 2017). Therefore, the observed spatial information is ignored, while the temporal changes are mainly driven by model simulations, 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 the soil moisture data retrieved from each sensor by using CDF matching followed by averaging the rescaled data during a common period, which is adopted in CCI, will result in problems when unifying the temporal variations in different soil moisture products. Three methods have been proposed that target the use of the information acquired by one sensor to produce soil moisture data compatible with those retrieved from another. Based on physical-based equations (Wigneron et al., 2004), the regression between SMOS soil moisture and the dual-polarized brightness temperature (Tb) data from AMSRE is used to calibrate the AMSRE soil moisture time series (R-square =0.36) (Al-Yaari et al., 2016). An example of the second method uses the Land Parameter Retrieval Model (LPRM) (Owe et al., 2008) to retrieve soil moisture from SMOS and then match these data 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 that links SMOS soil moisture to 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 has been proven to be the best choice according to the connection between precipitation and the changes in soil moisture, as evaluated through a data assimilation technique and triple collocation analysis result (Van der Schalie et al., 2018).

A global long-term observational-based soil moisture product was recently developed by building a neural network between the SMOS product and the Tb data from AMSRE (2003~September 2011) and AMSR2 (July 2012~2015) (Yao et al., 2017). Some environmental factors, including land surface temperature (LST) derived from Tb at 36.5 GHz (Holmes et al., 2009) and the microwave vegetation index (MVI, an indicator of vegetation cover), were also incorporated as ancillary inputs. The

training R-square value (R^2) of this product was only 0.45 (or correlation coefficient, r, equals 0.67), and the validation against in-situ measurements showed a temporal *r* of 0.52 and temporal RMSE of 0.084. 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 have become increasingly available, these data has been chosen as the training target, improving the training R^2 to 0.55, while the overall *r* and RMSE against measurements were 0.44 and 0.113 (Yao et al., 2019). Another 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 the environmental factors, while vegetation cover is not considered, elevation (DEM), IGBP land use cover type, grid location and the day of a year (DOY) were chosen as ancillary inputs. The training R^2 in this region reached 0.9, with a temporal accuracy higher than that of other products (temporal r=0.7; RMSE=0.07 in the unfrozen season). However, these data are regional (for Tibetan Plateau only), and with a temporal gap between AMSR-E and AMSR2 data (October 2011~June 2012). To be concluded, while previous studies have focused on developing long-term satellite-based surface moisture products using machine learning, there remain some major concerns remain that need to be solved. 1) The microwave observations from only three sensors at most are utilized, leading to large temporal and spatial gaps, and the limited training efficiency; 2) it remains unclear which environmental factors should be incorporated as ancillary inputs, and why; and 3) the training designed for soil moisture estimation at the global scale should more complex than that for only a specific region to ensure a satisfactory training efficiency. In this study, 11 high-quality microwave soil moisture products since 2003 are incorporated into 5 rounds of neural networks to achieve a spatially and temporally continuous simulation for 2003~2018, using as many sources of microwave observational data as possible as predictors in each neural network. The quality impact factors of microwave soil moisture retrievals are also determined and then utilized as ancillary inputs to improve the training efficiency. Moreover, we designed localized subnetworks instead of only one global-scale neural network to account for the regional differences in training rules.

2 Data and Methods

2.1 Data for the production of global long-term surface soil moisture data

2.1.1 Satellite-based surface soil moisture data products

SMAP currently has the highest quality of all remote sensing-based soil moisture products (Al-Yaari et al., 2019) 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, hereinafter SMAP_E for short), which was developed by improving the 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 often used Tb observations at various bands as network inputs (Rodríguez-Fernández et al., 2016). However, in this study, the well-acknowledged surface soil moisture products retrieved through mature algorithms (see Figure 1) are directly applied instead of Tb. This is because 1) the primary goal of this study is to calibrate and then fuse the existing popular microwave soil moisture products; and 2) the Tb signals at multiple bands contain too much information that is not related to soil moisture, which may weaken the training efficiency and lead to overfitting. Although 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 2.1.2). The first is the ASCAT soil moisture index product (ASCAT-SWI), which was developed by the ESA-Copernicus Land Monitoring Service (Albergel et al., 2008; Wagner et al., 1999). The saturation degree in 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 'SMAP Constant'). Second, AMSR2-JAXA is the AMSR2 soil moisture retrieved by the Japan Aerospace Exploration Agency (JAXA) using Tb at the X-band (10.65 GHz) (Fujii et al., 2009), and version 3 data on the Global Portal System (G-Portal) were used. Third, 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 were not applied due to high RFI (Njoku et al., 2005)), and is obtained from NASA's Earthdata Search web. The fourth predictor, SMOS-IC, is a new SMOS soil moisture product created by INRA and CESBIO with the main goal of being as independent as possible from the

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 has been proven to be higher than that of 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), which is used as the 5th predictor. Fengyun 3B is a Chinese meteorological satellite with a Microwave Radiation Imager (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' (the 6th predictor product). WindSat is flown on the Coriolis satellite (Gaiser et al., 2004), and the soil moisture retrieved by LPRM at the X-band (Parinussa et al., 2012) is provided by NASA (the 7th predictor). Three AMSRE products are also 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 (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 reprojected to the WGS-1984 reference coordinate system and resampled to 0.1°.

To reduce noise and fill the gaps between sensor observation tracks (it takes at least 3 days for a microwave sensor to cover the whole globe), for every soil moisture product, both the daytime and nighttime observations within each 10-day period are combined by data averaging (the relative superiority of daytime and nighttime retrievals is not considered). For example, for SMAP, 11% of the global land surface has data for only 5 days or less within a 10-day period.

2.1.2 The quality impact factors of soil moisture retrievals

Environmental factors, including DEM, LST and vegetation cover (indicated by NDVI, MVI, etc.), were used as ancillary neural network inputs to improve the soil moisture simulation (Lu et al., 2015; Qu et al., 2019; Yao et al., 2017). According to these studies, these factors alone may not predict surface soil moisture well without the incorporation of any microwave remote sensing data, which can also be justified by the contribution analysis results (Figure S1a). This is because although they are somewhat related to soil moisture (e.g. soil moisture is generally limited in areas with low vegetation cover but high in forests (McColl et al., 2017)), the relationships are rather uncertain (e.g., at smaller scales, leaf area index (LAI) may have

a negative influence on soil moisture due to the variation in evapotranspiration (Naithani et al., 2013), or without clear impacts (Zhao et al., 2010); also, soil moisture can be either high or low in summers when vegetation peaks (Baldocchi et al., 2006; Méndez-Barroso et al., 2009)). However, these factors are quite essential due to their direct impacts on soil moisture retrieval through the radiative transfer model using microwave remote sensing data (Fan et al., 2020), and are retrieval quality impact factors. The detailed explanations are as follows: 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 is overestimated by LPRM (Ma et al., 2019), whereas soil moisture is underestimated by JAXA (Kim et al., 2015a), and the magnitudes of the biases are 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., 2015a). Therefore, for improved data fusion, the weights assigned to different predictor soil moisture (or Tb) predictor data available at the same time should be determined by referring to these quality impact factors (Kim et al., 2015b).

In this study, 9 quality impact factors: LAI, water fraction, LST, land use cover, tree cover fraction, non-tree vegetation fraction, topographic complexity, and sand and clay fractions are selected and incorporated (see Figure 1). The reasons are as follows. Based on the two criteria above, the first environmental factor to be included is the 'vegetation factor' (i.e., vegetation water content, VWC). Plants can absorb or scatter 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 their longer wavelengths (Konings et al., 2017; Piles et al., 2018). On the other hand, although vegetation effects can be somewhat corrected (Jackson and Schmugge, 1991), different methods have different efficiencies. First-order radiative transfer models such as LPRM have difficulty describing the radiation attenuation by dense canopy (Crow et al., 2010), but the TU-Wien change detection algorithm applied to ASCAT can reduce vegetation impacts due to the implicit account of high-order scattering effects (Bartalis et al., 2007). Microwave vegetation indexes may contain large uncertainty and have coarse

resolutions (Liu et al., 2011a; Shi et al., 2008). NDVI becomes saturated at high vegetation cover (Huete et al., 2002). Because the 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 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 (Figure 1). The LAIs are averaged on a monthly scale and aggregated to 0.1° resolution. The second is the 'water fraction factor' (i.e., the fraction of water area in each pixel). Waters in land pixels dramatically decrease the Tb, leading to overestimation of soil moisture there. Because different methods are used to detect and correct small areas of water, either open water, wetlands or partly inundated wetlands and croplands (Entekhabi et al., 2010; Kerr et al., 2001; Mladenova et al., 2014; Njoku et al., 2003), microwave soil moisture data calibration and weight assignment based on the water fraction within land pixels make sense (Ye et al., 2015). In addition, the water fraction is a direct indicator of surface soil moisture. In this study, daily water area fraction derived from the Surface WAter Microwave Product Series (SWAMPS) v3.2 dataset (Schroeder et al., 2015) is applied. The third factor is the 'heat factor' (i.e., LST). Soil moisture retrievals from passive microwave sensors are based on the correlation between the soil dielectric constant, which 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 are used in the retrievals of different soil moisture products, while the bias of each LST estimate compared to the actual LST is influenced by the actual LST, we assume that that the actual LST can determine the accuracy of every LST estimate and finally the relative performances of various soil moisture products (Kim et al., 2015a). In this study, we averaged the MODIS monthly LST acquired from the ascending and descending passes of both TERRA and AQUA. The 4th factor 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 the 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 factors can influence the relationship between soil moisture and emissivity or the 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 various soil moisture products (Das and O'neill, 2010; Gao et al., 2006; Kim et al., 2015a). 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 (topographic complexity data are not available from SMAP Constant; soil texture is not provided by ASCAT Constant). The contribution analysis results (Figure S1) show that because various microwave soil moisture data have already been included, precipitation data are not an essential indicator of soil moisture, and are not utilized as a physically based 'quality impact factor' either (see Text S1 for detailed explanations).

2.2 Methods for the production of global long-term surface soil moisture data

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. Because the temporal coverage of SMAP does not overlap with that of TMI, FY-3B, WindSat or AMSRE, several rounds of simulations are performed to fully utilize the satellite-based soil moisture data. Hence, the simulated soil moisture may also be converted to the training target of the next round's neural network, meaning that some postprocessing steps are also preprocessing steps. The basic flow of this process is shown in Figure 2.

2.2.1 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; however, 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) were patchy, with clear boundaries between adjacent square-shaped zones over arid regions, while the patch size was exactly 1°×1°, which was probably due to the spatial distribution of parameters. This finding suggests that subnetworks should be built at the 1°×1° scale. Therefore, we divided the global extent except the polar areas (80°N~60°S) into 140×360 zones. Here, for a 0.1° pixel during a specific 10-day period, if all the input data (soil moisture products and quality impact factors) have valid values, one valid data point is provided. Therefore, the maximal number of valid data points applied to train a subnetwork = 100 × the number of 10-day periods within the training period. The subnetworks with less than 100 valid data points (e.g., those in oceans) were dropped, leaving usually >15,000 zonal subnetworks included in an independent neural network. The training was performed in MATLAB 2016a, and the number of nodes in the hidden layer (between the input and output layers (Stinchcombe and White, 1989)) of each subnetwork was set to 7.

2.2.2 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 (a specific 10-day period), within each 1°×1° zone, the soil moisture values are filtered to the level of three standard deviations relative to the mean in that zone. This preprocessing step is thus called '3σ denoising' (note that the denoise is conducted spatially, rather than temporally, so that the extreme events will not be treated). After neural network operation, boundary fuzzification is first applied, that 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 is applied, as shown in Figure S1. The soil moisture data with fuzzified boundaries are transformed into both 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 data masking. Because '3σ denoising' deleted suspicious soil moisture retrievals, the simulation outputs also contain few missing values, which can be simply 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 pixel in a specific ten-day period, if ice cover is reported by ASCAT-SWI in most years, it is also supposed to be covered by snow/ice, unless the thaw state is observed in the MEaSUREs Global Record of Daily Landscape Freeze/Thaw Status V4 dataset. The simulated soil moisture in the rainforests identified in the 'ASCAT Constant' is retained but not recommended due to the high 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, , we remove all suspicious values for every simulated result. This preprocessing step is performed by first obtaining 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 considered suspicious and is not used as a training target. Subsequently, '3σ denoising' is performed again before the simulated soil moisture becomes secondary training target, 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; the temporal spans of SIM-XT and SIM-X are the same, as shown in Figure 1).

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

The 11 available microwave soil moisture data products with different temporal spans are incorporated, and 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 the training efficiency, the spatial coverage of the simulation output is sacrificed because the overlapping area decreases with the increase in the number of soil moisture products. 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 the global scale (TMI is limited to [N40°, S40°]; SMOS lacks data in Asia). To solve that dilemma,

we classified all 0.1° pixels according to the available predictor soil moisture products over a 10-day period (for example, if there are at most four soil moisture data inputs in one round, there should be 4+3+2+1=10 combinations). However, to avoid soil moisture simulation under snow or ice cover (Section 2.2.2), not all combinations are considered. Then, corresponding to each selected combination, an independent neural network is trained. For data simulation in a 0.1° pixel, the most preferable independent neural network is expected to be trained using all the available soil moisture data sources in that pixel. However, in the 1° zone where it is located, the subnetwork belonging to that preferable independent neural network may not exist due to limited valid data points (see section 2.2.1). Then, an alternative subnetwork driven by the combination of fewer soil moisture data inputs should be applied instead. Hence, we should determine the neural network collocation that is the best choice for every pixel. Apart from applicability, the relative priority order of different neural networks was obtained 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 relative accuracy of the training targets (the training target quality declines monotonically: SMAP>SIM-1T>SIM-2T>SIM-3T>SIM-4T). Sometimes, the two most likely priority orders are given, with the simulation results of the corresponding two substeps integrated later. Specifically, when the LAI data source changes, the division of a single round into several substeps is also essential. Based on these principles, five rounds of neural networks are designed as follows, with 8 substeps containing a total of 67 independent neural networks. The training period for each neural network and the simulation period for each substep are shown in Figure 1 (below the timeline), and the details are as follows:

For the first round's neural network (labeled as NN1), the potential training period is 2015D10~2018 ('D' is the ordinal of the 10-day period, so '2015D10' represents the period from April 1st to April 10th in 2015) because SMAP soil moisture data during that period are applied as the training target, while ASCAT-SWI10 (abbreviated as ASCAT), SMOS-IC (SMOS), AMSR2-JAXA and AMSR2-LPRM-X (AMSR2-LPRM) are the four soil moisture products used as predictors (details are in Tables S1~S2). Because all the four predictors have data since 2012D19, the potential soil moisture simulation period is 2012D19~2018, which is further divided into two parts: one is 2014~2018 (substep1), for which the PROBA-V LAI data that begins in 2014 are applied, whereas the other is 2012D19~2013 (substep2), for which GLASS LAI data are used (note: because

GLASS LAI covers the period from the beginning of our study period until 2017, the training period for substep 2 is 2015D10~2017). Please refer to Tables S1~S2 for details. The simulation results of the two substeps (SIM-1-1 and SIM-1-2) are combined as SIM-1 and then transformed into a secondary training target, denoted as SIM-1T. In the second round of simulation, the training target can be either SMAP or SIM-1T, while the soil moisture input data are ASCAT, SMOS, TMI-LPRM-X (TMI) and FY-3B-NSMC (FY). The simulation output, SIM-2, covers the period of 2011D20~2012D18, which is constrained by the common period of the four predictors (Table S3~S4). In the third round of neural network operation, the simulation period is 2010D16~2011D19. SMAP, SIM-1T and SIM-2T are combined and used as the training targets (the training periods are within the range of 2011D20~2017D36), while the predictor soil moisture data are ASCAT, SMOS, TMI and WindSat-LPRM-X (WINDSAT). There are two substeps in round 3 that are distinguished by whether the priority order of the neural networks is determined mainly based on the training sample quantity and the training target quality (SIM-3-1), or by first considering the number of predictor soil moisture products (SIM-3-2, Table S5~S8). Because these two methods emphasize different aspects of neural network quality, in some pixels, SIM-3-1 will be advantageous, but in others, SIM-3-2 could be better. Hence, an algorithm is devised to combine the advantages of both simulations (SIM-3), which is described in Table S9. Next, the 4th round is for simulations during 2007D01~2010D15. SIM-2T and SIM-3T are combined to be the training target, and ASCAT, WINDSAT, TMI, AMSRE-JAXA, AMSRE-LPRM-X (AMSRE-LPRM) and AMSRE-NSIDC are all applied as predictors (LAI data now come from SPOT-VGT). Two substeps are also needed. In the first substep, neural networks are sorted by paying the greatest 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 create an alternative estimate (Tables S10~S13). Afterwards, SIM-4 is obtained by reasonably integrating 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 together are the training targets, while the predictor soil moisture data entering the neural networks consist of WINDSAT, TMI, AMSRE-JAXA, AMSRE-LPRM and AMSRE-NSIDC (Table S14~S15).

