



OPEN

DATA DESCRIPTOR

A 21-year dataset (2000–2020) of gap-free global daily surface soil moisture at 1-km grid resolution

Chaolei Zheng , Li Jia & Tianjie Zhao

Global soil moisture estimates from current satellite missions are suffering from inherent discontinuous observations and coarse spatial resolution, which limit applications especially at the fine spatial scale. This study developed a dataset of global gap-free surface soil moisture (SSM) at daily 1-km resolution from 2000 to 2020. This is achieved based on the European Space Agency - Climate Change Initiative (ESA-CCI) SSM combined product at 0.25° resolution. Firstly, an operational gap-filling method was developed to fill the missing data in the ESA-CCI SSM product using SSM of the ERA5 reanalysis dataset. Random Forest algorithm was then adopted to disaggregate the coarse-resolution SSM to 1-km, with the help of International Soil Moisture Network *in-situ* observations and other optical remote sensing datasets. The generated 1-km SSM product had good accuracy, with a high correlation coefficient (0.89) and a low unbiased Root Mean Square Error (0.045 m³/m³) by cross-validation. To the best of our knowledge, this is currently the only long-term global gap-free 1-km soil moisture dataset by far.

Background & Summary

Soil moisture (SM) is a key state variable in the climate system and hydrological cycle, and it controls the exchange of water, energy, and carbon fluxes between the land surface and atmosphere^{1–6}. SM datasets are essential for a wide range of applications in hydrology, meteorology, climatology, and water resource management^{7–13}. SM presents high spatial and temporal variability due to the complex interactions among various correlated variables such as soil texture and structure, topographic features, land cover patterns, vegetation properties, and meteorological forcing^{14–17}. These factors generally are difficult to isolate, and their coupled impacts on SM variability vary significantly over time and space domain^{13,18,19}.

Different ground observation techniques have been developed to measure SM, e.g., the gravimetric methods, time/frequency domain reflectometry, neutron probes, electrical resistivity measurements, heat pulse sensors, fiber optic sensors^{2,8,11,18}. Based on these point-scale ground observations, global SM networks, such as the International Soil Moisture Network (ISMN), have been established, and significant progress has been made in characterizing the spatial and temporal variation of SM to improve our understanding of the earth system²⁰. However, *in-situ* measurements are limited in terms of spatial representativeness, and extrapolation of such point-scale measurements to large spatial scale is usually complex and time-consuming, especially on land surface with high spatial heterogeneity^{15,21–24}. Spatial and temporal quantification of SM distributions at regional and global scales based on these ground-observed datasets remains challenging²⁵.

Satellite remote sensing technology can obtain surface SM (SSM) from regional to global scale, and the observations from active and passive microwave sensors are considered to be one of the best tools for SSM retrieval^{16,26,27}. Various satellites and algorithms have been developed with the ability to map SSM from satellite-based microwave sensors²⁸, such as the Advanced Microwave Scanning Radiometer–EOS (AMSR-E) and its successor AMSR2, Fengyun-3 MicroWave Radiation Imager (MWRI), the advanced scatterometer (ASCAT), Soil Moisture and Ocean Salinity (SMOS), Soil Moisture Active Passive (SMAP), and European Space Agency Sentinel-1 satellite^{29–33}, and global soil moisture products have been produced accordingly^{34–39}. Although significant progress has been made to merge various satellites data to improve the remote sensing SSM coverage, there are still many gaps in the daily SSM dataset due to the limited satellite orbit/swath and retrieving capability. For example, it is found that the widely used multi-sensor fusion SSM product from European Space

State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, 100101, China. ✉e-mail: zhengcl@aircas.ac.cn

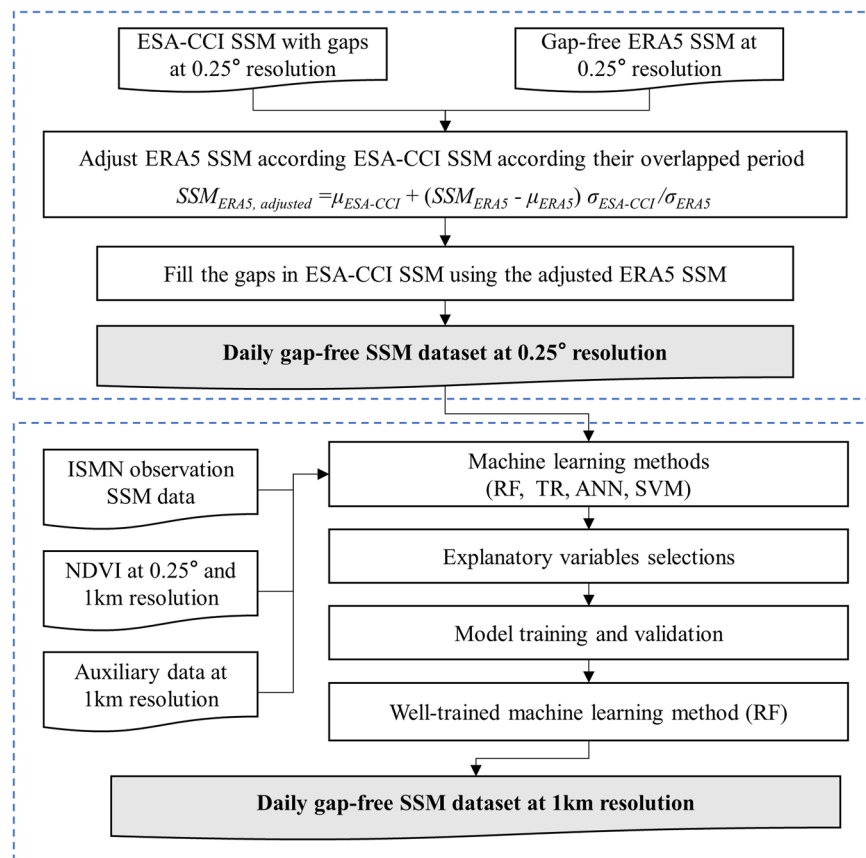


Fig. 1 Schematic overview of the methodology and data products for generating the daily gap-free SSM dataset at 1-km resolution.

Agency - Climate Change Initiative (ESA-CCI) product has a very low spatial coverage (roughly 20%) in the central and western Tibetan Plateau⁴⁰.

Another key limitation is that most of these global SSM products are at relative coarse spatial resolution, e.g., tens of kilometers, which limits the applications in regional hydrological and agricultural studies. Several downscaling approaches have been proposed to improve the spatial resolution of SSM product by considering the impacts of different environmental variables²⁵. The idea behind these downscaling methods is to establish either a statistical function or a physically based model between coarse-resolution SSM and fine-resolution auxiliary variables^{41–47}. However, several limitations were found for these downscaling methods, including the linear or nonlinear assumption to define the impact of spatial heterogeneity^{42–44}, the error of input fine-resolution data, the uncertainties associated with the model parameter estimates^{45–47}, and these may introduce large uncertainty. These downscaling methods generally use complex and computationally intensive disaggregation algorithms that are generally unsuitable for a global implementation due to complex and varying nonlinear relationships between soil moisture and the determinant variables used for downscaling⁴⁸. Consequently, the high-resolution SSM dataset with global coverage is still lacking. Machine learning algorithms are increasingly used to extract patterns and insights of different geospatial variables from the ever-increasing stream of Earth system science data⁴⁹, and have been proved to be a feasible method to disaggregate SSM (capture the complex nonlinear relationships) at coarse resolution hence to generate high resolution SSM at global scale^{50–55}.

Considering the importance of high-resolution gap-free SSM data, this study aims at generating a high-resolution SSM dataset at global scale with continuity at both space and time scales by developing a SSM downscaling algorithm based on machine learning method. A global gap-free SSM dataset at daily scale and 1-km spatial resolution from 2000 to 2020 is finally generated.

Methods

Experimental design. Previous studies on the evaluation of different soil moisture products^{56–58} conclude that ESA-CCI SSM has high accuracy and shows the best consistency with the ground observations. The top layer SM data from the European Centre for Medium-Range Weather Forecasts reanalysis v5 (ERA5) product shows good temporal correlation with ground observation, but with systematically large bias⁵⁷. Hence, it's reasonable to merge these two datasets by utilizing the high accuracy of ESA-CCI SM and good temporal variation and global gap-free coverage of ERA5 SM to generate a gap-free SSM dataset. The global daily gap-free SSM dataset at 1-km resolution was achieved by the following two steps in this study (Fig. 1). First, the ESA-CCI SSM product was gap-filled using ERA5 reanalysis product, and we achieved a daily gap-free SSM data at 0.25° resolution. Then,

