https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Open Access Earth System Science Data Discussions

A 1-km daily surface soil moisture dataset of enhanced coverage under all-weather conditions over China in 2003-2019

Peilin Song[1,4], Yongqiang Zhang[1], Jianping Guo[2], Jiancheng Shi[3], Tianjie Zhao[4], Bing Tong[2]

[1] Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, The Chinese Academy of Sciences, Beijing 100101, China

[2] State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081, China

[3] National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China

[4] State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences. Beijing 100101, China

Correspondence to: Yongqiang Zhang (yongqiang.zhang2014@gmail.com); Jianping Guo (jpguo@cma.gov.cn)

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Abstract:

Surface soil moisture (SSM) is crucial for understanding the hydrological process of our earth surface. Passive microwave (PM) technique has long been the primary tool for estimating global SSM from the view of satellite, while the coarse resolution (usually >~10 km) of PM observations hampers its applications at finer scales. Although quantitative studies have been proposed for downscaling satellite PM-based SSM, very few products have been available to public that meet the qualification of 1-km resolution and daily revisit cycles under all-weather conditions. In this study, we developed one such SSM product in China with all these characteristics. The product was generated through downscaling the AMSR-E/AMSR-2 based SSM at 36-km, covering all on-orbit time of the two radiometers during 2003-2019. MODIS optical reflectance data and daily thermal infrared land surface temperature (LST) that had been gap-filled for cloudy conditions were the primary data inputs of the downscaling model, so that the "all-weather" quality was achieved for the 1-km SSM. Daily images from this developed SSM product have quasi-complete coverage over the country during April-September. For other months, the national coverage percentage of the developed product is also greatly improved against the original daily PM observations, through a specifically developed sub-model for filling the gap between seams of neighboring PM swaths during the downscaling procedure. The product is well compared against *in situ* soil moisture measurements from 2000+ meteorological stations, indicated by station averages of the unbiased RMSD ranging from 0.052

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

vol/vol to 0.059 vol/vol. Moreover, the evaluation results also show that the developed product outperforms the SMAP-Sentinel (Active-Passive microwave) combined SSM product at 1-km, with a correlation coefficient of 0.55 achieved against that of 0.40 for the latter product. This indicates the new product has great potential to be used for hydrological community, agricultural industry, water resource and environment management.

1. Introduction

Surface soil moisture (SSM) is one of the most important variables that dominate the mass and energy cycles of earth surface system. Satellite-based SSM datasets of sufficiently fine spatio-temporal resolutions over large-scale areas have significant implication on improved investigations at various research fields including hydrological signature identification, agricultural yield production estimation, drought/waterlogging monitoring and warning, as well as weather prediction and future climate analysis. Microwave bands with centimeter-level or longer wavelengths (X-band, C-band, and L-band) are currently identified as the primary band channels suitable for SSM observations from view of satellite, due to their high penetration capabilities through cloud layers and vegetation canopies. In terms of sensor types, microwave SSM detection includes passive microwave (radiometer-based) techniques and active microwave (radar, scatterometer) techniques. Satellite-based passive microwave (PM) radiometers, e.g. the Soil Moisture Active Passive (SMAP), the Soil Moisture and Ocean Salinity (SMOS), and the Advance Microwave Scanning

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Radiometer-2 (AMSR-2), can obtain SSM observations at a revisit interval of 1-3 days, with relatively poor native spatial resolutions of tens of kilometers. Active microwave (AM) such as radar can achieve kilometer-level and even finer resolution of observations targeting at the earth surface. However, this usually sacrifices the swath width of radar configuration, because of which, most satellite-based synthetic aperture radars (SAR) have an obviously longer global revisit cycle (usually longer than 5 days, e.g. Sentinel-1 SAR data) than the typical radiometers. Moreover, AM radar backscatter signals are extremely sensitive to speckle noise (Entekhabi et al., 2016), as well as influence from soil roughness, vegetation canopy structure and water content (Piles et al., 2009). All above influential factors have seriously impeded the use of AM radar techniques or combination of passive/active microwave datasets for producing high spatial resolution SSM products with a frequent revisit.