2.3 Methods for the 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; Albergel et al., 2012; Dorigo et al., 2015; Fernandez-Moran et al., 2017b; Gao et al., 2020; Karthikeyan et al., 2017b; Kerr et al., 2016; Kim et al., 2015b; Kolassa et al., 2018; Lievens et al., 2017; Zhang et al., 2019). Because SMAP, the training target, is the soil moisture within 0~5 cm, the simulated soil moisture is used for that soil layer as well. Accordingly, the measurements used for validation are limited to ≤ 5 cm in depth (e.g., the Russian networks were not applicable for this reason). The quality flags of ISMN (Dorigo et al., 2013) are also checked to retain only the 'good quality' data. After data screening and processing (for example, in the case of high spatial variability in soil moisture, we excluded the pixels with average annual maximal water area fractions greater than 5%, 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 , which as a temporal resolution of 10 days. More than 90% of the stations are located in relatively flat areas with topographic complexity less than 10%. 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 3. The major climate types of the sites are determined from the Köppen-Geiger climate classification map (see Table 1 for the description (Kottek et al., 2006)).

Most ISMN networks are dense networks, as the stations are very close to each other, often within the same 0.1° pixel, whereas others are sparse networks (see Text S2 and Figure 3). In addition, various sensors are simultaneously operated at some stations. Hence, to make full use of all the high-quality records, and to reduce the problem caused by the scale difference between simulation and measurement, the site-scale 10-day averaged soil moisture data are further aggregated to a 0.1° pixel-scale by averaging all the data (different stations or different sensors) within the pixel (Gruber et al., 2020). Specifically, if soil moisture is not simulated due to snow or ice cover, the corresponding measurement is not useful. This process resulted in a final collection of ~40,000 pixel-scale 10-day period soil moisture records within the validation dataset.

The soil moisture datasets to be evaluated include the RSSSM product in this study (Remote Sensing Surface Soil Moisture, covering 2003~2018), SMAP_E (the primary training target, covering April 2015~2018), the longest record of satellite-based

soil moisture: 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, 10-day averaged and then evaluated during 2003~2018) as well as the soil moisture datasets developed by combining both satellite observations and model simulations: CCI v4.5 and GLEAM v3.3 (for v3.3a, the radiation and air temperature forcing data come from ERA5, whereas for v3.3b, all meteorological data are satellite-based, yet the data after September 2018 are not available). The overall performance of any soil moisture product is first evaluated using all of the validation datasets, with Pearson R-square (R^2) and RMSE values (unit: $m^3\ m^{-3}$) adopted as the main indicators. The next step is temporal pattern validation. For pixels with enough (>20) 10-day averaged in situ records, we compare the estimated soil moisture during all periods against the corresponding measurements, with the calculated Pearson correlation coefficient (*r*) and RMSE. Several supplementary indexes are also added, including bias, unbiased RMSE (ubRMSE) and the correlation coefficient between the anomalies (anomalies *r*, abbreviated here as 'A.R'; A.R can better indicate the simulation accuracy of interannual variations; soil moisture anomalies are calculated by Eq. 1). Next, we compare the means and medians of the above evaluation indexes for different soil moisture products and test whether the differences are significant. Moreover, the relative performances of various products in different climatic zones are analyzed. Finally, we perform spatial pattern validation. In detail, for every 10-day period, we compare all the soil moisture measurements that are upscaled to 0.1° during that period with the corresponding estimated values. The spatial pattern evaluation indexes include the correlation coefficient (*r*), RMSE, bias and ubRMSE values (Eq. 2). The relative superiority of all products during different 10-day periods in a year, and the changes in data coverage as well as data quality with time are also investigated.

$$\overline{SSM(k)} = \frac{\sum_{y=1}^{ny} SSM(y,k)}{ny} \quad (ny \geq 3)\ ;\ SSM\ is\ either\ estimated\ or\ measured$$

SSM: surface soil moisture; k: the ordinal of 10 day period in a year; y: a year with measured SSM in k^{th} 10 day period; ny: number of those years

$$SSM_{anom}(y,k) = SSM(y,k) - \overline{SSM(t)}$$

$SSM_{anom}(y,t)$: the anomalies of surface soil moisture during the t^{th} 10 day period in year y (Eq. 1)

$$\overline{SSM_{est}} = \frac{\sum_{i=1}^{ng} SSM_{est,i}}{ng}\ ;\ \overline{SSM_{act}} = \frac{\sum_{i=1}^{ng} SSM_{act,i}}{ng}\ (ng \geq 20)$$

i: a grid with upscaled surface soil moisture measurements during a specific 10 day period; ng: the number of those grids in the globe

$$ubRMSE_{spatial} = \sqrt{\sum_{i=1}^{ng}[(SSM_{est,i} - \overline{SSM_{est}}) - (SSM_{est,i} - \overline{SSM_{act}})]^2/ng} \quad \text{(Eq. 2)}$$

2.4 Methods for the intra-annual variation analysis of surface soil moisture

Because the original resolution of SMAP soil moisture is ~0.4° while that of most predictor soil moisture products is 0.25°, the intra-annual variation analysis of RSSSM is performed at 0.5° resolution. We also exclude high-latitude areas (60°N~90°N) where the available data are limited due to frequent ice cover. Fourier functions can characterize intra-annual variation well (Brooks et al., 2012; Hermance et al., 2007). Therefore, for the remaining areas (60°S~60°N), based on a total of 36×16 (years) =576 data points, we fit the intra-annual cycle of soil moisture using the Fourier function, 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 explore the relationship between the intra-annual cycles of precipitation and surface moisture using Fourier fitting (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 RSSSM indicates the average soil moisture condition during every 10-day period, we evaluate the surface soil moisture decline after 20 consecutive days (i.e., two adjacent 10-day periods) without effective precipitation to explore the impact of dry periods on surface soil moisture. Effective precipitation is calculated by precipitation minus canopy interception, which is estimated by the modified Merriam canopy interception model (Kozak et al., 2010; Merriam, 1960). 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 consider that the soil moisture change in the latter period compared to the previous period is mostly due to surface evaporation and percolation (capillary rise is negligible (McColl et al., 2017)), and thus should be negative. Hence, for a 0.5° grid, if the number of negative values does not meet two times the number of positive values, the precipitation threshold is reduced by 1 mm until that condition is satisfied. This loop

is terminated when there are less than 36 available data points in dry periods (the maximal number of data points is 576), and then the grid is 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 wet areas (e.g., rainforests), 20 days without effective precipitation seldom occurs, leading to no results over most areas. In the remaining areas, 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 training efficiency: a comparison between RSSSM and SMAP

To examine the training and simulation efficiency of the neural network, we compare the simulated surface soil moisture (RSSSM) 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 4a). If only the pixels with measured data are considered, the consistency between RSSSM and SMAP becomes even stronger, with an R^2 of 0.97 and an RMSE of 0.016 (Figure 4b). When validated against site measurements, the R^2 and RMSE values are 0.46 and 0.083, respectively, for both RSSSM and SMAP (Figure 4c and 4d). All these findings justify the high training and prediction efficiency of the neural network set designed in this study.

For temporal accuracy, according to Table 2, RSSSM is just slightly lower than SMAP (the differences in the five indicators, *r,* RMSE, bias, ubRMSE and A.R, are all nonsignificant). Figure 5 indicates generally the same level of temporal accuracy for RSSSM and SMAP under all climates. RSSSM cannot adequately characterize the temporal variation in soil moisture in the 'Dfc' (snow climate, fully humid, see Table 1) region because the training target, SMAP, does not have a high temporal accuracy in this area, probably due to frequent freezing and melting processes.

Next, we compare the spatial accuracy of RSSSM and SMAP. The spatial correlation of RSSSM is somewhat reduced compared to the training target, while the RMSE is slightly increased (Table 3), indicating a subtle loss of detailed spatial information through neural network operation. Because ISMN stations are mostly located in the middle to high latitudes of the Northern Hemisphere, Figure 6 shows that: 1) the accuracy of RSSSM is highest in summers (growing seasons) and lowest in winters, which is inherited from its origin, SMAP, probably due to the impact of freezing on soil moisture retrieval; and 2)

RSSSM has a similar spatial accuracy as SMAP in most periods, except for May to June and November to December.

3.2 The accuracy comparison between RSSSM and popular global long-term soil moisture products

3.2.1 The data quality comparison between RSSSM 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 RSSSM are 0.44 and 0.086, respectively (Figure 7), which appear to be much better than those for ASCAT-SWI (R^2=0.33, RMSE=0.100). If the data period of SMAP (2015D10~2018) is excluded, the overall R^2 and RMSE for RSSSM are 0.43 and 0.087, respectively, which are still better than those for ASCAT-SWI (R^2=0.33, RMSE=0.1). However, RSSSM overestimates soil moisture when it is low, which is a problem inherited from the SMAP product (Figure 4), and is a bit nonlinearly correlated with the measured values (Figure 7a).

According to the temporal validation results (Table 4), the evaluation indexes including *r*, RMSE, bias and ubRMSE for RSSSM are all significantly ($p<0.05$) better than those for ASCAT-SWI (anomalies *r* for RSSSM is also higher, but not significant). The temporal accuracy of RSSSM appears to be 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 are high for RSSSM (Figure 8).

The spatial accuracy of RSSSM is found significantly higher than ASCAT-SWI when any evaluation index is considered (Table 5). Moreover, the results show that RSSSM is generally superior to ASCAT-SWI throughout the year, especially during the growing seasons (Figure 9).

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

First, the overall accuracies of RSSSM and GLDAS Noah V2.1 surface soil moisture data during 2003~2018 are compared. While RSSSM is nonlinearly correlated with measured soil moisture, the relationship between GLDAS soil moisture and the measurements appears to be slightly more nonlinear, resulting in a smaller R^2 of 0.39 and higher RMSE of 0.097 for GLDAS product compared to RSSSM (R^2: 0.42; RMSE: 0.087, see Figure 10). When excluding the SMAP (training target) data period, the R^2 and RMSE for RSSSM are 0.41 and 0.089, respectively, which are also superior to those for GLDAS (R^2: 0.37; RMSE:

0.099).

The higher temporal accuracy of RSSSM than GLDAS can be justified by comparing the indicators including *r*, RMSE and ubRMSE (Table 6). The advantage of RSSSM over GLDAS could be identified in almost all climatic regions, especially the cold areas such as BWk, Dfa, Dfc, Dwc and ET (Figure 11), perhaps because the soil thawing and freezing processes are not simulated well. The spatial accuracy of RSSSM, indicated by *r*, RMSE, bias and ubRMSE, is found to be significantly higher than GLDAS as well (Table 7). The spatial correlation of RSSSM is somewhat higher than that of GLDAS during March to May and September to November, and the spatial RMSE is lower all year round except for January and February (Figure 12). ERA5-Land is a newly published reanalysis-based model product with 0.1° resolution. The overall quality validation (Figure S2) 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, only slightly lower than that of RSSSM (0.42), the RMSE for ERA5-Land, 0.123, which is much higher than that for RSSSM (0.087) during their common period. Without considering the SMAP period, the conditions are the same (the R^2 for RSSSM and ERA5-Land are 0.41 and 0.38; the RMSE values for these two products are 0.089 and 0.125, respectively). The temporal correlation indicated by *r* and A.R is somewhat higher for ERA5-Land in general (Table S17), but in most cold areas (Dfa, Dwc and ET), the opposite condition occurs (Figure S3a, S3d). The temporal ubRMSE values for RSSSM and ERA5-Land do not differ significantly, but RSSSM usually performs better in relatively arid places (Figure S3c). While the relative temporal accuracies of RSSSM and ERA5-Land are unclear, the spatial pattern of RSSSM is more accurate than that of ERA5-Land considering the significantly better spatial correlation, RMSE, bias and ubRMSE (Table S18). The considerable advantage of RSSSM over ERA5-Land exists throughout the year, especially during the growing seasons from March to November (Figure S4).

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

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 overall R^2 is only 0.31 with an RMSE value of up to 0.095 (Figure S5, when the SMAP data period is excluded, the R^2 and RMSE for CCI are 0.28 and 0.098,

compared to 0.41 and 0.089 for RSSSM). The temporal pattern of RSSSM, indicated by *r* and RMSE, is found to be significantly better than CCI (Table S19), and under all climate conditions (Figure S6). Our results indicate that RSSSM also shows a consistently higher spatial accuracy than CCI, especially during the growing seasons (Table S20 and Figure S7).

Next, we focus on the interannual change in data quality. According to Figure 13a~c, while the correlation coefficient for RSSSM does not vary significantly among different years, the RMSE and ubRMSE values in earlier periods are somewhat raised compared to those after 2012. Though the data quality of RSSSM can hardly be maintained as well, the degradation degree is much slighter than that of CCI. The comparison of the spatial coverages of the 10-day scale RSSSM and CCI data (rainforests are excluded) shows that RSSSM 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, whose data coverage before 2007 is limited (Figure 13d).