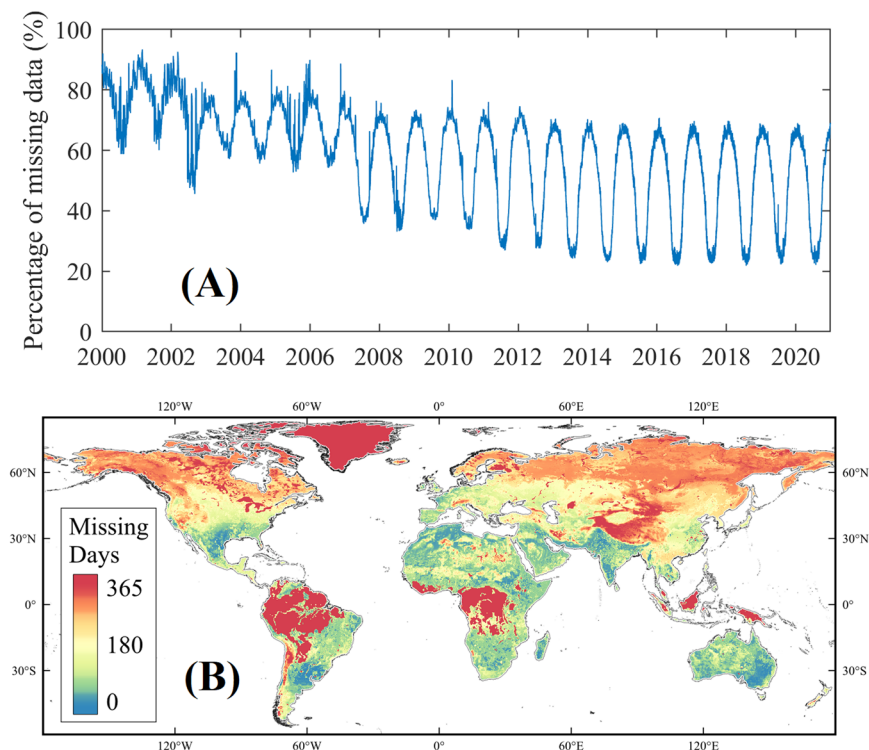


Fig. 2 Statistics of missing data in the original ESA-CCI SSM products. **(A)** Percentage of missing data globally in ESA-CCI SSM dataset from 2000 to 2020. **(B)** Missing days per year in the original ESA-CCI SSM dataset averaged over multi-years (from 2000 to 2020).

machine learning models were trained to downscale the daily gap-free SSM data at 0.25° resolution to 1-km resolution with the help of fine-resolution auxiliary data. We will introduce the data and algorithm used in this study in the following sections.

Satellite and auxiliary data. The ESA-CCI SSM product is currently the available dataset of satellite-based soil moisture with the longest data record to date. The ESA-CCI SSM v06.1 product at coarse resolution was adopted in current study for further gap-filling and downscaling. The ESA-CCI SSM products were obtained using various satellite-based observations from microwave sensors since 1978, with 0.25° grid resolution at daily interval. It provides three SSM datasets: the merged dataset from observations by active microwave sensors (“Active Product”), the merged dataset from observations by passive microwave sensors (“Passive Product”), and the combined dataset. The “Active Product” and the “Passive Product” were created by fusing soil moisture products from scatterometer and radiometer observations, respectively. The combined SM product was obtained by merging all active and passive SSM observations directly through temporal resampling, spatial resampling, Cumulative Distribution Function (CDF)-based rescaling, and triple collocation analysis-based merging algorithm. We selected the combined dataset in this study since it was supposed to inherit the advantages of both active and passive microwave observations, and it generally outperformed the products using single-sensor observation as input⁵⁹. Although enormous efforts have been conducted to obtain SSM at daily scale with global coverage, the daily ESA-CCI SSM products still could not fully cover the global land surface. The missing data percentage range from 21.8% to 94.41% at daily step from 2000 to 2020 according to our statistics, with an averaged value of 58.17% (Fig. 2). Even the global missing data percentage decreased dramatically with the increase of available satellite data after 2007, the minimum missing values of daily ESA-CCI SSM could still count for 21.8% of global land surface (Antarctica excluded). (Fig. 2). Large gaps are especially in winter time for the northern hemisphere due to frozen water content in soil, which is difficult to be detected by microwave bands^{60,61}. Studies intended to exclude the densely vegetated regions, since the retrieval errors are relatively large in these regions because the sensitivity of the radiometer to SSM is reduced due to the strong attenuation of the ground emission signal by vegetation⁶². Missing data days per year were also high in high elevations regions, e.g., the missing observation days was found larger than 200 days per year in the Tibetan Plateau.

The top layer SM data from the ERA5 product was downloaded from the Copernicus Climate Data Store (<https://cds.climate.copernicus.eu/>) and used to fill the missing values in ESA-CCI SSM dataset. The ERA5 was built upon its predecessor (ERA-Interim), and it combined more historical observations and run on finer resolutions⁶³. The ERA5 SSM has a globally spatial-temporal continuous coverage, with a spatial resolution of 0.25° and temporal resolution of 1 hr. In this study, daily ERA5 SSM was calculated by averaging the hourly ERA5 SSM. The ERA5 SSM has better performance in terms of correlation with ground observations than some other soil

moisture reanalysis products⁵⁶, and it could reasonably regenerate the monthly dynamics and annual cycles, especially the timings of the strong dry-wet transition⁵⁷.

The predictors used in the machine learning algorithm for downscaling SSM includes Normalized Difference Vegetation Index (NDVI), surface albedo, digital elevation model (DEM), and saturated soil moisture. To obtain the 1-km resolution SSM data, different optical remote sensing datasets at high resolution were collected and processed to obtain daily values of these predictor variables at 1-km resolution. The monthly NDVI data with 1-km resolution were from MOD13A2 product⁶⁴. The monthly 0.05° resolution NDVI from MOD13C1⁶⁵ were used and aggregated to 0.25° resolution to match the spatial resolution of ESA-CCI SSM data in the downscaling model. They were further interpolated linearly to daily temporal resolution to match the temporal resolution of SSM. The albedo data was from the Global Land Surface Satellite (GLASS) product^{66,67}, and it was reconstructed to daily step by linear interpolation for further application. The topographic information was retrieved from the SRTM30 DEM⁶⁸, and the DEM at 1-km resolution was retrieved from its native 30 Arc Seconds resolution using bilinear interpolation method. The global 1-km saturated soil moisture was obtained from previous study that produce a high-resolution global map of soil hydraulic properties by a hierarchical parametrization of a physically based water retention model⁶⁹, using the surface soil of SoilGrids dataset as input⁷⁰.

Ground observation data. The *in-situ* soil moisture observations datasets from ISMN were collected to train the machine learning method for SSM downscaling and validate the results. The ISMN is a soil moisture dataset network established and maintained through international cooperation. The SM observations have been collected around the world by different research teams and harmonized to make the data available for public research, through the coordination by the Global Energy and Water Exchanges Project²⁰. To date, the ISMN data consists of measurements from 2,879 sites in 68 networks (last access on February 10, 2022). These *in-situ* SM observations play an increasingly substantial role in evaluating satellite and model products^{2,20,71–74}. The ISMN data adopted in current study is listed in Supplementary Table S1. Further details about the instruments and the data quality control of the observations can be found in the network reports and references therein (available from <https://ismn.geo.tuwien.ac.at>).

ESA-CCI SSM gap-filling at 0.25° resolution. The applicability of SSM data is often hindered by spatio-temporal gaps. Global reanalysis data are featured by high spatial coverage and high temporal resolution, which could be used to fill the gaps in remote sensing SSM dataset. However, in many regions, global reanalysis data lack of accuracy and are biased from ground true or satellite retrievals^{62,64,74}. One solution is to make use of the consistency of temporal variation between the remote sensing SSM time series and the SSM of reanalysis data, while re-scale (or adjust) the magnitude of the reanalysis SSM according to the remote sensing SSM. In current study, the missing values in ESA-CCI SSM was gap-filled by using the ERA5 SSM to obtain daily gap-free ESA-CCI SSM at 0.25° resolution. To avoid the inconsistency between the ESA-CCI SSM and the ERA5 SSM, the daily ERA5 SSM was re-scaled (adjusted) according to the ESA-CCI SSM before it was used for gap-filling. The re-scaling of ERA5 SSM was done by establishing a linear relationship for each 0.25° grid between these two SSM time series using daily data on the overlapped days. A simple linear relationship between the ERA5 SSM and ESA-CCI SSM on overlapped days for each 0.25° grid could be built as below

$$SSM_{ESA-CCI} = a SSM_{ERA5} + b \quad (1)$$

where a and b were fitting parameters. Once the coefficients a and b are defined in Eq. (1), the re-scaled (or adjusted) ERA5 SSM, $SSM_{ERA5, adjusted}$ could be obtained as

$$SSM_{ERA5, adjusted} = a SSM_{ERA5} + b \quad (2)$$

Assuming the adjusted daily ERA5 SSM and the original daily ERA5 SSM depart the same way from their mean values of time series ($\mu_{ESA-CCI}$ and μ_{ERA5}) with the same standard deviations ($\sigma_{ESA-CCI}$ and σ_{ERA5}), the following equations could be obtained,

$$SSM_{ERA5, adjusted} - \mu_{ESA-CCI} = SSM_{ERA5} - \mu_{ERA5} \quad (3a)$$

$$(SSM_{ERA5, adjusted} - \mu_{ESA-CCI})/\sigma_{ESA-CCI} = (SSM_{ERA5} - \mu_{ERA5})/\sigma_{ERA5} \quad (3b)$$

Equation (3b) could be rearranged as

$$SSM_{ERA5, adjusted} = \sigma_{ESA-CCI}/\sigma_{ERA5} \cdot SSM_{ERA5} + \mu_{ESA-CCI} - \mu_{ERA5} \cdot \sigma_{ESA-CCI}/\sigma_{ERA5} \quad (4)$$

Hence, the pixel-wise a and b in Eq. (2) were obtained, and $SSM_{ERA5, adjusted}$ was estimated using the averaged values and standard deviation values of ERA5 SSM and ESA-CCI SSM over the overlapped days. $SSM_{ERA5, adjusted}$ estimated by Eq. (4) was then adopted to fill the missing values in the ESA-CCI SSM for each 0.25° grid. For the 0.25° grids where no overlapped data were found (roughly 10% of global land surface), mainly in the tropical rainforest regions, the ERA5 SSM was directly adopted to fill the missing values in the ESA-CCI SSM dataset. Even it may be controversy, ERA5 also gives the SSM value for the water and snow/ice covered pixels, and it was directly used to fill the missing values of water and snow/ice covered pixels during the SSM gap-filling phase in this study. Finally, the global daily gap-free SSM at 0.25° resolution was achieved.