Apart from microwave signals, solar reflectance or ground emission signals originated from optical and infrared band domains also have the potential to reflect SSM variation. Based on optical/infrared bands, however, SSM is typically estimated based on indirect relationships through intermediate variables like soil evaporation, vegetation condition, or soil thermal inertia. To overcome the spatio-temporally instable performance on SSM modelling that might be brought by such indirect relationships, they are typically fused with the PM SSM datasets. In this manner, it can well reconcile the advantage of PM observations with respect to its high sensitivity to SSM variation, as well as that of optical/infrared observations with respect to its finer spatial resolutions at kilometer- or even hectometer-levels. Such data fusion techniques

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

are also known as downscaling techniques of PM remote sensing SSM. Archetypal downscaling models include the "triangle feature space (UTF)"-based models (Chauhan et al., 2003; Choi and Hur, 2012; Sanchez-Ruiz et al., 2014), the "DISaggregation based on a Physical And Theoretical scale CHange (DISPACTH)" model (Merlin et al., 2010; Merlin et al., 2005; Merlin et al., 2013; Merlin et al., 2008), and the UCLA model (Peng et al., 2016). The physics of these models are mainly based on the response of SSM variation to changes in soil evaporation or land surface evapotranspiration. Another significant branch of such downscaling models are based on the sensitivity of SSM to soil thermal inertia, which is quantified by diurnal LST difference estimated from thermal-infrared wave bands (Fang and Lakshmi, 2013; Fang et al., 2018).

Sabaghy et al. (2020) have shown that using optical and infrared data can achieve finer-resolution SSM estimates which are better consistent with ground soil moisture records, compared with using the radar datasets. Moreover, considering the short revisit cycle (daily) of optical/infrared sensors onboard typical polar-orbit satellites, using optical/infrared datasets to downscale PM SSM should be among the optimal methods for obtaining SSM data with high spatio-temporal resolutions over national, continental, or global scales. On the other hand, satellite remote sensing SSM products that are characterized by 1-km resolution of daily revisit intervals and stable long time series dating back to at least 15-20 years ago, are urgently required for accelerating the development of various research fields, especially agriculture industry, water resources management, and hydrological disaster monitoring (Sabaghy et al., 2020; Mendoza et

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

al., 2016). However, very seldom sets of such data products are publicly available to the remote sensing research community because of the following drawbacks,. First, there is a serious lack of cloud-free optical/infrared imagery, which means the method cannot deliver any SSM downscaling under cloudy/rainy weather. Second, most of the above-mentioned optical/infrared-data-based downscaling methods were mainly evaluated at regional or even smaller scales. This might raise concern on the universality of those methods. For example, the DISPATCH method has been recognized to be less effective in humid (energy-limited) regions compared with in arid and semi-arid (water-limited) regions (Molero et al., 2016; Song et al., 2021; Zheng et al., 2021), whilst another experiment (Kim and Hogue, 2012) shows that the UTF-based methods are found even inferior to the DISPATCH in a typical water-limited region in North America.

To improve the above-mentioned issues, we produced all-weather daily SSM data products at 1-km resolution all over China during 2003-2019, based on fusion of multiple remote sensing techniques, including reconstruction of optical/infrared observations under cloud as well as an improved PM SSM downscaling methodology proposed in our previous study (Song et al., 2021). The major objectives of this study include

(i) to better serve and investigate the land surface hydrology processes and their sophisticated interactions to human society at multi-scale (from national to regional) resolutions in China because the country covers about 1/15 of the global terrestrial area with about 1/5 of the world population, and

(ii) to provide a methodology framework that can inspire future studies on generating similar SSM datasets all over the globe, based on the plentifulness of resources on climate type, land covers, and topography in China.

2. Methods and Materials

2.1 Datasets

2.1.1 PM SSM data

Spatial downscaling of PM SSM is the fundamental theory for constructing the target finer-resolution data product in this study. Therefore, the native retrieval accuracy of the coarse-resolution PM SSM data product, based on which the downscaling procedures are performed, is considerably crucial to the performance of the downscaled data product (Busch et al., 2012; Im et al., 2016; Kim and Hogue, 2012). Although the L-band PM brightness temperature (TB) observed by satellite missions such as SMAP or SMOS are considered more suitable for SSM retrieval compared with C- or X-band TB, both above missions started their space operations after 2010s. This means that to obtain downscaled SSM of longer historical periods, we still require to rely on other C-/X-band-based radiometers which started their operations earlier than SMAP and SMOS. An optimal satellite PM TB observation system dating back to earlier years of this century is composed of the "Advanced Microwave Scanning Radiometer of the Earth Observing System (AMSR-E)", together with its successor of AMSR-2. AMSR-E operated during 2002-2011 onboard the Aqua satellite which is

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

governed by National Aeronautics and Space Administration (NASA), whilst AMSR-2 is operating onboard the Global Change Observation Mission1-Water (GCOM-W1) satellite developed by the Japan Aerospace Exploration Agency (JAXA) since 2012.