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 the measured values (Figure S8). Therefore, with an R^2 of 0.42 and an RMSE of 0.087, RSSSM is found to be superior to GLEAM v3.3a/b in general (if the SMAP data period is excluded, RSSSM's R^2 and RMSE values are 0.41 and 0.089, respectively, which are still better than both GLEAM v3.3a (R^2: 0.35; RMSE: 0.141) and GLEAM v3.3a (R^2: 0.34; RMSE: 0.128)). The temporal and spatial accuracies of GLEAM products and RSSSM are compared in Tables S21~S24. The advantage of GLEAM is its ability to characterize the temporal variations in soil moisture, with higher temporal correlation achieved in most climatic regions (Figure S9a and S9d). However, the main potential disadvantage is the obvious overestimation, which leads to significantly higher RMSE values than RSSSM in all regions and all periods (Figure S9b and Figure S10b). Moreover, the spatial pattern of GLEAM products is less convincing than that of RSSSM, considering the lower spatial correlation coefficients, especially in spring (March to May) and autumn (September to November) (Figure S10a). Therefore, the potential advantages of RSSSM can exceed those of GLEAM.

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

temporal accuracy, but their absolute values and spatial patterns are relatively unreliable, whereas RSSSM 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 RSSSM (applicable to all studies)> GLEAM (suitable for temporal variation studies)> 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^2 and RMSE are 0.2 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 improvements also to the temporal *r* and RMSE (0.69 and 0.08) as well as overall R^2 and RMSE (0.42 and 0.087) values. In addition, our 10-day period average product is both spatially and temporally continuous over 16 years, with a high spatial resolution, and covers all land except for frozen ground. Hence, our product could be more useful than previous machine learning 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 14a (the maps for separate months are shown in Figure S11a). The above validation results show that except for RSSSM, GLDAS has the highest spatial accuracy, so the spatial map of GLDAS surface moisture is attached below (Figure S11b). By comparison, the spatial patterns of RSSSM and GLDAS are similar, but some differences also exist (see the regions circled in red). Obviously, RSSSM has a higher spatial heterogeneity and probably more reflections on wetlands and irrigated fields (e.g., the Hetao Irrigation Area in China), whereas GLDAS appears patchy in arid areas. The latitudinal pattern comparison in Figure S12a also implies that RSSSM contains more detailed spatial information.

For the interannual variation, because the GLEAM v3.3a product is proven to have the best accuracy in characterizing the temporal anomalies of soil moisture, and also covers the whole world, this product is selected as the reference to justify our calculation. According to Figure S12b, both GLEAM and RSSSM support a significant rising trend in global mean surface soil moisture during 2003~2018, while the average rates are both approximately 0.03 $m^3\ m^{-3}\ yr^{-1}$ (Figure S12b). The spatial

patterns of the interannual trends in RSSSM and GLEAM are shown in Figure 14c~d, which are generally consistent. Soil moisture gains are found over the border between the 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 lies in Central Africa, the Arabian Peninsula and northwestern Canada.

Because the validation against measurements proves that the intra-annual soil moisture variation in the 'Dfc' climate region cannot be captured by SMAP or RSSSM, 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 midlatitude 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 (nongrowing seasons) and low in summers (Figure 15a 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 it is lowest in arid places (Figure 15b; Figure S13b~c). Precipitation is a direct driver of surface soil moisture changes (Figure S14a~b), and the intra-annual cycle of soil moisture often strictly follows that of precipitation as long as it exists (Figure 15c and Figure S14c). Therefore, considering that precipitation is highest in summer at 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 15d), which is expected to be related to the 'time lag' effect. On dry days, the fastest surface moisture decline is expected in summers when evapotranspiration is high. However, this study reveals that at midlatitudes, the opposite condition occurs: the surface water loss without rain is lowest in summer (Figure 15e and Figure S15a). Further analysis proves a positive correlation between surface moisture and its rate of decline, with r>0.8 over 85% of the area (Figure S15b~c), indicating that because soil moisture in the westerlies is often high in winters, the available surface water for evaporation and percolation loss is limited in summer, and plants tend 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 15f). Therefore, if sufficient water during rainy seasons is lacking there, significant water loss (Figure S15d) may destroy the local ecosystem.

4 Discussion and conclusions

In this study, an improved global long-term satellite-based surface soil moisture dataset, RSSSM, was developed mainly based on 11 microwave soil moisture products. Our product is more comparable to the in situ measurements at ISMN stations than most existing popular surface soil moisture datasets. Our product is temporally continuous during 2003~2018, and covers the whole globe except for frozen grounds (CCI has limited spatial coverage before 2007, when ASCAT data are unavailable), ensuring its applicability to global long-term studies or ecosystem modeling. However, the achieved accuracy (R^2=0.42; RMSE=0.087) for surface soil moisture is still lower than that for many other terrestrial essential climate variables. The target RMSE for surface soil moisture set by GCOS is 0.04 m^3 m^{-3}, which is much lower than the value met in this study, indicating the need to further improve the global soil moisture data quality.

Fortunately, this study provides a novel approach that has the potential to create increasingly better soil moisture products in future. The RMSE and ubRMSE values in earlier periods are somewhat higher than those after 2012, which is because: 1) five rounds of simulations were performed, with the output converted into the training target of the next round's neural networks, leading to a little error propagation as the simulation period extended to the past; and 2) the quality of microwave soil moisture data is generally lower in earlier periods due to the relatively unadvanced microwave sensors with low signal-to-noise ratio (SNR). However, due to the elaborate design of the neural network set (localized networks, full use of 11 microwave soil moisture products, the determination of quality impact factors and the organization of 67 independent neural networks), high training efficiency is achieved, resulting in limited amplification of noise and high maintenance of valid information during 16 years of simulation. This method turns out to be better than the simple CDF matching algorithm, which may not efficiently calibrate the low-quality soil moisture data retrieved from earlier sensors. The overall data accuracy of RSSSM is only slightly lower than that of SMAP, the primary training target. Therefore, if microwave sensors with higher SNR or better penetration

of vegetation canopy than SMAP are launched in the future (for example, the upcoming P-band microwave sensors (Etminan et al., 2020; Ye et al., 2020)), we can develop a temporally continuous soil moisture dataset beginning in 2003 by using the soil moisture or Tb retrieved from the new sensors as the reference. This upcoming product is expected to have even higher accuracy than the SMAP product (we will update the complete RSSSM product then). In that sense, the data fusion algorithm proposed here will be very meaningful in the future.

Another way to improve global surface soil moisture data accuracy as well as the temporal resolution 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 reanalysis-based models have advantages in characterizing temporal variations, and even on a daily scale, except for irrigated croplands. Furthermore, because root-zone soil moisture is the direct factor influencing vegetation growth, it often plays a more important role than surface moisture in ecosystems; however, this factor 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 incorporated only very limited observational information. We propose that one possible approach is to use the pixel-specific confidence range and the spatial pattern of satellite-based soil moisture (e.g., our product: RSSSM) to constrain the model parameters or add supplementary modules if necessary. In detail, RSSSM 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 RSSSM. In addition, whether the simulated value falls within the confidence range (e.g., ±20%) of that reported by RSSSM should also be tested. By recurrent adjustments, the model parameters in each pixel can be optimized. For irrigated croplands, if irrigation is not considered in the models, the simulated surface soil moisture will soon fall below the confidence range, and the spatial correlation will also decline regardless of the parameters that are provided. Therefore, a well-designed irrigation module (Chen et al., 2019) should be introduced. 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, RSSSM, is available at: 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 corresponding author, Xiaoming Feng, supervised the research and revised the draft. Bojie Fu administered the project and funded the research. All co-authors reviewed the manuscript and contributed to the writing process.

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.

Acknowledgments

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 Sens. Environ., 224, 289-303, https://doi.org/10.1016/j.rse.2019.02.008, 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 Sens. Environ., 180, 453-464, https://doi.org/10.1016/j.rse.2015.11.022, 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 Sens. Environ., 118, 215-226, https://doi.org/10.1016/j.rse.2011.11.017, 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, Hydrol. Earth Syst. Sci., 12, 1323-1337, http://doi.org/10.5194/hess-12-1323-2008, 2008

Baldocchi, D., Tang, J., and Xu, L.: How switches and lags in biophysical regulators affect spatial-temporal variation of soil respiration in an oak-grass savanna, Journal of Geophysical Research: Biogeosciences, 111, https://doi.org/10.1029/2005JG000063, 2006

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, Hydrol. Earth Syst. Sci., 19, 389-407, http://doi.org/10.5194/hess-19-389-2015, 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 Sens. Environ., 137, 299-309, https://doi.org/10.1016/j.rse.2012.12.027, 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), Geophys. Res. Lett., 34, http://doi.org/10.1029/2007GL031088, 2007

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, J. Geophys. Res.-Atmos., 121, 2658-2678, http://doi.org/10.1002/2015JD024131, 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, Bull. Amer. Meteorol. Soc., 95, 1431-1443, http://doi.org/10.1175/BAMS-D-13-00047.1, 2014

Brooks, E. B., Thomas, V. A., Wynne, R. H., and Coulston, J. W.: Fitting the Multitemporal Curve: A Fourier Series Approach to the Missing Data Problem in Remote Sensing Analysis, IEEE Trans. Geosci. Remote Sensing, 50, 3340-3353, http://doi.org/10.1109/TGRS.2012.2183137, 2012

Burgin, M., Colliander, A., Njoku, E., Chan, S., Cabot, F., Kerr, H. Y., Bindlish, R., Jackson, T., Entekhabi, D., and Yueh, S.:

A Comparative Study of the SMAP Passive Soil Moisture Product With Existing Satellite-Based Soil Moisture Products, IEEE Trans. Geosci. Remote Sensing, 55, 2959-2971, http://doi.org/10.1109/TGRS.2017.2656859, 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 Sens. Environ., 137, 310-329, https://doi.org/10.1016/j.rse.2013.02.030, 2013

Cammalleri, C., Verger, A., Lacaze, R., and Vogt, J. V.: Harmonization of GEOV2 fAPAR time series through MODIS data for global drought monitoring, Int. J. Appl. Earth Obs. Geoinf., 80, 1-12, https://doi.org/10.1016/j.jag.2019.03.017, 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., Crow, W. T., and Kerr, Y.: Development and assessment of the SMAP enhanced passive soil moisture product, Remote Sens. Environ., 204, 931-941, https://doi.org/10.1016/j.rse.2017.08.025, 2018

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 Sens. Environ., 214, 1-13, https://doi.org/10.1016/j.rse.2018.05.008, 2018

Chen, Y.: A new dataset of satellite observation-based global surface soil moisture covering 2003-2018, 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 Resour. Res., 55, 3708-3738, http://doi.org/10.1029/2018WR023573, 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, J. Geophys. Res.-Atmos., 118, 4466-4475, http://doi.org/10.1002/jgrd.50301, 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, 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 Sens. Environ., 191, 215-231, https://doi.org/10.1016/j.rse.2017.01.021, 2017

Crow, W. T., Wagner, W., and Naeimi, V.: The Impact of Radar Incidence Angle on Soil-Moisture-Retrieval Skill, IEEE Geosci. Remote Sens. Lett., 7, 501-505, https://doi.org/10.1109/LGRS.2010.2040134, 2010

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

Dobson, M. C., Ulaby, F. T., Hallikainen, M. T., and El-rayes, M. A.: Microwave Dielectric Behavior of Wet Soil-Part II: Dielectric Mixing Models, IEEE Trans. Geosci. Remote Sensing, GE-23, 35-46, https://doi.org/10.1109/TGRS.1985.289498, 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, Geophys. Res. Lett., 39, https://doi.org/10.1029/2012GL052988, 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 Sens. Environ., 203, 185-215, https://doi.org/10.1016/j.rse.2017.07.001, 2017

Dorigo, W. A., Gruber, A., De Jeu, R. A. M., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R. M., and Kidd, R.: Evaluation of the ESA CCI soil moisture product using ground-based observations, Remote Sens. Environ., 162, 380-395, https://doi.org/10.1016/j.rse.2014.07.023, 2015

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, Hydrol. Earth Syst. Sci., 14, 2605-2616, https://doi.org/10.5194/hess-14-2605-2010, 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, Hydrol. Earth Syst. Sci., 15, 1675-1698, https://doi.org/10.5194/hess-15-1675-2011, 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 J., 12, https://doi.org/10.2136/vzj2012.0097, 2013

Draper, C. S., Reichle, R. H., De Lannoy, G. J. M., and Liu, Q.: Assimilation of passive and active microwave soil moisture retrievals, Geophys. Res. Lett., 39, http://doi.org/10.1029/2011GL050655, 2012

Du, Y., Ulaby, F. T., and Dobson, M. C.: Sensitivity to soil moisture by active and passive microwave sensors, IEEE Trans. Geosci. Remote Sensing, 38, 105-114, https://doi.org/10.1109/36.823905, 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, Proc. IEEE, 98, 704-716, https://doi.org/10.1109/JPROC.2010.2043918, 2010

Etminan, A., Tabatabaeenejad, A., and Moghaddam, M.: Retrieving Root-Zone Soil Moisture Profile From P-Band Radar via Hybrid Global and Local Optimization, IEEE Trans. Geosci. Remote Sensing, doi: https://doi.org/10.1109/TGRS.2020.2965569, 2020. 1-9, https://doi.org/10.1109/TGRS.2020.2965569, 2020

Fan, X., Liu, Y., Gan, G., and Wu, G.: SMAP underestimates soil moisture in vegetation-disturbed areas primarily as a result of biased surface temperature data, Remote Sens. Environ., 247, 111914, https://doi.org/10.1016/j.rse.2020.111914, 2020

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 Sens. Environ., 196, 253-264, https://doi.org/10.1016/j.rse.2017.05.012, 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 Sens., 9, https://doi.org/10.3390/rs9050457, 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, Int. J. Appl. Earth Obs. Geoinf., 62, 27-38, https://doi.org/10.1016/j.jag.2017.05.013, 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, J. Meteorol. Soc. Japan, 29, 282-292, https://doi.org/10.11440/rssj.29.282, 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 Trans. Geosci. Remote Sensing, 42, 2347-2361, https://doi.org/10.1109/TGRS.2004.836867, 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, J. Hydrometeorol., 7, 23-38, https://doi.org/10.1175/JHM473.1, 2006

Gao, L., Sadeghi, M., and Ebtehaj, A.: Microwave retrievals of soil moisture and vegetation optical depth with improved resolution using a combined constrained inversion algorithm: Application for SMAP satellite, Remote Sens. Environ., 239, 111662, https://doi.org/10.1016/j.rse.2020.111662, 2020

Griend, A. A. V. d. and Wigneron, J.: On the measurement of microwave vegetation properties: some guidelines for a protocol, IEEE Trans. Geosci. Remote Sensing, 42, 2277-2289, https://doi.org/10.1109/TGRS.2004.832243, 2004

Gruber, A., De Lannoy, G., Albergel, C., Al-Yaari, A., Brocca, L., Calvet, J. C., Colliander, A., Cosh, M., Crow, W., Dorigo, W., Draper, C., Hirschi, M., Kerr, Y., Konings, A., Lahoz, W., McColl, K., Montzka, C., Muñoz-Sabater, J., Peng, J., Reichle, R., Richaume, P., Rüdiger, C., Scanlon, T., van der Schalie, R., Wigneron, J. P., and Wagner, W.: Validation practices for satellite soil moisture retrievals: What are (the) errors?, Remote Sens. Environ., 244, 111806, https://doi.org/10.1016/j.rse.2020.111806, 2020