Spatial downscaling. The daily gap-free ESA-CCI SSM at 0.25° resolution was disaggregated to 1-km using machine learning method. The disaggregation strategy first learns the nonlinear relationships between the *in-situ*

| Test | Explanatory variables |
|-------|--|
| Test1 | SSM_{25km} , $NDVI_{25km}$, $NDVI_{1km}$ |
| Test2 | SSM_{25km} , $NDVI_{25km}$, $NDVI_{1km}$, α_{1km} |
| Test3 | SSM_{25km} , $NDVI_{25km}$, $NDVI_{1km}$, DEM_{1km} |
| Test4 | SSM_{25km} , $NDVI_{25km}$, $NDVI_{1km}$, $\theta_{S, 1km}$ |
| Test5 | SSM_{25km} , $NDVI_{25km}$, $NDVI_{1km}$, α_{1km} , DEM_{1km} |
| Test6 | SSM_{25km} , $NDVI_{25km}$, $NDVI_{1km}$, α_{1km} , $\theta_{S, 1km}$ |
| Test7 | SSM_{25km} , $NDVI_{25km}$, $NDVI_{1km}$, DEM_{1km} , $\theta_{S, 1km}$ |
| Test8 | SSM_{25km} , $NDVI_{25km}$, $NDVI_{1km}$, DEM_{1km} , α_{1km} , $\theta_{S, 1km}$ |

Table 1. Experiment design to explore the performances by using different explanatory variables for SSM downscaling.

observations of SSM and the ESA-CCI soil moisture at 0.25° resolution and NDVI at both 0.25° and 1-km resolutions, to predict the fine resolution (1-km) SSM. The ISMN observation data were adopted to train the machine learning model for SSM downscaling. We also tested the performance of different machine learning methods and different combinations of explanatory variables, which allowed us to select the most suitable model and explanatory variables for SSM downscaling.

We first explored the performances of using different explanatory variables, several tests were conducted with different variables included in the machine learning models (Table 1). The SSM_{25km} , $NDVI_{25km}$ and $NDVI_{1km}$ was employed in Test1. The surface albedo (α_{1km}), digital elevation model (DEM_{1km}), and surface saturated soil moisture ($\theta_{S, 1km}$), were added successively from Test2 to Test8. These variables were selected because of their physical relationships with the spatial variation of SSM. For instance, high SSM values generally associated with good vegetation conditions (high NDVI and low albedo). The NDVI both at coarse and fine resolutions were used in previous studies⁵¹. The albedo showed exponential relationship with SSM and systematic decrease in albedo in response to rainfall were observed widely^{75,76}. DEM_{1km} and $\theta_{S, 1km}$ somehow related to the soil water holding capacity⁴⁰. Land surface temperature (LST) was highly related to SSM, but it was not selected for SSM downscaling since the gap-free LST at moderate (e.g., 1-km) resolution were not available due to the impact of cloudiness. It's also noted that precipitation was not selected as an explanatory variable considering that the global moderate resolution precipitation dataset was not available. We will consider to update the auxiliary datasets in the future when they are available.

We tested and compared four machine learning algorithms to select the most accurate algorithm for SSM downscaling, including the Random Forest (RF), the Support Vector Machine (SVM), the Tree-based Regression (TR), and the Artificial Neural Networks (ANN). We used k -fold ($k = 10$ in current study) cross validation to validate and compare the downscaled SSM by different models and tests. The SSM observation data from ISMN was divided into k -fold (or groups) randomly, with one of the folds (10% of the observation data) was left as 'unmeasured' and the remaining $k-1$ folds (90% of the observation data) were used for training the models. The trained models were validated based on the 'unmeasured' data. The training and validation procedures were repeated 10 times, using a different fold as the holdout set for each time; hence all data was selected for validation. This validation could explore the transferability of the downscaling model from known *in-situ* SSM observation sites to any other sites for global applications. The correlation coefficient (R) and unbiased root-mean-square error ($ubRMSE$) were used to evaluate the performances of the four different machine learning methods.

Figure 3 illustrated the performance of different machine learning models using different combinations of explanatory variables by cross validation. Generally, with the inclusion of more explanatory variables in all the four models, R was increasing, and $ubRMSE$ was decreasing. Best performance (with high R and low $ubRMSE$) was found in Test8, when all six explanatory variables were selected. This indicated that the SSM can be predicted accurately when all the selected explanatory variables were included in the downscaling model. Hence, all six explanatory variables (SSM_{25km} , $NDVI_{25km}$, $NDVI_{1km}$, DEM_{1km} , α_{1km} , $\theta_{S, 1km}$) were adopted in the SSM downscaling model. All these four machine learning algorithms showed good performance in predicting SSM ($R > 0.6$ for all). The Tree and RF methods showed much better performance than ANN and SVM, the best results were given by the RF method with the highest R (0.89) and lowest $ubRMSE$ ($0.05 \text{ m}^3/\text{m}^3$) than the other three algorithms (Fig. 3). Based on the performance analyzed above, the RF model was applied to downscale SSM product and to generate global daily/1-km resolution SSM product using the explanatory variables in Test8, i.e., SSM and NDVI at 25-km resolution, NDVI, surface albedo, DEM and surface saturated soil moisture at 1-km resolution.

Data Records

The final daily/1-km SSM product accounts for more than 1 TB of data capacity. Due to the storage limitation of the online repositories, we provide the monthly averaged 1-km SSM data for download from the data portal of National Tibetan Plateau/Third Pole Environment Data Center⁷⁷ (<https://doi.org/10.11888/RemoteSen.tpd.c.272760>) after user registration. Data are freely available in this data portal. For easy read and manipulation, the monthly 1-km SSM data are stored in geotiff format with one file for each month with global coverage. Users can use most Geographic Information Systems (GIS) and remote sensing software packages to read and manipulate the data. The file names follow the structure of "SM.1km.Month.YYYYMM.Global.v001.tif", where "SM.Month.1km." represents the 1-km monthly averaged SSM product, "YYYY" is the year, "MM" represents the month, "Global" represents the global coverage, and "v001" indicates the product version.

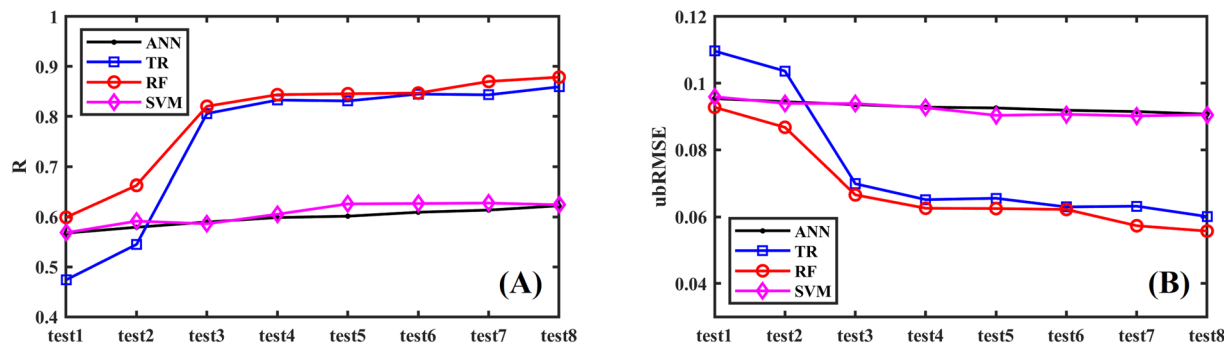


Fig. 3 Comparison of the performance of SSM_{1km} predictions based on the four machine learning models using different explanatory variables.

| Data | Long name | Number type | Unit | Fill value | Scale factor | Added offset | Valid range |
|------|-----------------------------|-------------|-----------|------------|--------------|--------------|-------------|
| SM | Daily soil moisture | Int16 | m^3/m^3 | -1 | 0.001 | 0 | 0 ~ 1000 |
| flag | Flag of daily soil moisture | Int8 | None | -1 | none | none | 0 ~ 8 |

Table 2. Detailed information about the Scientific DataSet in the daily SSM product.