Gruber, A., Dorigo, W. A., Crow, W., and Wagner, W.: Triple Collocation-Based Merging of Satellite Soil Moisture Retrievals, IEEE Trans. Geosci. Remote Sensing, 55, 6780-6792, https://doi.org/10.1109/TGRS.2017.2734070, 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 Syst. Sci. Data, 11, 717-739, https://doi.org/10.5194/essd-11-717-2019, 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, J. Geophys. Res.-Atmos., 124, 1474-1495, https://doi.org/10.1029/2018JD029304, 2019

Hansen, M. and Song, X. P.: Vegetation Continuous Fields (VCF) Yearly Global 0.05 Deg, NASA EOSDIS Land Processes DAAC, https://doi.org/10.5067/MEaSUREs/VCF/VCF5KYR.001, 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, https://doi.org/10.5194/bg-12-4121-2015, 2015

Hermance, J. F., Jacob, R. W., Bradley, B. A., and Mustard, J. F.: Extracting Phenological Signals From Multiyear AVHRR NDVI Time Series: Framework for Applying High-Order Annual Splines With Roughness Damping, IEEE Trans. Geosci. Remote Sensing, 45, 3264-3276, http://doi.org/10.1109/TGRS.2007.903044, 2007

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, Atmos. Chem. Phys., 19, 3097-3124, https://doi.org/10.5194/acp-19-3097-2019, 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, J. Geophys. Res.-Atmos., 114, https://doi.org/10.1029/2008JD010257, 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 Sens. Environ., 83, 195-213, https://doi.org/10.1016/S0034-4257(02)00096-2, 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 Sens. Environ., 36, 203-212, https://doi.org/10.1016/0034-4257(91)90057-D, 1991

Jackson, T. J., Schmugge, T. J., and Wang, J. R.: Passive microwave sensing of soil moisture under vegetation canopies, Water Resour. Res., 18, 1137-1142, https://doi.org/10.1029/WR018i004p01137, 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 Sens., 8, https://doi.org/10.3390/rs8020156, 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, Adv. Water Resour., 109, 106-120, https://doi.org/10.1016/j.advwatres.2017.09.006, 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, Adv. Water Resour., 109, 236-252, https://doi.org/10.1016/j.advwatres.2017.09.010, 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 Trans. Geosci. Remote Sensing, 41, 184-194, https://doi.org/10.1109/TGRS.2002.808331, 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 Sens. Environ., 180, 40-63, https://doi.org/10.1016/j.rse.2016.02.042, 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 Trans. Geosci. Remote Sensing, 39, 1729-1735, https://doi.org/10.1109/36.942551, 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 Sens. Environ., 204, 260-275, https://doi.org/10.1016/j.rse.2017.10.026, 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 Sens. Environ., 161, 43-62, https://doi.org/10.1016/j.rse.2015.02.002, 2015a

Kim, S., Parinussa, R. M., Liu, Y. Y., Johnson, F. M., and Sharma, A.: A framework for combining multiple soil moisture retrievals based on maximizing temporal correlation, Geophys. Res. Lett., 42, 6662-6670, http://doi.org/10.1002/2015GL064981, 2015b

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, https://doi.org/10.2208/prohe.48.217, 2004

Kolassa, J., Reichle, R. H., Liu, Q., Alemohammad, S. H., Gentine, P., Aida, K., Asanuma, J., Bircher, S., Caldwell, T., Colliander, A., Cosh, M., Holifield Collins, C., Jackson, T. J., Martínez-Fernández, J., McNairn, H., Pacheco, A., Thibeault, M., and Walker, J. P.: Estimating surface soil moisture from SMAP observations using a Neural Network technique, Remote Sens. Environ., 204, 43-59, https://doi.org/10.1016/j.rse.2017.10.045, 2018

Konings, A. G., Piles, M., Das, N., and Entekhabi, D.: L-band vegetation optical depth and effective scattering albedo estimation from SMAP, Remote Sens. Environ., 198, 460-470, https://doi.org/10.1016/j.rse.2017.06.037, 2017

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World Map of the Köppen-Geiger climate classification updated, Meteorol. Z., 15, 259-263, https://doi.org/10.1127/0941-2948/2006/0130, 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, Hydrol. Process., 21, 229-241, https://doi.org/10.1002/hyp.6235, 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, Hydrol. Earth Syst. Sci., 19, 4463-4478, https://doi.org/10.5194/hess-19-4463-2015, 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 Sens., 1, https://doi.org/10.3390/rs1020080, 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 Sens. Environ., 205, 85-99, https://doi.org/10.1016/j.rse.2017.11.002, 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 Trans. Geosci. Remote Sensing, 52,

1562-1571, https://doi.org/10.1109/TGRS.2013.2252468, 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 Sens. Environ., 189, 194-210, https://doi.org/10.1016/j.rse.2016.11.022, 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, Geophys. Res. Lett., 38, https://doi.org/10.1029/2011GL048684, 2011a

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 Sens. Environ., 123, 280-297, https://doi.org/10.1016/j.rse.2012.03.014, 2012

Liu, Y. Y., Parinussa, R. M., Dorigo, W. A., De Jeu, R. A. M., Wagner, W., van Dijk, A. I. J. M., McCabe, M. F., and Evans, J. P.: Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals, Hydrol. Earth Syst. Sci., 15, 425-436, http://doi.org/10.5194/hess-15-425-2011, 2011b

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, https://doi.org/10.1109/IGARSS.2015.7325932.

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 Sens. Environ., 231, 111215, https://doi.org/10.1016/j.rse.2019.111215, 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, Geosci. Model Dev., 10, 1903-1925, https://doi.org/10.5194/gmd-10-1903-2017, 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 Sens. Environ., 177, 277-286, https://doi.org/10.1016/j.rse.2016.02.064, 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, Nat. Geosci., 10, 100, https://doi.org/10.1038/ngeo2868, 2017

Méndez-Barroso, L. A., Vivoni, E. R., Watts, C. J., and Rodríguez, J. C.: Seasonal and interannual relations between precipitation, surface soil moisture and vegetation dynamics in the North American monsoon region, J. Hydrol., 377, 59-70, https://doi.org/10.1016/j.jhydrol.2009.08.009, 2009

Merriam, R. A.: A note on the interception loss equation, Int. J. Digit. Earth, 65, 3850-3851, https://doi.org/10.1029/jz065i011p03850, 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, Hydrol. Earth Syst. Sci., 15, 453-469, https://doi.org/10.5194/hess-15-453-2011, 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 Sens. Environ., 144, 197-213, https://doi.org/10.1016/j.rse.2014.01.013, 2014

Naithani, K. J., Baldwin, D. C., Gaines, K. P., Lin, H., and Eissenstat, D. M.: Spatial Distribution of Tree Species Governs the Spatio-Temporal Interaction of Leaf Area Index and Soil Moisture across a Forested Landscape, PLOS ONE, 8, e58704, https://doi.org/10.1371/journal.pone.0058704, 2013

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, https://doi.org/10.1109/IGARSS.2011.6049703.

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 Trans. Geosci. Remote Sensing, 43, 938-947, https://doi.org/10.1109/TGRS.2004.837507, 2005

Njoku, E. G. and Chan, T. K.: Vegetation and surface roughness effects on AMSR-E land observations, Remote Sens. Environ., 100, 190-199, https://doi.org/10.1016/j.rse.2005.10.017, 2006

Njoku, E. G., Jackson, T. J., Lakshmi, V., Chan, T. K., and Nghiem, S. V.: Soil moisture retrieval from AMSR-E, IEEE Trans. Geosci. Remote Sensing, 41, 215-229, https://doi.org/10.1109/TGRS.2002.808243, 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 Trans. Geosci. Remote Sensing, 50, 1427-1439, https://doi.org/10.1109/TGRS.2012.2182775, 2012

Owe, M., de Jeu, R., and Holmes, T.: Multisensor historical climatology of satellite-derived global land surface moisture, J. Geophys. Res.-Earth Surf., 113, https://doi.org/10.1029/2007JF000769, 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 Trans. Geosci. Remote Sensing, 39, 1643-1654, https://doi.org/10.1109/36.942542, 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 Sens. Environ., 113, 435-444, https://doi.org/10.1016/j.rse.2008.10.010, 2009

Parinussa, R. M., Holmes, T. R. H., and Jeu, R. A. M. d.: Soil Moisture Retrievals From the WindSat Spaceborne Polarimetric Microwave Radiometer, IEEE Trans. Geosci. Remote Sensing, 50, 2683-2694, https://doi.org/10.1109/TGRS.2011.2174643, 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, J. Hydrometeorol., 16, 932-947, https://doi.org/10.1175/JHM-D-13-0200.1, 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 detection from passive microwave observations, Hydrol. Earth Syst. Sci., 15, 3135-3151, https://doi.org/10.5194/hess-15-3135-2011, 2011

Piles, M., Schalie, R. v. d., Gruber, A., Muñoz-Marí, J., Camps-Valls, G., Mateo-Sanchis, A., Dorigo, W., and Jeu, R. d.: Global Estimation of Soil Moisture Persistence with L and C-Band Microwave Sensors, 22-27 July 2018 2018, 8259-8262, https://doi.org/10.1109/IGARSS.2018.8518161.

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, Int. J. Appl. Earth Obs. Geoinf., 48, 17-27, https://doi.org/10.1016/j.jag.2015.11.012, 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 Soil Moisture Product Using Random Forest Adopting AMSR-E/AMSR2 Brightness Temperature and SMAP over the Qinghai–Tibet Plateau, China, Remote Sens., 11, https://doi.org/10.3390/rs11060683, 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, Bull. Amer. Meteorol. Soc., 85, 381-394, https://doi.org/10.1175/BAMS-85-3-381, 2004

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 Sens., 8, https://doi.org/10.3390/rs8110959, 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 Trans. Geosci. Remote Sensing, 53, 5991-6007, https://doi.org/10.1109/TGRS.2015.2430845, 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, Nat. Clim. Chang., 8, 421-426, https://doi.org/10.1038/s41558-018-0138-5, 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 Sens., 7, https://doi.org/10.3390/rs71215843, 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 Sens. Environ., 112, 4285-4300, https://doi.org/10.1016/j.rse.2008.07.015, 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 Trans. Geosci. Remote Sensing, 56, 6296-6305, https://doi.org/10.1109/TGRS.2018.2835316, 2018

Stinchcombe and White: Universal approximation using feedforward networks with non-sigmoid hidden layer activation functions, 1989 1989, 613-617 vol.611, http://doi.org/10.1109/IJCNN.1989.118640.

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, Nat. Geosci., 12, 264-270, https://doi.org/10.1038/s41561-019-0318-6, 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 Trans. Geosci. Electronics, 16, 286-295, https://doi.org/10.1109/TGE.1978.294586, 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 Effect of Three Different Data Fusion Approaches on the Quality of Soil Moisture Retrievals from Multiple Passive

Microwave Sensors, Remote Sens., 10, https://doi.org/10.3390/rs10010107, 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 Sens. Environ., 189, 180-193, https://doi.org/10.1016/j.rse.2016.11.026, 2017

Verger, A., Baret, F., and Weiss, M.: Near Real-Time Vegetation Monitoring at Global Scale, IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens., 7, 3473-3481, https://doi.org/10.1109/JSTARS.2014.2328632, 2014

Wagner, W., Lemoine, G., and Rott, H.: A Method for Estimating Soil Moisture from ERS Scatterometer and Soil Data, Remote Sens. Environ., 70, 191-207, https://doi.org/10.1016/S0034-4257(99)00036-X, 1999

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 Geosci. Remote Sens. Lett., 1, 277-281, https://doi.org/10.1109/LGRS.2004.834594, 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 Sens. Environ., 107, 639-655, https://doi.org/10.1016/j.rse.2006.10.014, 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 Trans. Geosci. Remote Sensing, 52, 209-223, https://doi.org/10.1109/TGRS.2013.2237780, 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 Trans. Geosci. Remote Sensing, 54, 5301-5318, https://doi.org/10.1109/TGRS.2016.2560522, 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 Trans. Geosci. Remote Sensing, 49, 4552-4560, https://doi.org/10.1109/TGRS.2011.2148200, 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, Int. J. Digit. Earth, 5, 251-265, https://doi.org/10.1080/17538947.2012.658666, 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, https://doi.org/10.1109/IGARSS.2019.8898152, 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 Sens., 9, https://doi.org/10.3390/rs9010035, 2017

Ye, N., Walker, J. P., Guerschman, J., Ryu, D., and Gurney, R. J.: Standing water effect on soil moisture retrieval from L-band passive microwave observations, Remote Sens. Environ., 169, 232-242, https://doi.org/10.1016/j.rse.2015.08.013, 2015

Ye, N., Walker, J. P., Yeo, I., Jackson, T. J., Kerr, Y., Kim, E., McGrath, A., PopStefanija, I., Goodberlet, M., and Hills, J.: Toward P-Band Passive Microwave Sensing of Soil Moisture, IEEE Geosci. Remote Sens. Lett., 169, 232-242,

https://doi.org/10.1109/LGRS.2020.2976204, 2020

Yilmaz, M. T., Hunt, E. R., and Jackson, T. J.: Remote sensing of vegetation water content from equivalent water thickness using satellite imagery, Remote Sens. Environ., 112, 2514-2522, https://doi.org/10.1016/j.rse.2007.11.014, 2008

Zhang, R., Kim, S., and Sharma, A.: A comprehensive validation of the SMAP Enhanced Level-3 Soil Moisture product using ground measurements over varied climates and landscapes, Remote Sens. Environ., 223, 82-94, https://doi.org/10.1016/j.rse.2019.01.015, 2019

Zhao, Y., Peth, S., Wang, X. Y., Lin, H., and Horn, R.: Controls of surface soil moisture spatial patterns and their temporal stability in a semi-arid steppe, Hydrol. Process., 24, 2507-2519, https://doi.org/10.1002/hyp.7665, 2010

970 **Tables**

Table 1: 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
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

Table 2: The mean and median values of the five evaluation indexes (correlation coefficient: r, RMSE, bias, unbiased RMSE (ubRMSE), and the anomalies r (A.R)) on the temporal accuracy of the surface soil moisture simulated in this study (RSSSM) and the other surface soil moisture products, when validated using the ISMN in-situ measurements. Note: 1) for the comparison of RSSSM against the SMAP_E (SMAP) product, the validation period is from April 2015 to 2018; 2) for the comparison between RSSSM and ASCAT-SWI (ASCAT), the period is 2007~2018; 3) the comparison period for RSSSM and GLDAS Noah v2.1 (GLDAS), ERA5-Land (ERA5-L), CCI or GLEAM v3.3a (GLE-a) surface soil moisture product are 2003~2018; 4) the common period for RSSSM and GLEAM v3.3b (GLE-b) is from 2003 to September 2018.

Index	*r*		RMSE		bias		ubRMSE		A.R	
Product	RSSSM	SMAP	RSSSM	SMAP	RSSSM	SMAP	RSSSM	SMAP	RSSSM	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
Product	RSSSM	ASCAT	RSSSM	ASCAT	RSSSM	ASCAT	RSSSM	ASCAT	RSSSM	ASCAT
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
Product	RSSSM	GLDAS	RSSSM	GLDAS	RSSSM	GLDAS	RSSSM	GLDAS	RSSSM	GLDAS
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
Product	RSSSM	ERA5-L	RSSSM	ERA5-L	RSSSM	ERA5-L	RSSSM	ERA5-L	RSSSM	ERA5-L
Mean	0.689	0.734	0.080	0.112	0.001	0.082	0.047	0.050	0.620	0.648
Median	0.737	0.758	0.075	0.094	-0.002	0.073	0.048	0.049	0.661	0.672
Product	RSSSM	CCI	RSSSM	CCI	RSSSM	CCI	RSSSM	CCI	RSSSM	CCI
Mean	0.690	0.642	0.080	0.091	0.002	-0.002	0.047	0.049	0.620	0.530
Median	0.735	0.666	0.074	0.080	-0.002	0.006	0.049	0.047	0.658	0.552
Product	RSSSM	GLE-a	RSSSM	GLE-a	RSSSM	GLE-a	RSSSM	GLE-a	RSSSM	GLE-a
Mean	0.689	0.735	0.080	0.126	0.001	0.093	0.047	0.047	0.620	0.681
Median	0.737	0.771	0.075	0.119	-0.002	0.104	0.048	0.046	0.661	0.715
Product	RSSSM	GLE-b	RSSSM	GLE-b	RSSSM	GLE-b	RSSSM	GLE-b	RSSSM	GLE-b
Mean	0.688	0.729	0.080	0.117	0.001	0.077	0.047	0.046	0.618	0.670
Median	0.730	0.762	0.075	0.112	-0.002	0.091	0.048	0.045	0.659	0.705

Table 3: The mean and median values of the four evaluation indexes (r, RMSE, bias and ubRMSE) on the spatial pattern accuracy of RSSSM and the other global long-term surface soil moisture products (SMAP_E, ASCAT-SWI, GLDAS Noah v2.1, ERA5-Land, CCI, GLEAM v3.3a and GLEAM v3.3b) in every 10-day period. For each pair of comparisons, the evaluation indexes are for the common period of the two products, which is the same as Table 2. The abbreviations for the products are also the same as those in Table 2.

Index	*r*		RMSE		bias		ubRMSE	
Product	RSSSM	SMAP	RSSSM	SMAP	RSSSM	SMAP	RSSSM	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
Product	RSSSM	ASCAT	RSSSM	ASCAT	RSSSM	ASCAT	RSSSM	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
Product	RSSSM	GLDAS	RSSSM	GLDAS	RSSSM	GLDAS	RSSSM	GLDAS
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
Product	RSSSM	ERA5-L	RSSSM	ERA5-L	RSSSM	ERA5-L	RSSSM	ERA5-L
Mean	0.616	0.575	0.090	0.125	-0.005	0.077	0.086	0.095
Median	0.641	0.633	0.089	0.125	0.001	0.082	0.086	0.092
Product	RSSSM	CCI	RSSSM	CCI	RSSSM	CCI	RSSSM	CCI
Mean	0.618	0.497	0.090	0.099	-0.004	0.003	0.086	0.093
Median	0.647	0.554	0.089	0.098	0.002	0.006	0.086	0.093
Product	RSSSM	GLE-a	RSSSM	GLE-a	RSSSM	GLE-a	RSSSM	GLE-a
Mean	0.617	0.576	0.090	0.139	-0.005	0.105	0.086	0.089
Median	0.643	0.616	0.089	0.142	0.001	0.112	0.086	0.088
Product	RSSSM	GLE-b	RSSSM	GLE-b	RSSSM	GLE-b	RSSSM	GLE-b
Mean	0.616	0.560	0.090	0.128	-0.005	0.088	0.086	0.090
Median	0.643	0.613	0.089	0.130	0.001	0.094	0.086	0.089

Figures

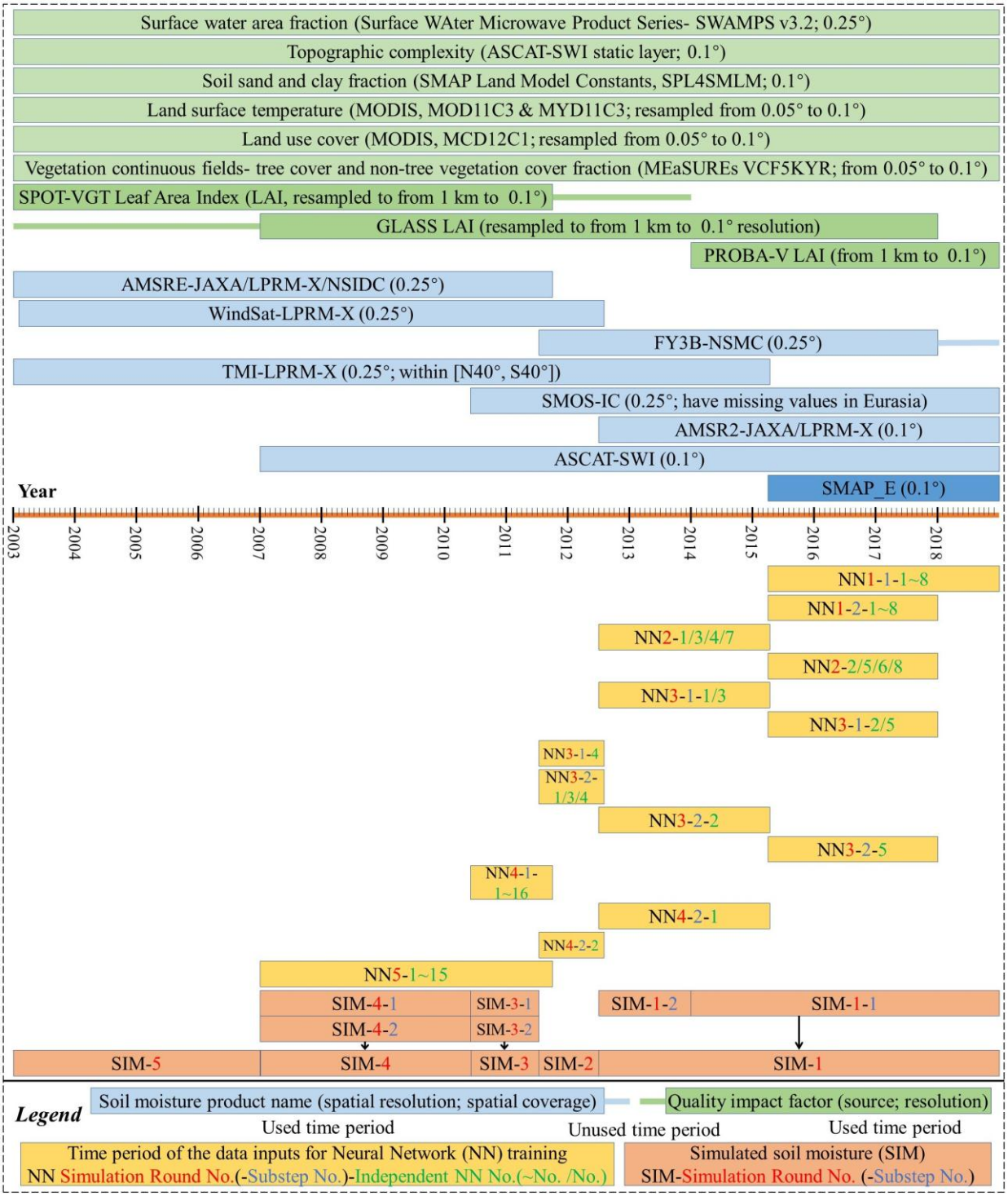

Surface water area fraction (Surface WAter Microwave Product Series- SWAMPS v3.2; 0.25°)
Topographic complexity (ASCAT-SWI static layer; 0.1°)
Soil sand and clay fraction (SMAP Land Model Constants, SPL4SMLM; 0.1°)
Land surface temperature (MODIS, MOD11C3 & MYD11C3; resampled from 0.05° to 0.1°)
Land use cover (MODIS, MCD12C1; resampled from 0.05° to 0.1°)
Vegetation continuous fields- tree cover and non-tree vegetation cover fraction (MEaSUREs VCF5KYR; from 0.05° to 0.1°)
SPOT-VGT Leaf Area Index (LAI, resampled to from 1 km to 0.1°)
GLASS LAI (resampled to from 1 km to 0.1° resolution)
PROBA-V LAI (from 1 km to 0.1°)
AMSRE-JAXA/LPRM-X/NSIDC (0.25°)
WindSat-LPRM-X (0.25°)
FY3B-NSMC (0.25°)
TMI-LPRM-X (0.25°; within [N40°, S40°])
SMOS-IC (0.25°; have missing values in Eurasia)
AMSR2-JAXA/LPRM-X (0.1°)
ASCAT-SWI (0.1°)
SMAP_E (0.1°)
Year
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
NN1-1-1~8
NN1-2-1~8
NN2-1/3/4/7
NN2-2/5/6/8
NN3-1-1/3
NN3-1-2/5
NN3-1-4
NN3-2-1/3/4
NN3-2-2
NN3-2-5
NN4-1-1~16
NN4-2-1
NN4-2-2
NN5-1~15
SIM-4-1
SIM-3-1
SIM-1-2
SIM-1-1
SIM-4-2
SIM-3-2
SIM-5
SIM-4
SIM-3
SIM-2
SIM-1
Legend
Soil moisture product name (spatial resolution; spatial coverage)
Quality impact factor (source; resolution)
Used time period
Unused time period
Used time period
Time period of the data inputs for Neural Network (NN) training
NN Simulation Round No.(-Substep No.)-Independent NN No.(~No. /No.)
Simulated soil moisture (SIM)
SIM-Simulation Round No. (-Substep No.)

Figure 1: The timeline figure showing the time periods of the soil moisture datasets and the 'quality impact factor' products (e.g., LAI dataset) used in this study (listed above the timeline), as well as the periods of data applied for the training of the 67 independent neural networks and the neural network simulation outputs (i.e., simulated soil moisture) of eight substeps (listed below the timeline).

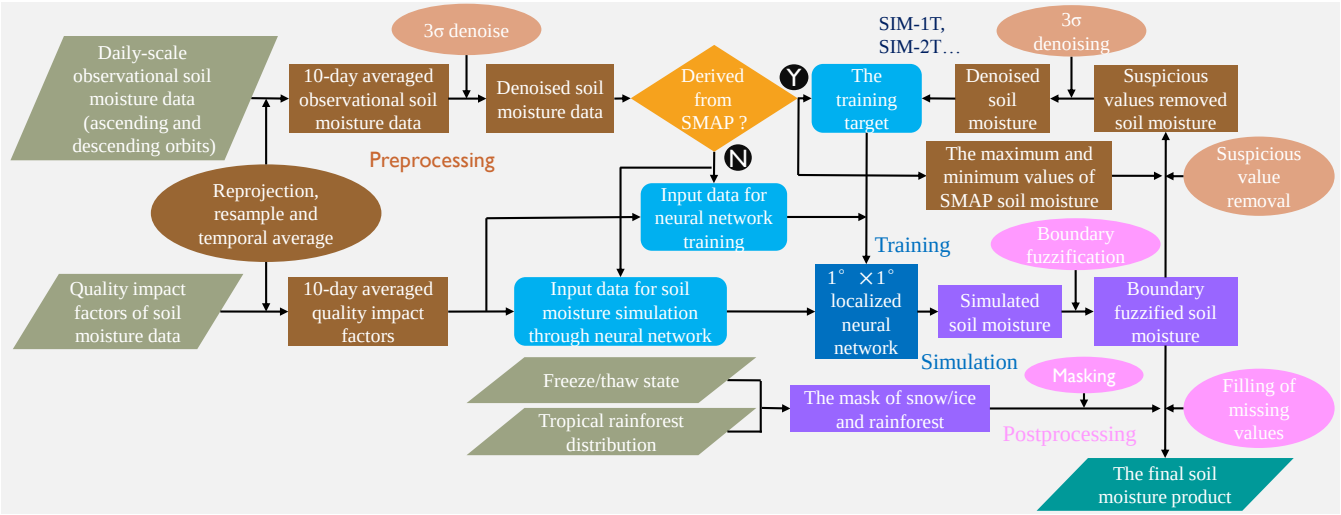


Figure 2: The flow chart for the production of global surface soil moisture data (RSSSM).

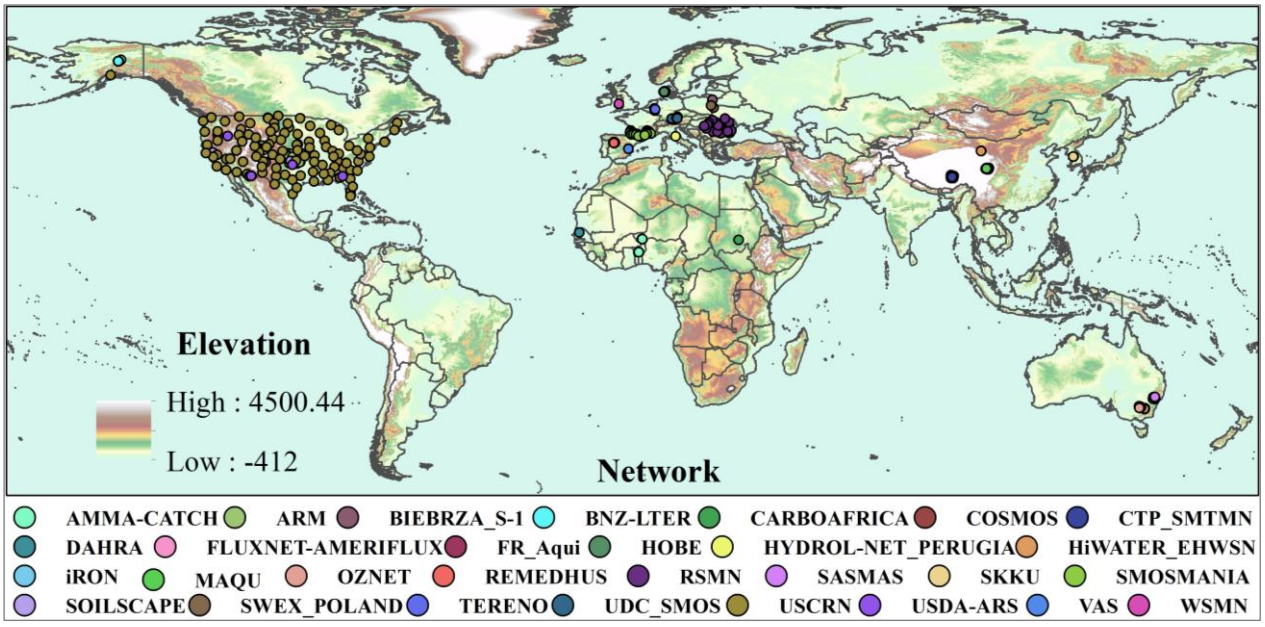


Figure 3. The global distribution of ISMN networks and stations.