Daily SSM in customized regions are available on request to the corresponding author (zhengcl@aircas.ac.cn) with details of the intended and desired spatial and temporal resolution, domain, and period of interest in user's request. The detailed information of the daily SSM product is shown in Table 2. The daily 1-km SSM data are stored in hdf5 format (<https://www.hdfgroup.org/>), and global data was divided into tiles with sinusoidal grid projection following the data structure of MODIS products. Each file covers roughly $10^\circ \times 10^\circ$ area, and more details on the tiles and projection could be seen from https://modis-land.gsfc.nasa.gov/MODLAND_grid.html. The file names of daily 1-km SSM data follow the structure of "SM.1km.Daily.YYYYDOY.tiles.v001.h5", where "SM.Daily.1km" represents the daily 1-km SSM product, "YYYY" is the year, "DOY" represents the day of the year (from 001 to 365 or 366), "tiles" represents the tile number (e.g., h24v05) according to the MODIS sinusoidal grid, and "v001" indicates the product version. A quality flag was provided in the daily SSM dataset, ranging from 0 to 8 (0: original ESA-CCI SSM was used; 1: original ESA-CCI SSM was not available and the gap-filled SSM was used; 2 ~ 3: RF algorithm was failed, and value was from simple linear or nearest interpolation; 4 ~ 7: input data was missing, e.g. albedo, NDVI; 8: non-soil pixels). It should be noticed that SSM retrieval/downscaling is not available for water and snow/ice surfaces, and those pixels with NDVI below zero were set as non-soil pixels and marked in quality flag for the 1-km SSM. Users can use Python, IDL, MATLAB, etc., to read and manipulate the data.

Technical Validation

Validation of the gap-filled SSM at resolution of 0.25° . Theoretically, the percentage of SSM data gaps can be reduced to zero after the gap-filling procedure in this study. Figure 4 shows the examples of the gap-filled SSM on typical winter and summer days. Given that SSM does not apply to non-soil pixels such as water and snow/ice cover surfaces, these pixels are masked in Fig. 4. Generally, the gap-filled SSM data could capture the global SSM spatial variation while retaining the original information of ESA-CCI SSM.

Figure 5 represents the temporal behaviours of the original ESA-CCI SSM and the gap-filled ESA-CCI SSM in 0.25° grids at selected ISMN sites. Large gaps could be found in the early years of the time series of the original ESA-CCI SSM in some selected grids. The average and standard deviation values of ESA-CCI SSM and ERA5 SSM during their overlapped days are also shown in Fig. 5. Although relatively large difference could be found between ESA-CCI SSM and ERA5 SSM in some sites grids, their temporal variations are generally consistent with each other. The gap-filled ESA-CCI SSM followed the temporal variation of ERA5 SSM and original ESA-CCI SSM, and showed consistent magnitude with the original ESA-CCI SSM. The averaged values of gap-filled ESA-CCI SSM are also stable, and systematic error introduced by the gap-filling method is negligible.

To validate the reasonability of the gap-filling method, a k -fold ($k = 10$ in current study) validation was conducted. The daily ESA-CCI SSM values in each 0.25° grid in 2000–2020 were randomly divided into 10 folds, and each fold of the 10-folds was taken out and predicted by the remaining 9 folds. This procedure was repeated 10 times until all the 10 folds of ESA-CCI SSM data were traversed (predicted). All the predicted SSM values were gathered to compare with the original ESA-CCI SSM series, the results were shown in Fig. 6. It generally shows the reliability of the gap-filled ESA-CCI SSM with overall high R values (0.98) and low bias ($0.001 m^3/m^3$) globally between the gap-filled ESA-CCI SSM and the original ESA-CCI SSM (Fig. 6A). The temporal variation of annual global mean values of the predicted SSM showed very close pattern to the original ESA-CCI SSM (Fig. 6B). Lower R was found in high-latitude cold regions in the northern hemisphere and the extreme arid region (e.g. the Sahel desert and the western Tibetan Plateau) (Fig. 6C). There were more missing data in the

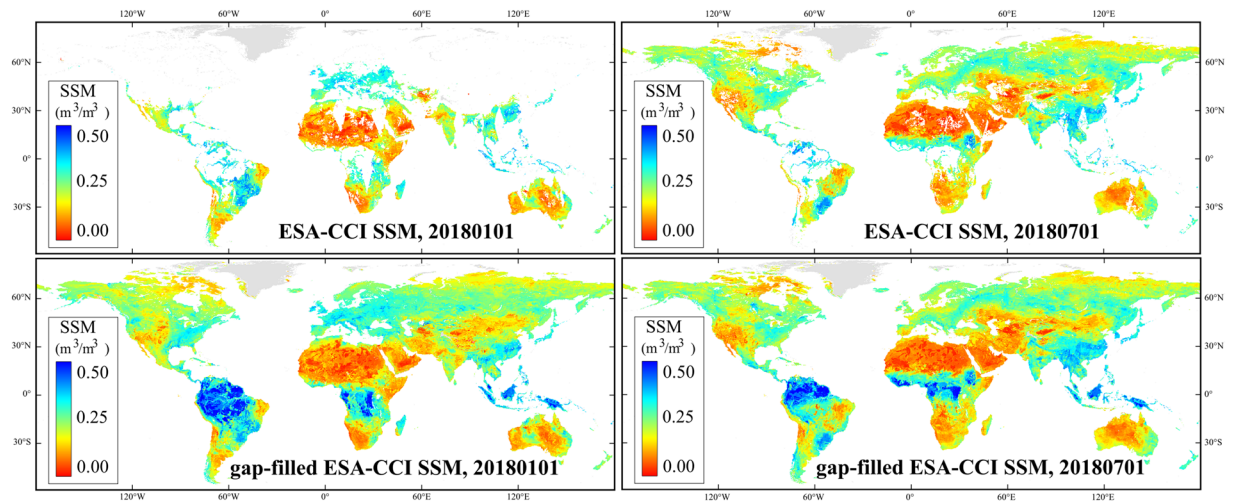


Fig. 4 Global map examples of the original ESA-CCI SSM product and gap-filled ESA-CCI SSM on January 1st and July 1st, 2018 (m^3/m^3).

ESA-CCI SSM dataset in the high-latitude cold regions in the northern hemisphere, which partly explained the lower R and larger bias (Fig. 6C and Fig. 6D). In the desert regions, SSM was very low, and low R was anticipated due to larger uncertainty in the retrieved SSM under very dry condition.

Validation of the downscaled SSM at 1-km resolution based on ISMN observation. As demonstrated in the Method section, the RF model outperformed other methods, it was therefore used to produce global daily/1-km SSM from 2000 to 2020. Figure 7 presents some examples of monthly averaged downscaled SSM at 1-km resolution in January, April, July, and October of 2018. For better illustration of the regional SSM distribution, zoom-in views of different selected sub-regions are also shown in Fig. 7. The selected sub-regions in Fig. 7 cover roughly $5^\circ \times 5^\circ$ with ISMN network included, which are 1) USDA-ARS network of America, 2) SMOSMANIA network in Europe, 3) AMMA-CATCH network in Africa, 4) OZNET network in Australia, and 5) HiWATER-EHWSN in China. The global 1-km SSM can capture well the overall spatial variations of global SSM, and the spatial features of SSM are well illustrated by the high-resolution SSM as shown in the sub-region maps.

Supplementary Table S2 lists the accuracy metrics of the downscaled 1-km SSM in each ISMN network compared with the ISMN observations. The overall bias of the downscaled 1-km SSM is $0 \text{ m}^3/\text{m}^3$ with the range from -0.065 to $0.015 \text{ m}^3/\text{m}^3$, R is 0.89 with the range from 0.325 to 0.962, $ubRMSE$ is $0.045 \text{ m}^3/\text{m}^3$ with the range from $0.015 \text{ m}^3/\text{m}^3$ to $0.069 \text{ m}^3/\text{m}^3$.

Figure 8 shows the temporal variation of downscaled 1-km SSM at selected ISMN sites. The 1-km SSM are very close to the ground observations, and it can trace the seasonal variation of SSM very well. Although the ground measured data used in Fig. 8 were included in training the models (hence not independent), it is acceptable to take them as reference for evaluation of the performance of the downscaled SSM. The results demonstrated the prediction ability of the RF model used for downscaling of SSM. Meanwhile, to illustrate the ability of the 1-km soil moisture data to detect rainfall events, precipitation information was also shown in Fig. 8. Clearly, the soil moisture fluctuates with precipitation, especially in the arid land, e.g., the OZNET Yanco-Research station, where the dry-down process (soil moisture depletes following the precipitation event) could be well captured by the 1-km soil moisture data.

Usage Notes

In this study, we provided a dataset of global spatiotemporally continuous daily surface soil moisture at 1-km resolution from 2000 to 2020 for various applications and studies. The 1-km SSM dataset generated in this study has several potential applications. For example, it was successfully applied to ETMonitor model for global high-resolution evapotranspiration estimation⁷⁸, in which soil moisture was used to parameterize soil surface resistance to soil evaporation and canopy resistance to plant transpiration, so as to better consider the influence of soil moisture on land surface evapotranspiration. Compared with those evapotranspiration dataset that did not use soil moisture information, e.g., the MOD16 evapotranspiration product, the error of estimated evapotranspiration by ETMonitor decreased significantly after using SM information at 1-km resolution, e.g., RMSE of global 8-days evapotranspiration decreased from 1.34 mm/d by MOD16 to 0.83 mm/d by ETMonitor⁷⁸.