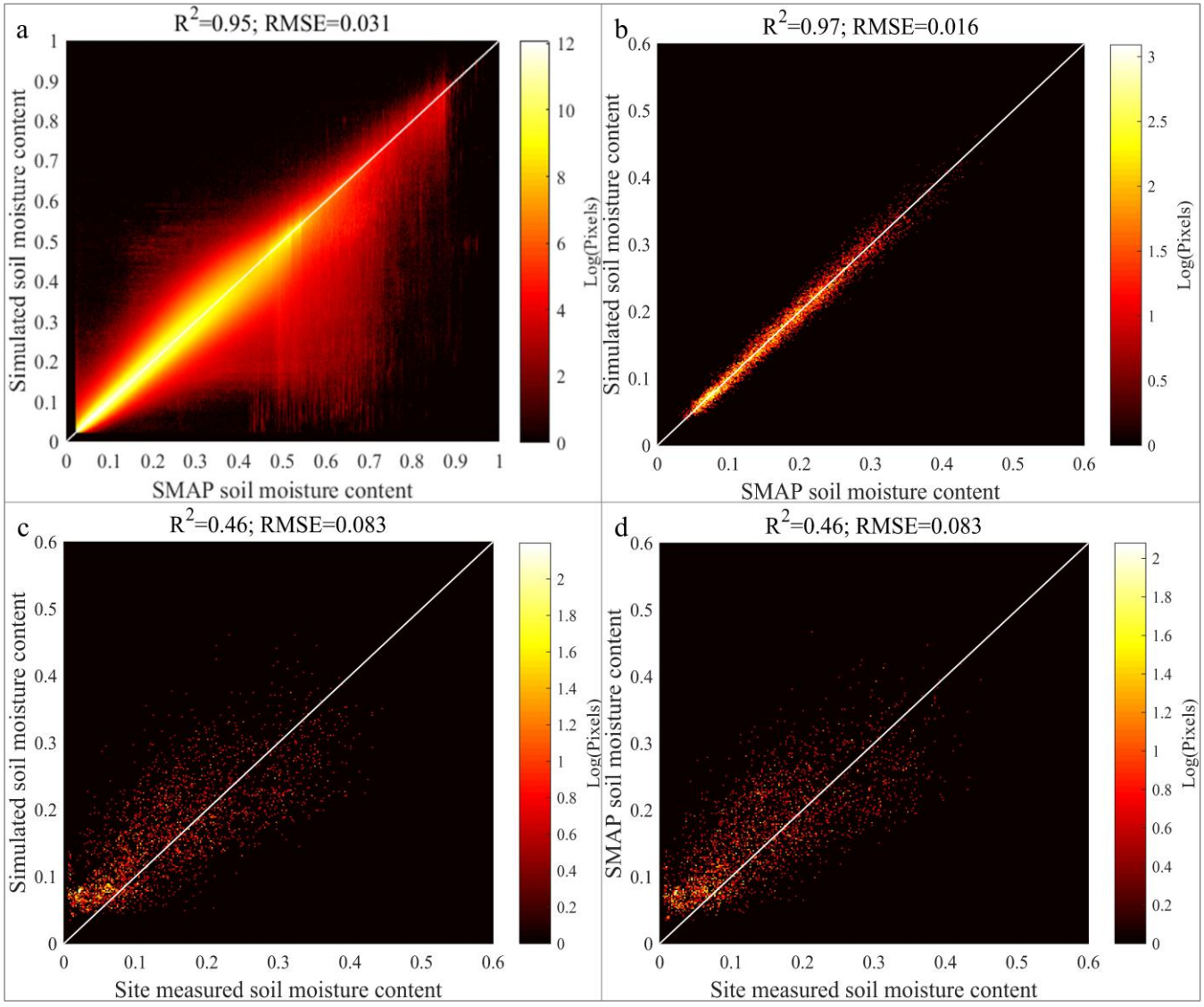


Figure 4: Comparison between the neural network simulated surface soil moisture (RSSSM) and SMAP data. The scatter plots are between: (a) RSSSM and SMAP values at all pixels; b) RSSSM and SMAP values at only the pixels with measurements; (c) RSSSM and the site measured soil moisture; and (d) SMAP and the site measurements during April 2015~2018. All plots are represented as the point density on a logarithmic scale, while the units for soil moisture content and RMSE values are m^3 m^{-3}.

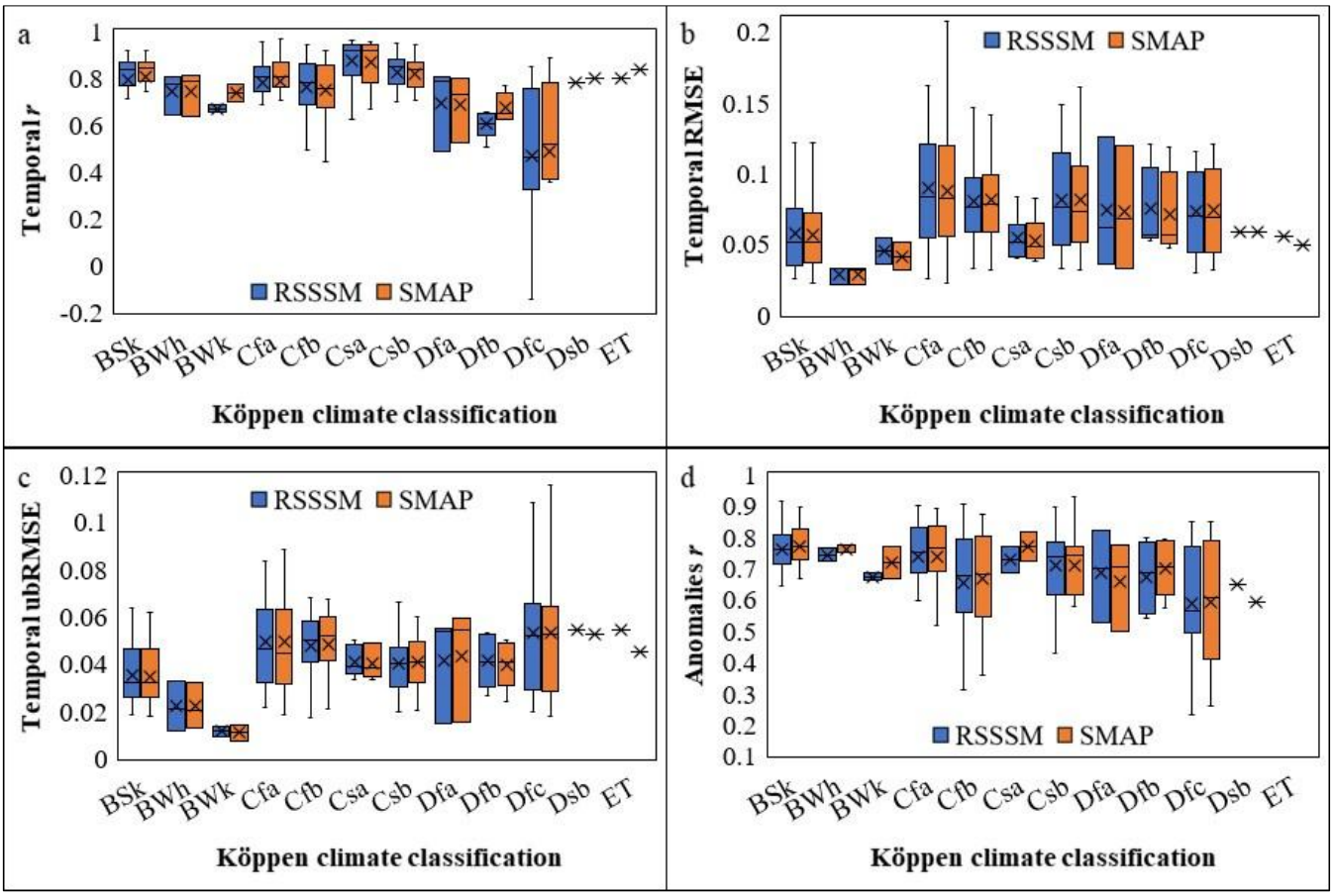


Figure 5: Comparison between the temporal accuracy of RSSSM and SMAP in regions with different Köppen-Geiger climate types. The four indexes are (a) *r*, (b) RMSE, (c) ubRMSE and (d) Anomalies *r* (A.R). The lengths of the error bars are 1.5 times that of the interquartile range, while the upper and 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).

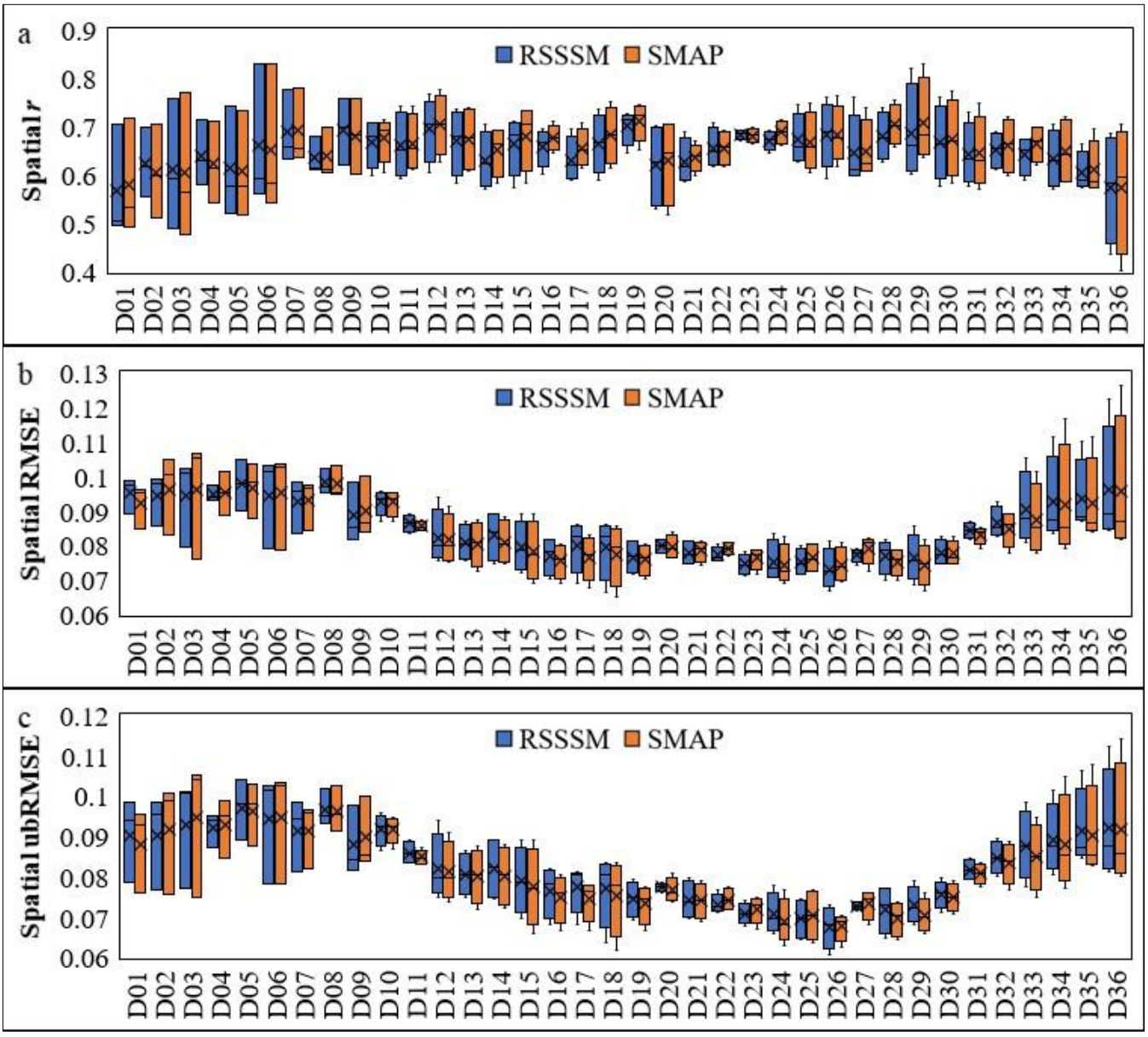


Figure 6: Comparison between the spatial pattern accuracy of RSSSM and SMAP in different 10-day periods during April 2015~2018. The three evaluation indexes are (a) *r*, (b) RMSE and (c) ubRMSE. The length of each box/error bar is determined from the evaluation index values in three (January to March) or four (April to December) years.

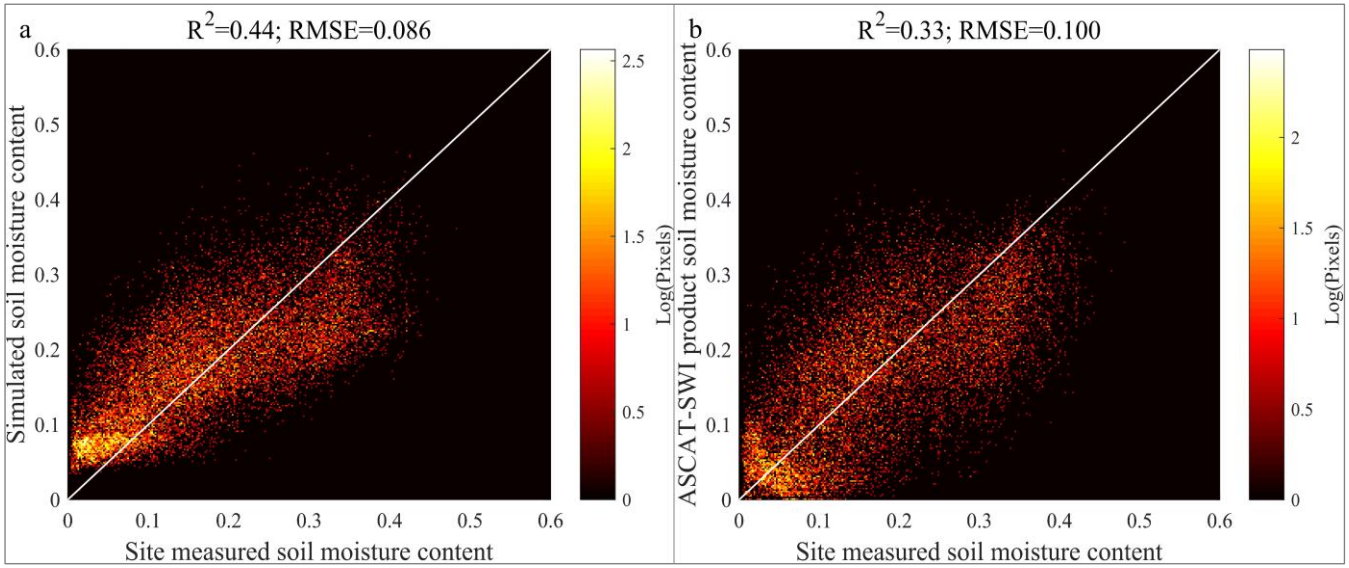


Figure 7: The overall data accuracy comparison between RSSSM and the ASCAT-SWI data product. The scatter plot is between (a) RSSSM or (b) ASCAT-SWI soil moisture and the site measured values during 2007~2018. The unit of all plots is the density of points on a logarithmic scale.