The high-resolution SSM dataset also has the potential in distinguishing irrigated fields, inferring irrigation water use, improving wildfire danger prediction, etc. However, the size of farmland is generally small, and higher spatial resolution (e.g., 30-m) may be more appropriate for distinguishing irrigation and non-irrigation fields. Furthermore, it is not practical to assess the performance of identifying irrigation fields using the 1-km SM data due to lack of *in-situ* irrigation information. Further evaluation will be needed to assess the capability of the downscaled 1-km SSM for distinguishing between irrigated and non-irrigated fields. An alternative way is that the irrigation information could be achieved by comparing satellite derived and modelled SM (the latter does not include

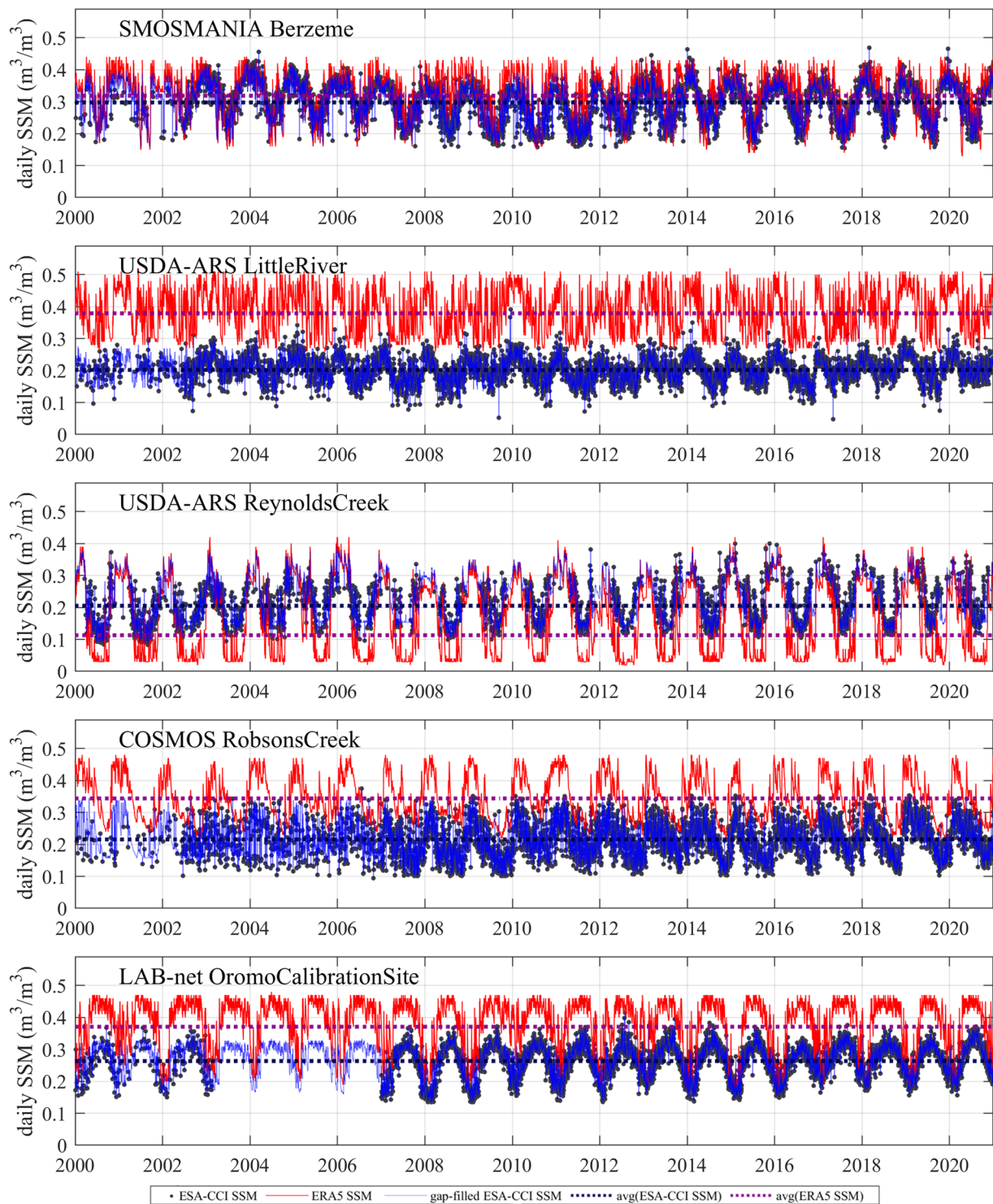


Fig. 5 Examples of the time series of the gap-filled ESA-CCI SSM and the original ESA-CCI SSM product, ERA5 SSM products in 0.25° grids at the selected ISMN sites in 2000–2020.

irrigation information)⁷⁹, or by inverting soil water balance⁸⁰ equation to derive the total in-flow water in the soil⁸⁰. However, we should notice these retrieve algorithms need to be calibrated carefully to achieve good accuracy⁸¹.

Notably, the 1-km resolution SSM in this study is obtained by downscaling the low-resolution SSM data, which is essentially to spatially redistributed microwave-based soil moisture in the coarse grid (0.25°) to enclosed pixels (grids) at high resolution (1-km in this study), hence the high-resolution SSM inherits the uncertainty of the low-resolution SSM product. Although comparison with the *in-situ* observations from the ISMN at global scale shows satisfactory accuracy, considering that the *in-situ* observation sites used for validation are relatively sparse and the distribution of ISMN is extremely uneven, it is impossible to guarantee the same quality

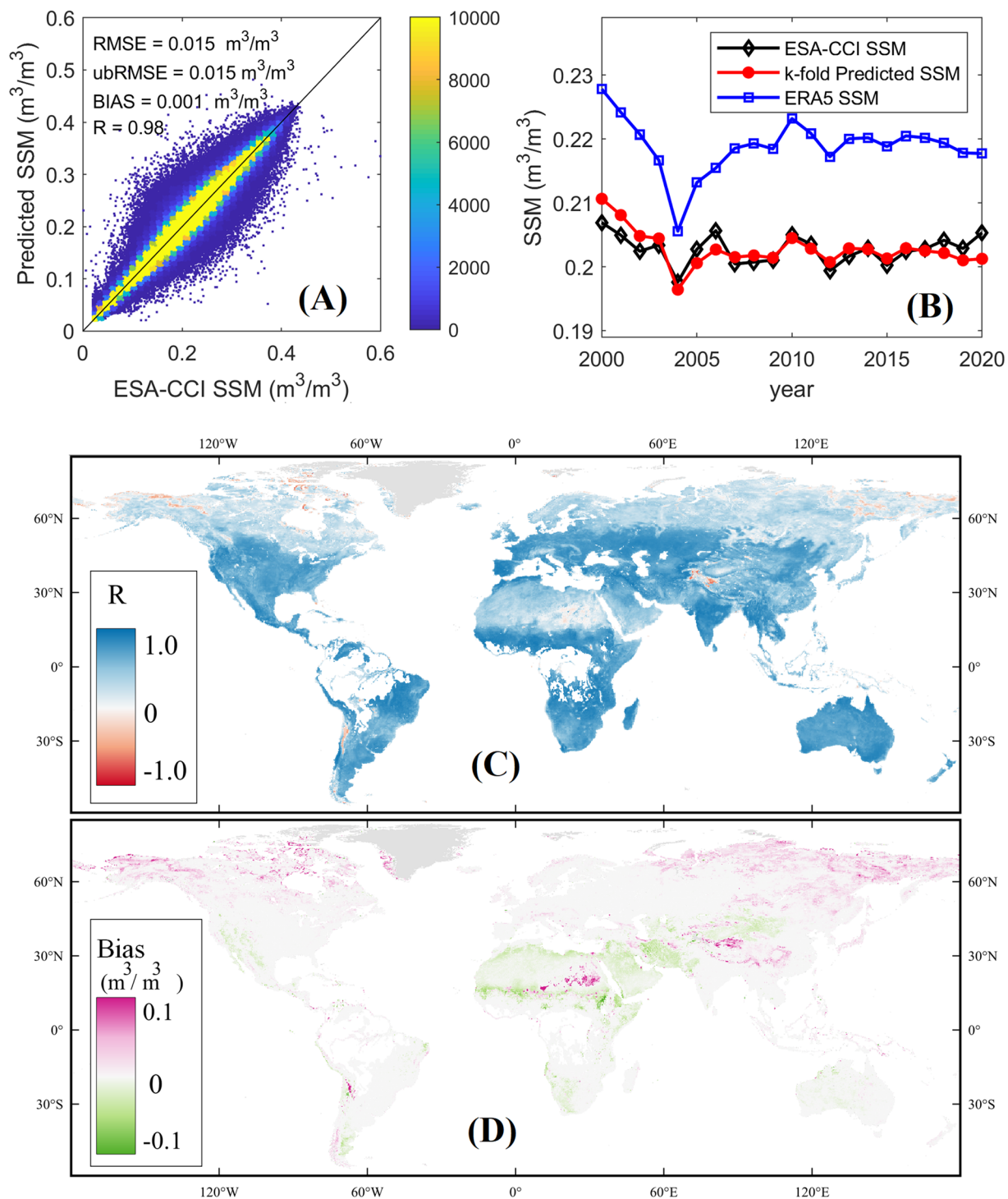


Fig. 6 Results of *k*-fold cross-validation of the gap-filled ESA-CCI SSM: **(A)** Scatterplot of the predicted ESA-CCI SSM against the original ESA-CCI SSM; **(B)** Temporal variation of annual mean values of the predicted ESA-CCI SSM, original ESA-CCI SSM, and ERA5 SSM; **(C)** R of predicted ESA-CCI SSM, **(D)** Bias of predicted ESA-CCI SSM.

in different regions of the world. Moreover, one should be aware of the limitation of machine-learning-based model, which cannot always correctly capture the variations in SM. Previous study has reported that machine-learning-based model failed to track the ‘tipping points’ (where a slowly changing soil moisture triggered a sudden shift to a new soil moisture) when applied to SSM prediction⁸². In additions to the inherent capability of machine-learning-based models, the choice of explanatory variables has a significant impact on the results, and the uncertainty in the employed input datasets will certainly be propagated into the downscaled

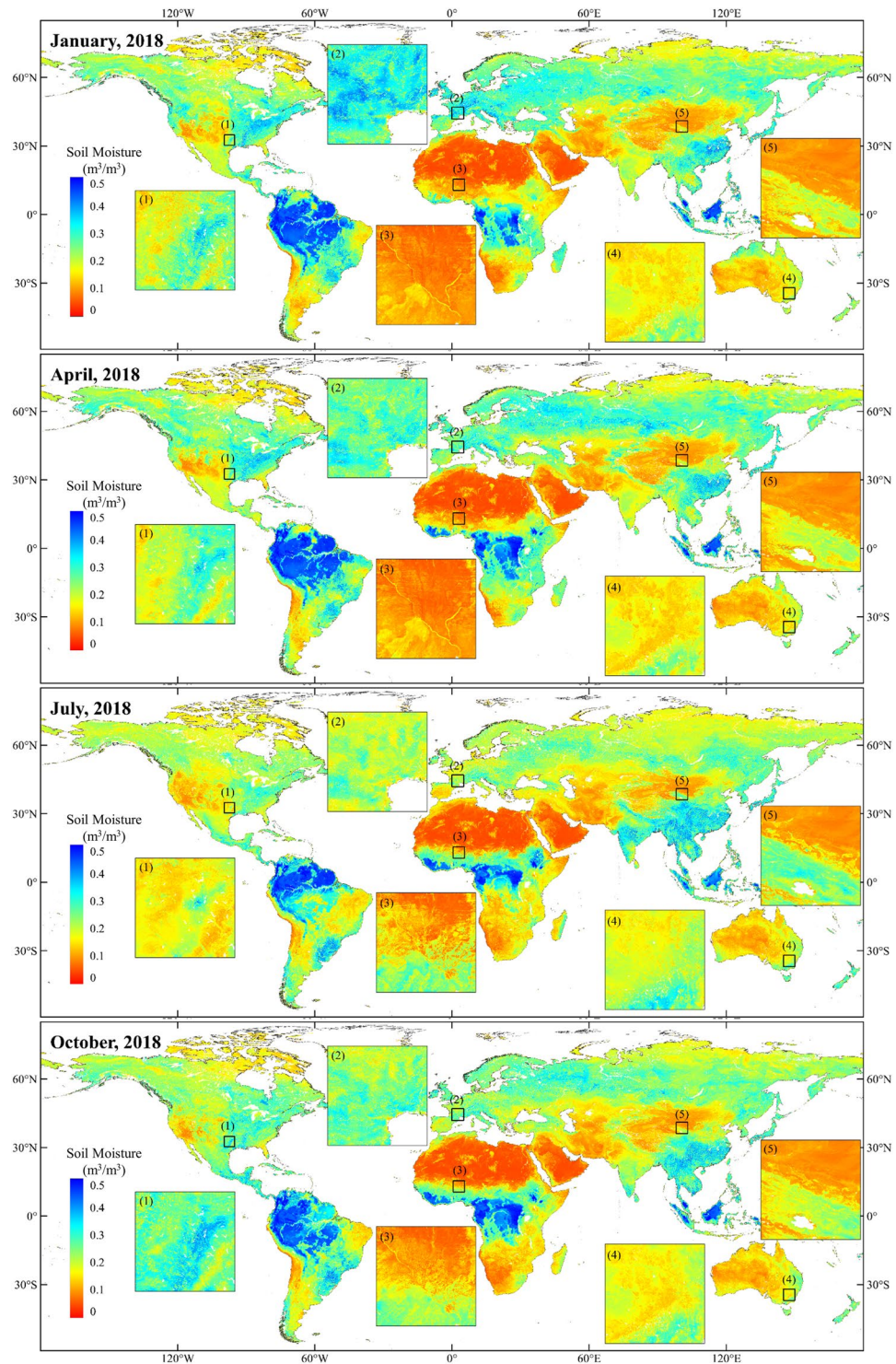


Fig. 7 Global distribution of monthly averaged SSM at 1-km resolution in January, April, July and October of 2018. The zoom-in views of the selected sub-regions cover different ISMN networks: (1) USDA-ARS network of America, (2) SMOSMANIA network in Europe, (3) AMMA-CATCH network in Africa, (4) OZNET network in Australia, and (5) HiWATER-EHWSN in China.

1-km SSM. The temporal resolution of the dataset is achieved by our method using ERA5 dataset for day-by-day filling. However, it is difficult to analyse the actual physical spatial resolution, which can be very complex and related to all microwave and optical datasets used in the study. It can be inferred that the actual physical resolution would vary from location to location. Therefore, cautions should be paid when applying the downscaled SSM dataset for further analysis, e.g., for detection of convective rainfall events, for prediction of flood and landslide risk at high resolution.

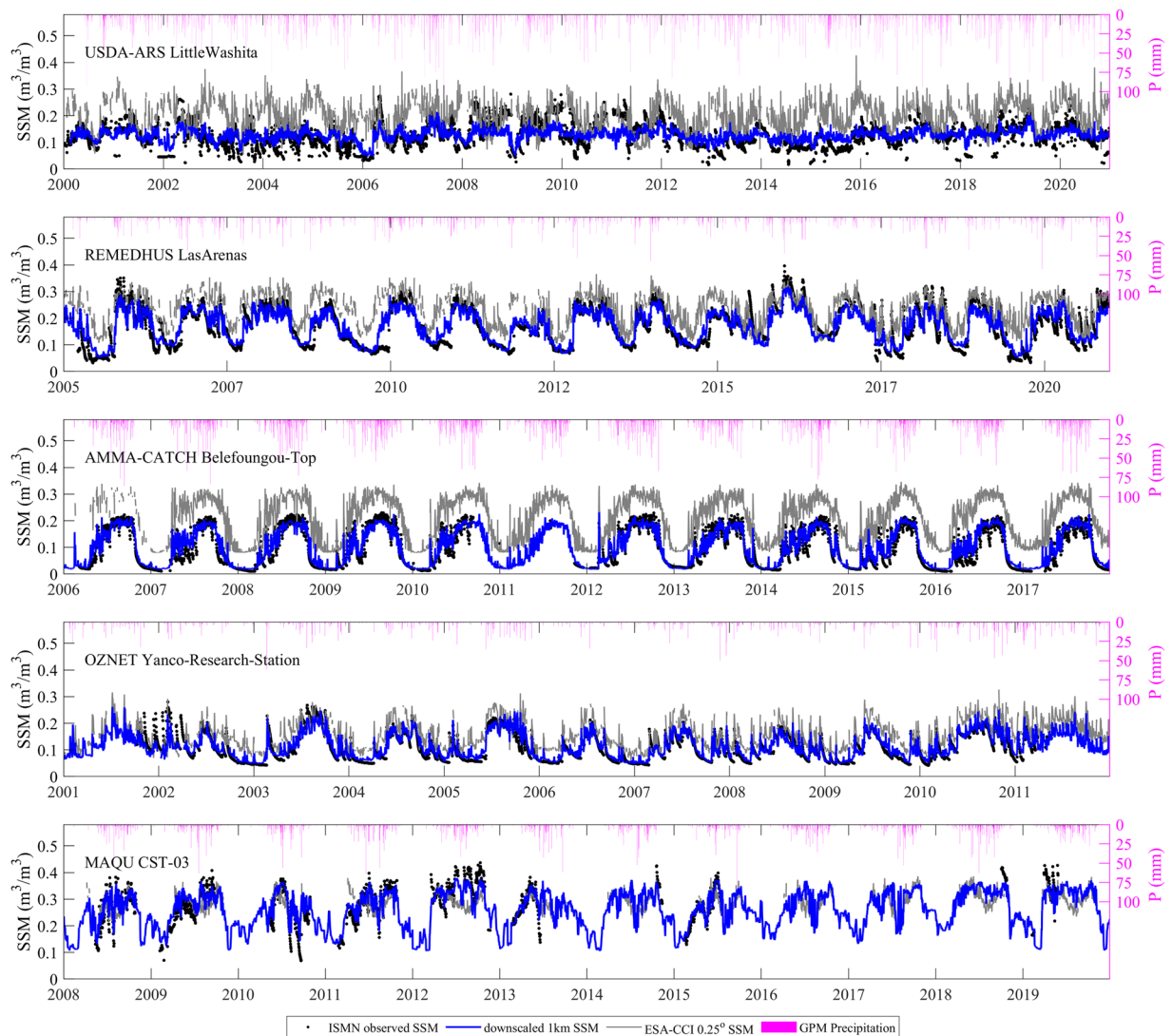


Fig. 8 Time series of the downscaled 1-km SSM from the grids at the selected ISMN sites. Precipitation (P) from GPM product was also shown.