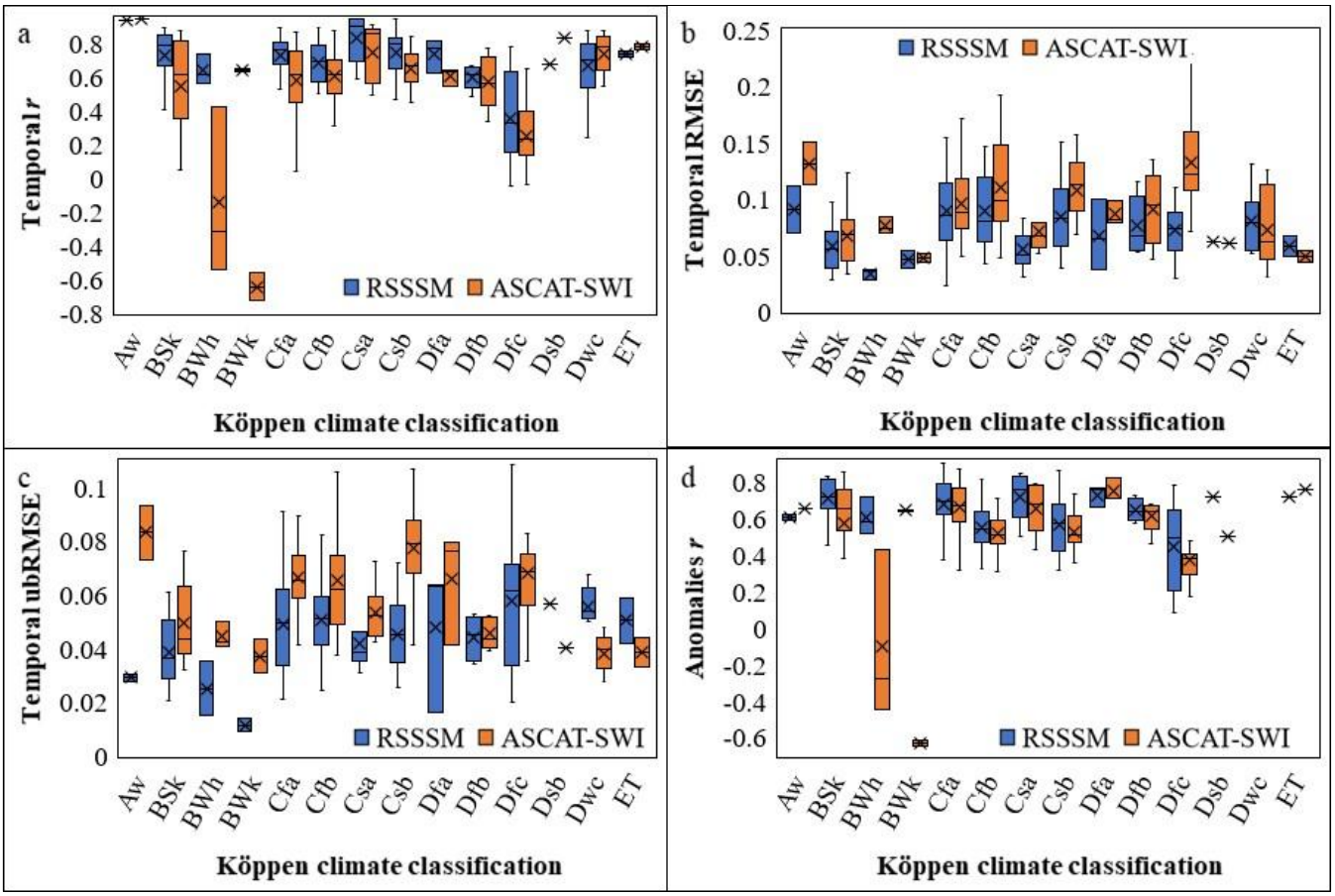


Figure 8: Comparison between the temporal accuracy of RSSSM and ASCAT-SWI in different Köppen-Geiger climatic regions. The four indexes are (a) *r*, (b) RMSE, (c) ubRMSE, and (d) Anomalies R (A.R).

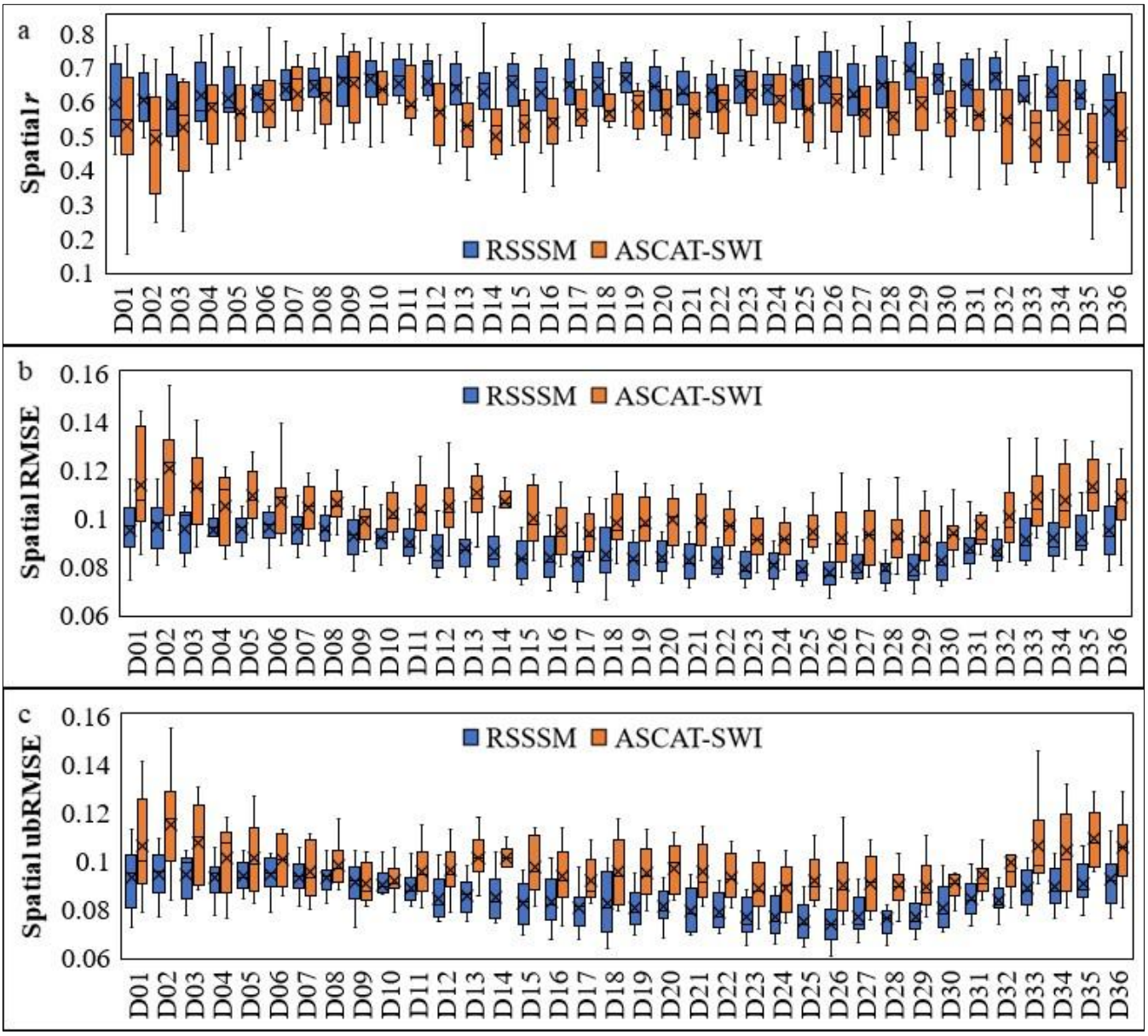


Figure 9: Comparison between the spatial accuracy of RSSSM and ASCAT-SWI during different 10-day periods. The evaluation indexes are (a) *r*, (b) RMSE, and (c) ubRMSE.

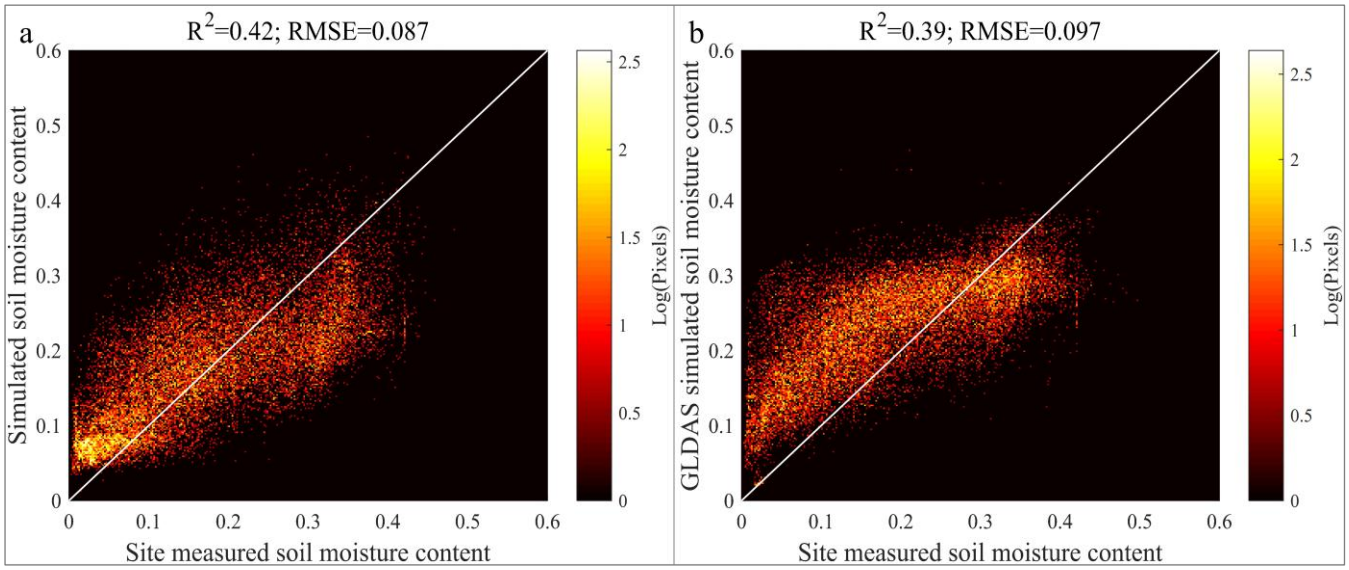
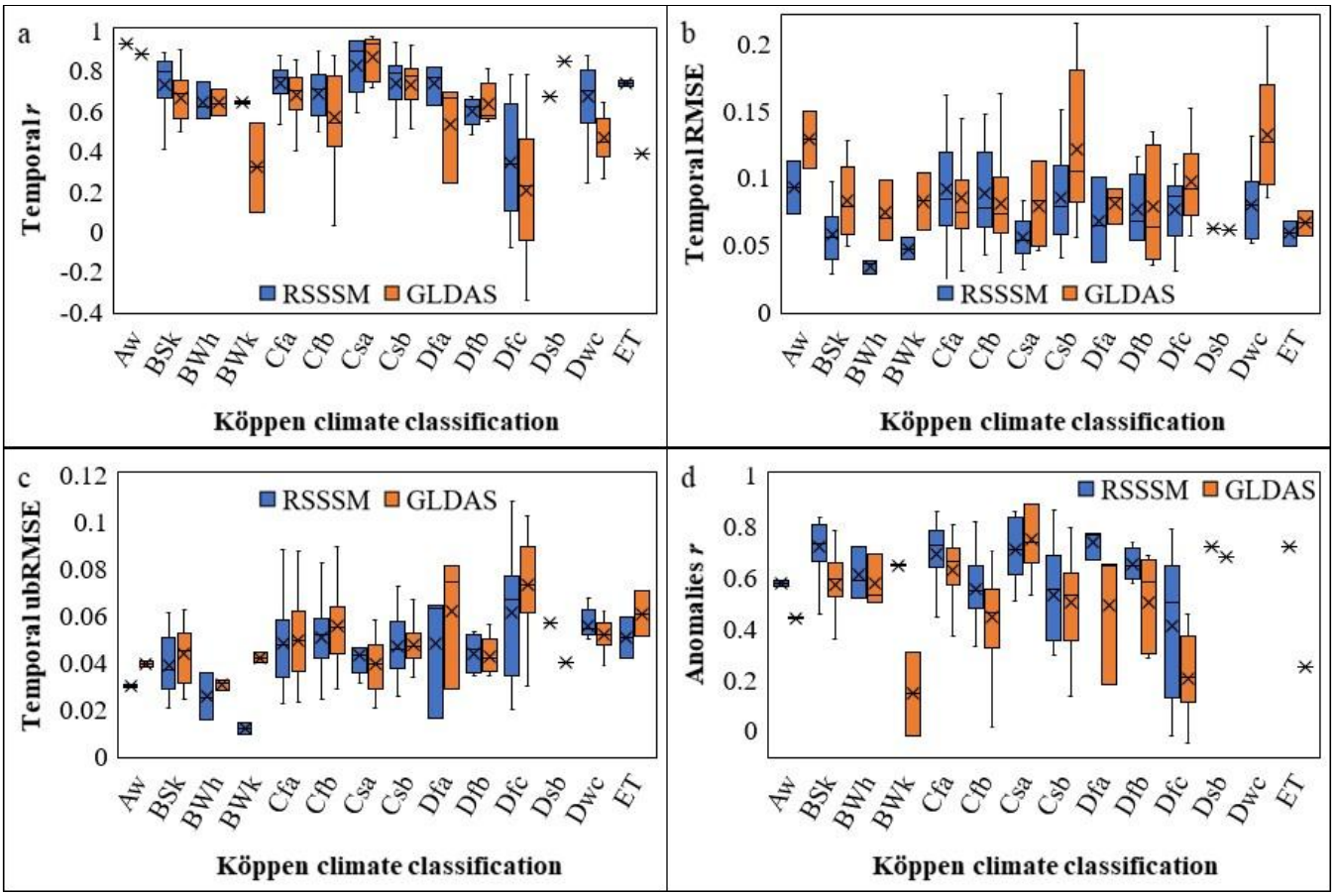


Figure 10: The overall data accuracy comparison between RSSSM and the surface soil moisture simulated by GLDAS Noah V2.1. The scatter plot is between the (a) RSSSM or (b) GLDAS soil moisture and the measured soil moisture during 2003~2018.



1035

Figure 11: Comparison between the temporal accuracy of RSSSM 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*.