All the monthly 1-km SSM data are stored in geotiff in the data portal of National Tibetan Plateau/Third Pole Environment Data Center⁷⁷ (<https://doi.org/10.11888/RemoteSen.tpd.272760>). The daily 1-km SSM data, stored in hdf5 format, are available on request to the corresponding author (zhengcl@aircas.ac.cn). Users can freely choose the spatial and temporal coverage of SSM dataset according to their specific research objectives. Users can use Python, IDL, MATLAB, and popular Geographic Information Systems (GIS) or remote sensing software packages to read and manipulate the data. It should be noted that the data must be multiplied by their corresponding scale factors (in Table 2). Instructions for data post-processing (converting to geographic coordinates, etc.) is provided with the data upon request. The dataset will be updated in the future when new or better input data become available.

Code availability

The codes used in this study will be available at <https://github.com/zhengchaolei/GlobalSSMGapfillDownscaling.git> after this work is accepted.

Received: 23 August 2022; Accepted: 27 January 2023;

Published online: 15 March 2023

References

- Ochsner, T. E. *et al.* State of the Art in Large-Scale Soil Moisture Monitoring. *Soil Sci. Soc. Am. J.* **77**, 1888–1919 (2013).
- Robock, A. *et al.* The Global Soil Moisture Data Bank. *Bull. Am. Meteorol. Soc.* **81**, 1281–1300 (2000).
- Wagner, W. *et al.* Operational readiness of microwave remote sensing of soil moisture for hydrologic applications. *Hydrol. Res.* **38**, 1–20 (2007).
- Western, A. W. & Blöschl, G. On the spatial scaling of soil moisture. *J. Hydrol.* **217**, 203–224 (1999).

5. Western, A. W., Grayson, R. B. & Blöschl, G. Scaling of soil moisture: A hydrologic perspective. *Annu. Rev. Earth Planet. Sci.* **30**, 149–180 (2002).
6. Seneviratne, S. I. *et al.* Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Sci. Rev.* **99**, 125–161 (2010).
7. Pauwels, V. R. N., Hoeben, R., Verhoest, N. E. C., De Troch, F. P. & Troch, P. A. Improvement of TOPLATS-based discharge predictions through assimilation of ERS-based remotely sensed soil moisture values. *Hydrol. Process.* **16**, 995–1013 (2002).
8. Robinson, D. A. *et al.* Soil Moisture Measurement for Ecological and Hydrological Watershed-Scale Observatories: A Review. *Vadose Zo. J.* **7**, 358–389 (2008).
9. Dai, A., Trenberth, K. E. & Qian, T. A global dataset of Palmer Drought Severity Index for 1870–2002: Relationship with soil moisture and effects of surface warming. *J. Hydrometeorol.* **5**, 1117–1130 (2004).
10. Koster, R. D. *et al.* Regions of strong coupling between soil moisture and precipitation. *Science*. **305**, 1138–1140 (2004).
11. Dobriyal, P., Qureshi, A., Badola, R. & Hussain, S. A. A review of the methods available for estimating soil moisture and its implications for water resource management. *J. Hydro.* **458–459**, 110–117 (2012).
12. Hu, G. & Jia, L. Monitoring of evapotranspiration in a semi-arid inland river basin by combining microwave and optical remote sensing observations. *Remote Sens.* **7**, 3056–3087 (2015).
13. Zhao, T. *et al.* Soil moisture experiment in the Luan River supporting new satellite mission opportunities. *Remote Sens. Environ.* **240**, 111680 (2020).
14. Brocca, L., Morbidelli, R., Melone, F. & Moramarco, T. Soil moisture spatial variability in experimental areas of central Italy. *J. Hydrol.* **333**, 356–373 (2007).
15. Crow, W. T. *et al.* Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. *Rev. Geophys.* **50**, RG2002 (2012).
16. Mohanty, B. P. & Skaggs, T. H. Spatio-temporal evolution and time-stable characteristics of soil moisture within remote sensing footprints with varying soil, slope, and vegetation. *Adv. Water Resour.* **24**, 1051–1067 (2001).
17. Vereecken, H. *et al.* On the value of soil moisture measurements in vadose zone hydrology: A review. *Water Resour. Res.* **46**, W00D06 (2008).
18. Vereecken, H. *et al.* On the spatio-temporal dynamics of soil moisture at the field scale. *J. Hydrol.* **516**, 76–96 (2014).
19. Vanderlinden, K. *et al.* Temporal Stability of Soil Water Contents: A Review of Data and Analyses. *Vadose Zo. J.* **11**, vzt2011.0178 (2012).
20. Dorigo, W. A. *et al.* The International Soil Moisture Network: A data hosting facility for global *in situ* soil moisture measurements. *Hydrol. Earth Syst. Sci.* **15**, 1675–1698 (2011).
21. Collow, T. W., Robock, A., Basara, J. B. & Illston, B. G. Evaluation of SMOS retrievals of soil moisture over the central United States with currently available *in situ* observations. *J. Geophys. Res. Atmos.* **117**, D09113 (2012).
22. Loew, A. Impact of surface heterogeneity on surface soil moisture retrievals from passive microwave data at the regional scale: The Upper Danube case. *Remote Sens. Environ.* **112**, 231–248 (2008).
23. Vinnikov, K. Y., Robock, A. & Speranskaya, N. A. & Adam Schlosser, C. Scales of temporal and spatial variability of midlatitude soil moisture. *J. Geophys. Res. Atmos.* **101**, 7163–7174 (1996).
24. Zreda, M. *et al.* COSMOS: The cosmic-ray soil moisture observing system. *Hydrol. Earth Syst. Sci.* **16**, 4079–4099 (2012).
25. Peng, J., Loew, A., Merlin, O. & Verhoest, N. E. C. A review of spatial downscaling of satellite remotely sensed soil moisture. *Rev. Geophys.* **55**, 341–366 (2017).
26. Jeu, R. A. M. *et al.* Global soil moisture patterns observed by space borne microwave radiometers and scatterometers. *Surv. Geophys.* **29**, 399–420 (2008).
27. Schmugge, T. J., Kustas, W. P., Ritchie, J. C., Jackson, T. J. & Rango, A. Remote sensing in hydrology. *Adv. Water Resour.* **25**, 1367–1385 (2002).
28. Zhao, T. *et al.* Retrievals of soil moisture and vegetation optical depth using a multi-channel collaborative algorithm. *Remote Sens. Environ.* **257**, 112321 (2021).
29. Njoku, E. G., Jackson, T. J., Lakshmi, V., Chan, T. K. & Nghiem, S. V. Soil moisture retrieval from AMSR-E. *IEEE Trans. Geosci. Remote Sens.* **41**, 215–229 (2003).
30. Bartalis, Z. *et al.* Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT). *Geophys. Res. Lett.* **34**, L20401 (2007).
31. Kerr, Y. H. *et al.* The SMOS L: New tool for monitoring key elements of the global water cycle. *Proc. IEEE* **98**, 666–687 (2010).
32. Entekhabi, D. *et al.* The soil moisture active passive (SMAP) mission. *Proc. IEEE* **98**, 704–716 (2010).
33. Kang, C. S. *et al.* Global Soil Moisture Retrievals from the Chinese FY-3D Microwave Radiation Imager. *IEEE Trans. Geosci. Remote Sens.* **59**, 4018–4032 (2021).
34. Owe, M., de Jeu, R. & Holmes, T. Multisensor historical climatology of satellite-derived global land surface moisture. *J. Geophys. Res. Earth Surf.* **113**, F01002 (2008).
35. Naeimi, V., Scipal, K., Bartalis, Z., Hasenauer, S. & Wagner, W. An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations. *IEEE Trans. Geosci. Remote Sens.* **47**, 1999–2013 (2009).
36. Kerr, Y. H. *et al.* Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Trans. Geosci. Remote Sens.* **39**, 1384–1403 (2001).
37. Liu, Y. Y. *et al.* Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. *Hydrol. Earth Syst. Sci.* **15**, 425–436 (2011).
38. Wagner, W. *et al.* The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications. *Meteorologische Zeitschrift.* **22**, 5–33 (2013).
39. Yao, P. *et al.* A long term global daily soil moisture dataset derived from AMSR-E and AMSR2 (2002–2019). *Sci. Data* **8**, 143 (2021).
40. Cui, Y. *et al.* A two-step fusion framework for quality improvement of a remotely sensed soil moisture product: A case study for the ECV product over the Tibetan Plateau. *J. Hydrol.* **587**, 124993 (2020).
41. Song, C., Jia, L. & Menenti, M. Retrieving high-resolution surface soil moisture by downscaling AMSR-E brightness temperature using MODIS LST and NDVI data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **7**, 935–942 (2014).
42. Colliander, A. *et al.* Spatial Downscaling of SMAP Soil Moisture Using MODIS Land Surface Temperature and NDVI during SMAPVEX15. *IEEE Geosci. Remote Sens. Lett.* **14**, 2107–2111 (2017).
43. Merlin, O., Chehbouni, A. G., Kerr, Y. H., Njoku, E. G. & Entekhabi, D. A combined modeling and multipectral/multiresolution remote sensing approach for disaggregation of surface soil moisture: Application to SMOS configuration. *IEEE Trans. Geosci. Remote Sens.* **43**, 2036–2050 (2005).
44. Merlin, O. *et al.* Disaggregation of SMOS soil moisture in Southeastern Australia. *IEEE Trans. Geosci. Remote Sens.* **50**, 1556–1571 (2012).
45. Merlin, O., Chehbouni, A., Kerr, Y. H. & Goodrich, D. C. A downscaling method for distributing surface soil moisture within a microwave pixel: Application to the Monsoon '90 data. *Remote Sens. Environ.* **101**, 379–389 (2006).
46. Ines, A. V. M., Mohanty, B. P. & Shin, Y. An unmixing algorithm for remotely sensed soil moisture. *Water Resour. Res.* **49**, 408–425 (2013).
47. Shin, Y. & Mohanty, B. P. Development of a deterministic downscaling algorithm for remote sensing soil moisture footprint using soil and vegetation classifications. *Water Resour. Res.* **49**, 6208–6228 (2013).
48. Zheng, J. *et al.* Soil moisture downscaling using multiple modes of the DISPATCH algorithm in a semi-humid/humid region. *Int. J. Appl. Earth Obs. Geoinf.* **104**, 102530 (2021).
49. Reichstein, M. *et al.* Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019).

50. Wei, Z., Meng, Y., Zhang, W., Peng, J. & Meng, L. Downscaling SMAP soil moisture estimation with gradient boosting decision tree regression over the Tibetan Plateau. *Remote Sens. Environ.* **225**, 30–44 (2019).
51. Alemohammad, S., Kolassa, J., Prigent, C., Aires, F. & Gentile, P. Global downscaling of remotely sensed soil moisture using neural networks. *Hydrol. Earth Syst. Sci.* **22**, 5341–5356 (2018).
52. Srivastava, P. K., Han, D., Ramirez, M. R. & Islam, T. Machine Learning Techniques for Downscaling SMOS Satellite Soil Moisture Using MODIS Land Surface Temperature for Hydrological Application. *Water Resour. Manag.* **27**, 3127–3144 (2013).
53. Meng, X. *et al.* A fine-resolution soil moisture dataset for China in 2002–2018. *Earth Syst. Sci. Data* **13**, 3239–3261 (2021).
54. Xu, W., Zhang, Z., Long, Z. & Qin, Q. Downscaling SMAP Soil Moisture Products with Convolutional Neural Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **14**, 4051–4062 (2021).
55. Liu, Y., Jing, W., Wang, Q. & Xia, X. Generating high-resolution daily soil moisture by using spatial downscaling techniques: a comparison of six machine learning algorithms. *Adv. Water Resour.* **141**, 103601 (2020).
56. Zheng, J. *et al.* Assessment of 24 soil moisture datasets using a new *in situ* network in the Shandian River Basin of China. *Remote Sens. Environ.* **271**, 112891 (2022).
57. Li, M., Wu, P. & Ma, Z. A comprehensive evaluation of soil moisture and soil temperature from third-generation atmospheric and land reanalysis data sets. *Int. J. Climatol.* **40**, 5744–5766 (2020).
58. Ling, X. *et al.* Comprehensive evaluation of satellite-based and reanalysis soil moisture products using *in situ* observations over China. *Hydrol. Earth Syst. Sci.* **25**, 4209–4229 (2021).
59. Dorigo, W. *et al.* ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sens. Environ.* **203**, 185–215 (2017).
60. Zhang, L., Zhao, T., Jiang, L. & Zhao, S. Estimate of phase transition water content in freeze-thaw process using microwave radiometer. *IEEE Trans. Geosci. Remote Sens.* **48**, 4248–4255 (2010).
61. Zhao, T. *et al.* A new soil freeze/thaw discriminant algorithm using AMSR-E passive microwave imagery. *Hydrol. Process.* **25**, 111958 (2011).
62. Zhao, T. *et al.* Soil moisture retrievals using L-band radiometry from variable angular ground-based and airborne observations. *Remote Sens. Environ.* **248** (2020).
63. Hersbach, H. *et al.* The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146** (2020).
64. Didan, K. MOD13A2 MODIS/Terra Vegetation Indices 16-Day L3 Global 1km SIN Grid. NASA EOSDIS Land Processes DAAC <https://doi.org/10.5067/MODIS/MOD13A2.006> (2015).
65. Didan, K. MOD13C1 MODIS/Terra Vegetation Indices 16-Day L3 Global 0.05Deg CMG. NASA EOSDIS Land Processes DAAC <https://doi.org/10.5067/MODIS/MOD13C1.006> (2015).
66. Liu, Q. *et al.* Preliminary evaluation of the long-term GLASS albedo product. *Int. J. Digit. Earth* **6**, 69–95 (2013).
67. Liu, N. F. *et al.* A statistics-based temporal filter algorithm to map spatiotemporally continuous shortwave albedo from MODIS data. *Hydrol. Earth Syst. Sci.* **17**, 2121–2129 (2013).
68. Becker, J. J. *et al.* Global Bathymetry and Elevation Data at 30 Arc Seconds Resolution: SRTM30_PLUS. *Mar. Geod.* **32**, 355–371 (2009).
69. Zhang, Y., Schaap, M. G. & Zha, Y. A High-Resolution Global Map of Soil Hydraulic Properties Produced by a Hierarchical Parameterization of a Physically Based Water Retention Model. *Water Resour. Res.* **54**, 9744–9790 (2018).
70. Hengl, T. *et al.* SoilGrids1km - Global soil information based on automated mapping. *PLoS One* **9**, e114788 (2014).
71. Gruber, A. *et al.* Validation practices for satellite soil moisture retrievals: What are (the) errors? *Remote Sens. Environ.* **244**, 111806 (2020).
72. Beck, H. E. *et al.* Evaluation of 18 satellite- And model-based soil moisture products using *in situ* measurements from 826 sensors. *Hydrol. Earth Syst. Sci.* **25**, 17–40 (2021).
73. Balenzano, A. *et al.* Sentinel-1 soil moisture at 1 km resolution: a validation study. *Remote Sens. Environ.* **263**, 112554 (2021).
74. Chen, Y., Feng, X. & Fu, B. An improved global remote-sensing-based surface soil moisture (RSSSM) dataset covering 2003–2018. *Earth Syst. Sci. Data* **13**, 1–31 (2021).
75. Duchon, C. E. & Hamm, K. G. Broadband albedo observations in the southern Great Plains. *J. App. Meteorol. Clim.* **45**(1), 210–235 (2006).
76. Sugathan, N., Biju, V. & Renuka, G. Influence of soil moisture content on surface albedo and soil thermal parameters at a tropical station. *J. Earth. Syst. Sci.* **123**, 1115–1128 (2014).
77. Zheng, C., Jia, L. & Zhao, T. Global daily surface soil moisture dataset at 1-km resolution (2000 - 2020). *National Tibetan Plateau/Third Pole Environment Data Center* <https://doi.org/10.11888/RemoteSen.tpd.c.272760> (2022).
78. Zheng, C., Jia, L. & Hu, G. Global Land Surface Evapotranspiration Monitoring by ETMonitor Model Driven by Multi-source Satellite Earth Observations. *J. Hydro.* **613**, 128444 (2022).
79. Zaussinger, F. *et al.* Estimating irrigation water use over the contiguous United States by combining satellite and reanalysis soil moisture data. *Hydrol. Earth Syst. Sci.* **23**, 897–923 (2019).
80. Brocca, L. *et al.* How much water is used for irrigation? A new approach exploiting coarse resolution satellite soil moisture products. *Int. J. App. Earth Observ. Geoinf.* **73**, 752–766 (2018).
81. Filippucci, P. *et al.* Soil moisture as a potential variable for tracking and quantifying irrigation: A case study with proximal gamma-ray spectroscopy data. *Adv. Water Resour.* **136**, 103502 (2020).
82. Li, Q. *et al.* A 1km daily soil moisture dataset over China using *in situ* measurement and machine learning. *Earth Syst. Sci. Data* **14**, 5267–5286 (2022).

Acknowledgements

This work is funded by the National Natural Science Foundation of China (NSFC) (Grant no. 42090014, 42171039, 41801346).

Author contributions

Chaolei Zheng: conceptualization, methodology, formal analysis, investigation, data curation, writing - original draft. Li Jia: writing—review and editing, funding acquisition. Tianjie Zhao: writing—review and editing.

Competing interests

The authors declare no conflict of interest for this article. The funding agencies had no role in the design of the study, the data processing and writing or in the decision to publish the results.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-023-01991-w>.

Correspondence and requests for materials should be addressed to C.Z.

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

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



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

© The Author(s) 2023