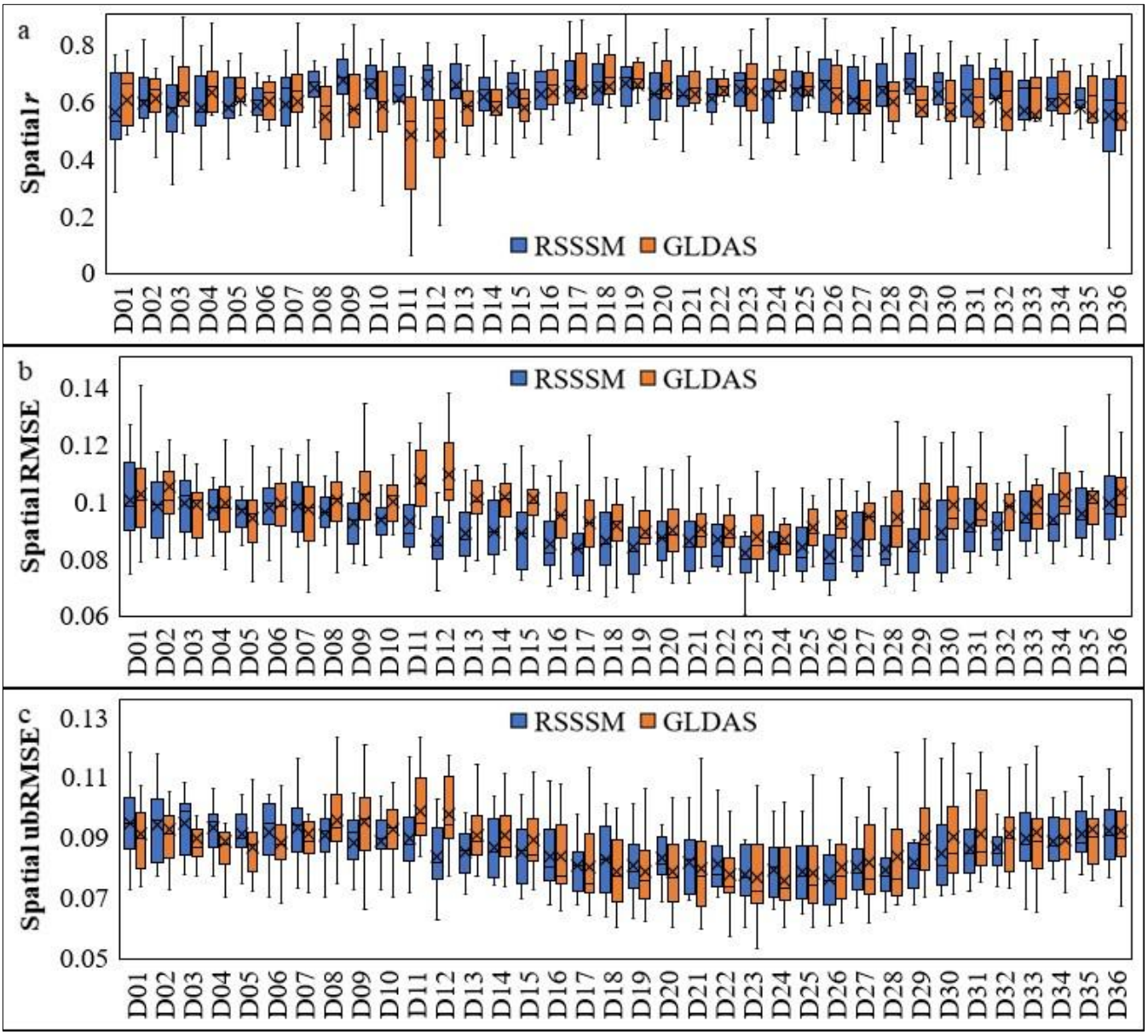


Figure 12: Comparison between the spatial accuracy of RSSSM and GLDAS during different 10-day periods. The evaluation indexes are the same as those in Figure 7.

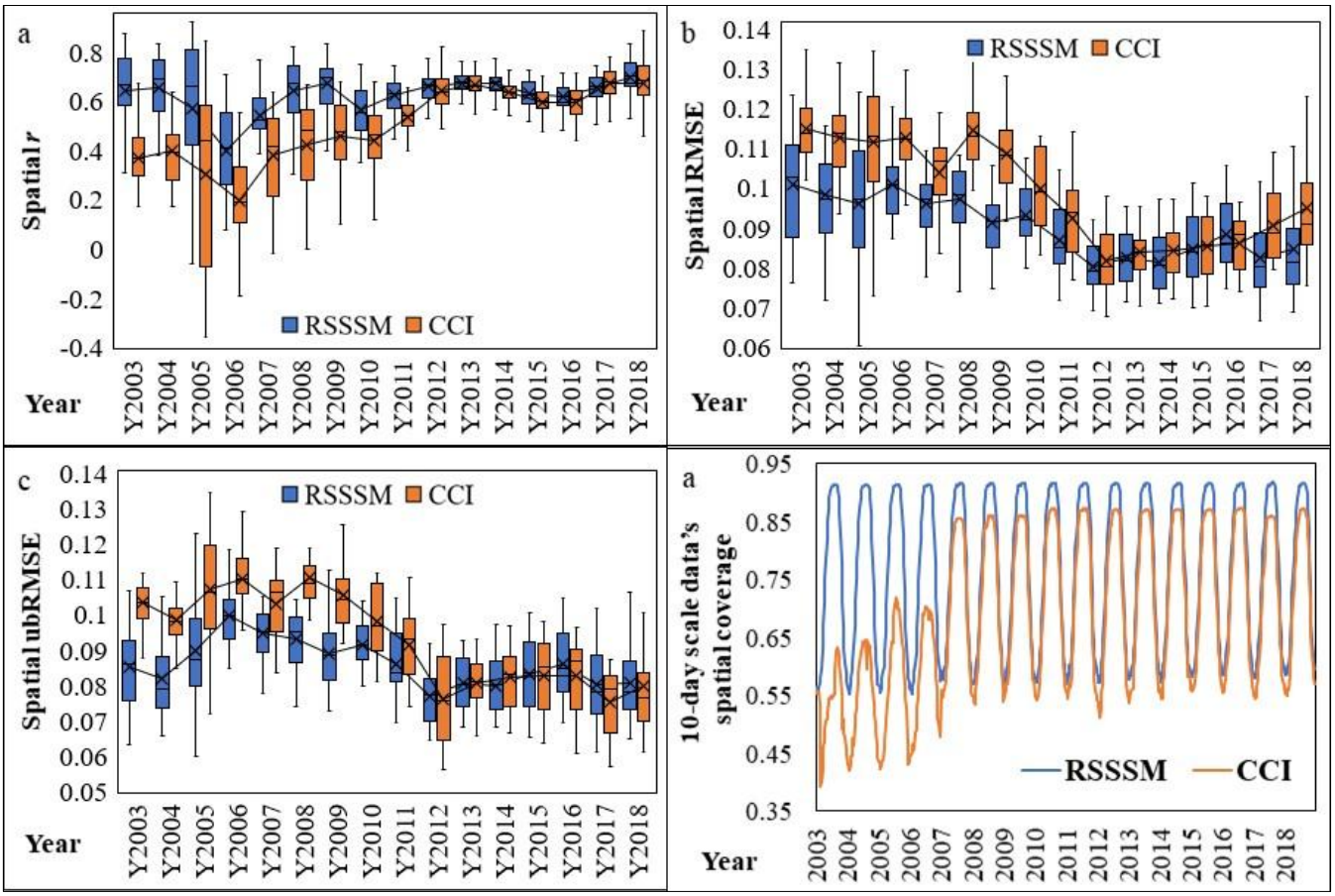


Figure 13: Changes in the data quality and data spatial coverages of RSSSM and CCI soil moisture with year. The interannual changes in (a) spatial correlation coefficients (*r*), (b) spatial RMSE, (c) spatial ubRMSE values, and (d) the spatial coverages of 10-day period data of RSSSM and CCI.

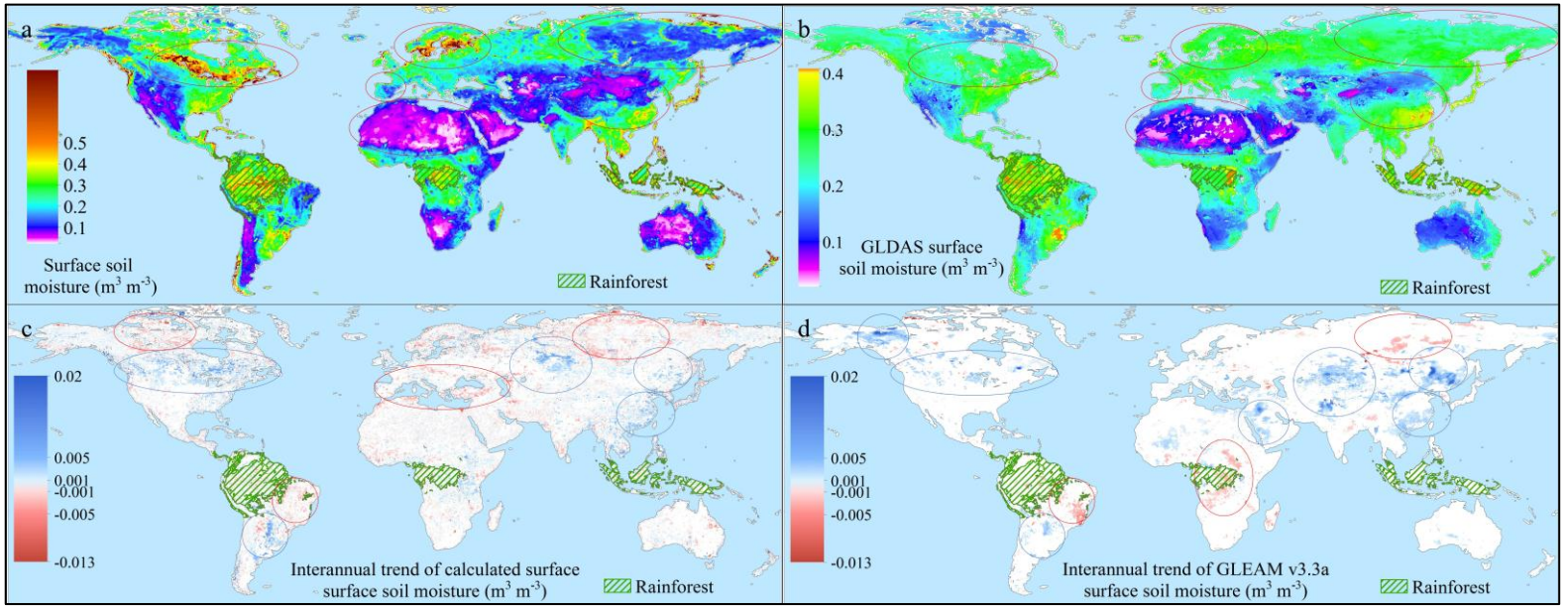


Figure 14: The spatial and temporal patterns of the neural network simulated surface soil moisture (RSSSM) and comparison against other products: (a~b) the global map of (a) calculated RSSSM and (b) GLDAS Noah V2.1 soil moisture (averaged during 2003~2018); (c~d) the interannual trend map of (c) calculated RSSSM and (d) GLEAM v3.3a soil moisture from 2003 to 2018. The circled regions in (a~b) are the places with obvious differences between RSSSM and the other products, while the circled regions in (c~d) are those with significant trends.

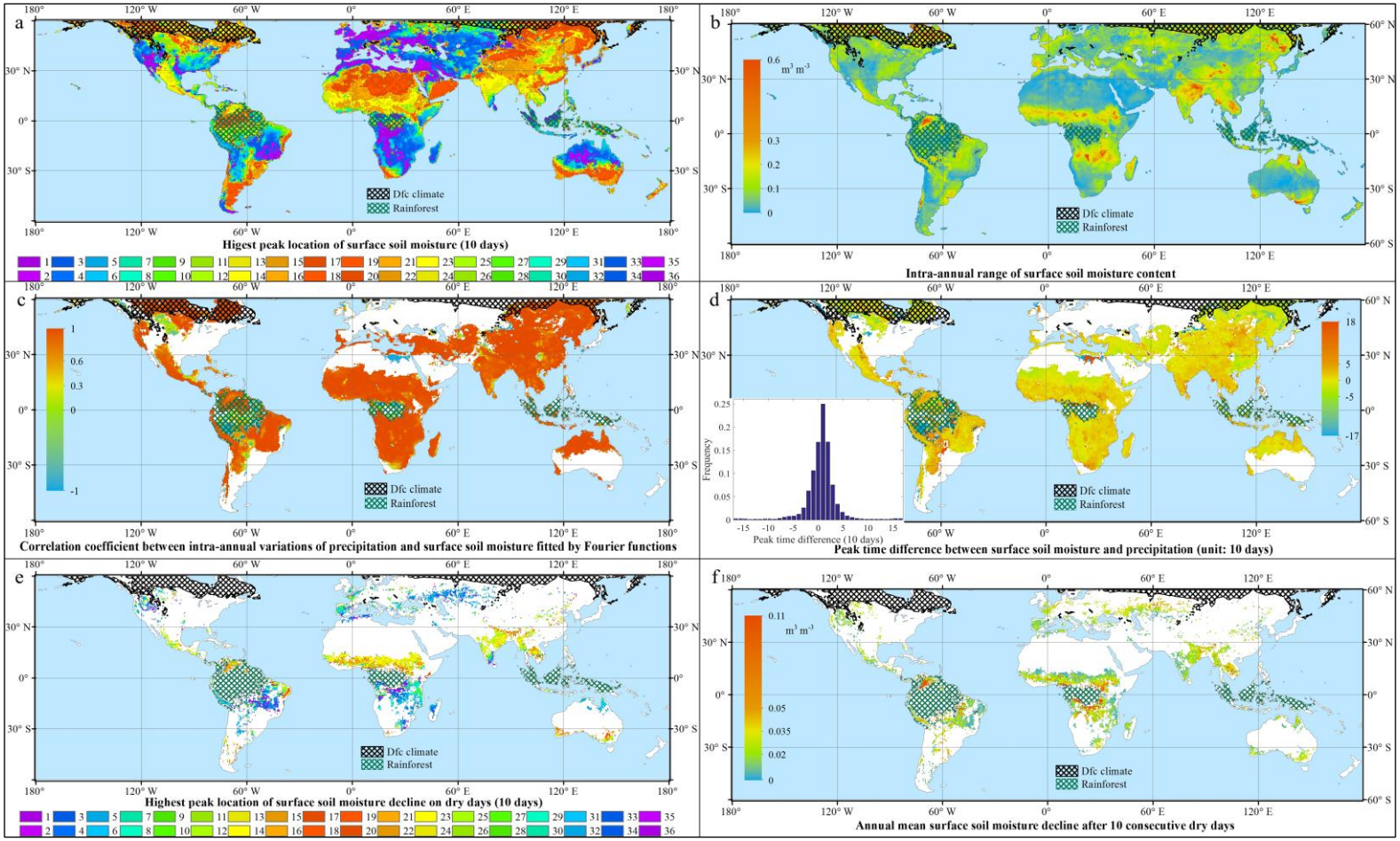


Figure 15: The intra-annual variation in global surface soil moisture and its relationship with precipitation. (a) The spatial pattern of the time when surface soil moisture reaches its maximum in a year (unit: 10 days, note that the seasons are opposite in the Northern and Southern Hemispheres); (b) the intra-annual variation range of surface soil moisture; (c) the map of the correlation coefficient between the intra-annual variations in precipitation and surface soil moisture (both are fitted by Fourier periodic functions); (d) the peak time difference between the surface soil moisture and precipitation (unit: 10 days), with the frequency histogram shown as the inset; (e) the 10-day period with the fastest surface soil moisture loss on rainless days in every 0.5° grid over the world; and (f) map of the annual mean surface soil moisture decline after 10 consecutive dry days (if assuming that the dry period occurs randomly throughout a year).