Several classical PM SSM retrieval algorithms have been applied to the afore-mentioned "AMSR series (including AMSR-E and AMSR-2)" TB for generating long-term global SSM products at 25 km (Table 1), including the JAXA algorithm (Fujii et al., 2009; Koike et al., 2004), the "Land Parameter Retrieval Model (LPRM)" algorithm (Song et al., 2019b; Meesters et al., 2005; Owe et al., 2001), and the algorithm developed by the University of Montana (UMT) (Jones et al., 2009; Du et al., 2016). A recent AMSR-based night-time SSM product during 2002-2019 has been produced through a neural network trained against SMAP descending SSM (hereafter referred to as "NN-SM product") (Yao et al., 2021). The global validation results show that this NN-SM product is better than the JAXA and LPRM products.

Besides, the NN-SM has also been compared with another long-term ~25-km all-weather SSM dataset generated through the European Space Agency (ESA)'s Climate Change Initiative (CCI) program. The ESA-CCI SSM product is different from the rest products mentioned above in that it was implemented by fusion of observations from comprehensive AM- and PM-based satellite sensors, rather than only relying on the radiometers of AMSR series. According to Yao et al. (2021), the ESA-CCI SSM has slightly better validation accuracy than the NN-SM product, but the number of available observations per pixel cell in an entire year is much smaller for the ESA-CCI SSM in Southeast China. In view of all above coarse-resolution SSM data products, we finally

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Open Access Earth System Science Data Discussions

selected the NN-SM product to implement the following spatial downscaling procedures rather than the ESA-CCI SSM, to make a balance between data accuracy and data availability per year. We have also made additional evaluations within China in Section Appendix-A to ensure the relatively outstanding performance of the NN-SM product as described above.

Table 1 Information of all-weather microwave remote sensing coarse-resolution SSM data products that can be potentially downscaled to obtain high resolution SSM.

Name	**Resolution**	**Satellite radiometers involved**	**Data availability (url)**
NN-SM product	36 km (by the EASE Grid projection)	AMSR-E/ AMSR-2 (2002-2011, 2012-present)	https://data.tpdc.ac.cn/en/data/c26201fc-526c-465d-bae7-5f02fa49d738/
ESA-CCI v6.1 product	0.25°	AMSR-E/ AMSR-2/ SMOS/ WindSat/ SMMR/ SSM/I/ TMI (1978-2020)	https://www.esa-soilmoisture-cci.org/v06.1_release
JAXA product	0.25° / 0.1°	AMSR-E/ AMSR-2 (2002-2011, 2012-present)	https://gportal.jaxa.jp/
LPRM product	0.25° / 0.1°	AMSR-E/ AMSR-2 (2002-2011, 2012-present)	https://search.earthdata.nasa.gov/
UMT product	25 km (by the EASE Grid projection)	AMSR-E/ AMSR-2 (2002-present)	http://files.ntsg.umt.edu/data/LPDR_v2/

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

2.1.2 Optical remote sensing data and digital elevation model (DEM)

Optical remote sensing datasets provide finer spatial texture information on the daily basis for the downscaling purpose of PM SSM. Such data that can be used as inputs of our SSM product processing line are mainly provided by the Moderate-resolution Imaging Spectroradiometer (MODIS) onboard the Terra and Aqua satellites. Specifically, they involve the 1-km daily night-time Aqua MODIS LST product (MYD21A1N.v061) and the 500-m daily "Bidirectional Reflectance Distribution Function (BRDF)" - Adjusted Reflectance dataset (MCD43A4.v061). MYD21A1 LST data can be recognized as a crucial proxy of land surface thermal capacity (Fang et al., 2013) and soil evaporative rate (Merlin et al., 2008). The MCD43A4 nadir reflectance product, with view angle effect corrected using the BRDF model, is capable to provide observations from visible to shortwave-infrared bands that can characterize water content variation of the bare soils as well as the vegetation canopy. Overall, the above-mentioned datasets were selected primarily because they deliver indicators (land surface thermal capacity, soil evaporative rate, or vegetation condition) that can well response to soil moisture dynamics from different aspects. Prior to be employed for SSM downscaling, conventional pre-processing procedure of pixel quality check was applied for both optical products by screening out pixels not classed as "good quality", according to the 8-bit "Quality Assessment (QA)" field of each spectral band. Moreover, to normalize their natively different spatial resolutions, all MCD43A4 based reflectance values at the 500-m pixel level were upscaled to the sinusoidally projected MODIS 1-km grids using their spatial averages.

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Apart from MODIS optical remote sensing data, all 90-m DEM tiles generated by the NASA Shuttle Radar Topography Mission (SRTM; http://srtm.csi.cgiar.org/, last access: July 10, 2020) were mosaicked over the entire China and then employed as another essential input variable for the procedures as described by Section 2.2.2 below. Similar to that applied to the MCD43A4 product, spatial upscaling in correspondence to the MODIS 1-km grids is also an indispensable pre-processing step for the mosaicked DEM data.

2.1.3 Ground validation data

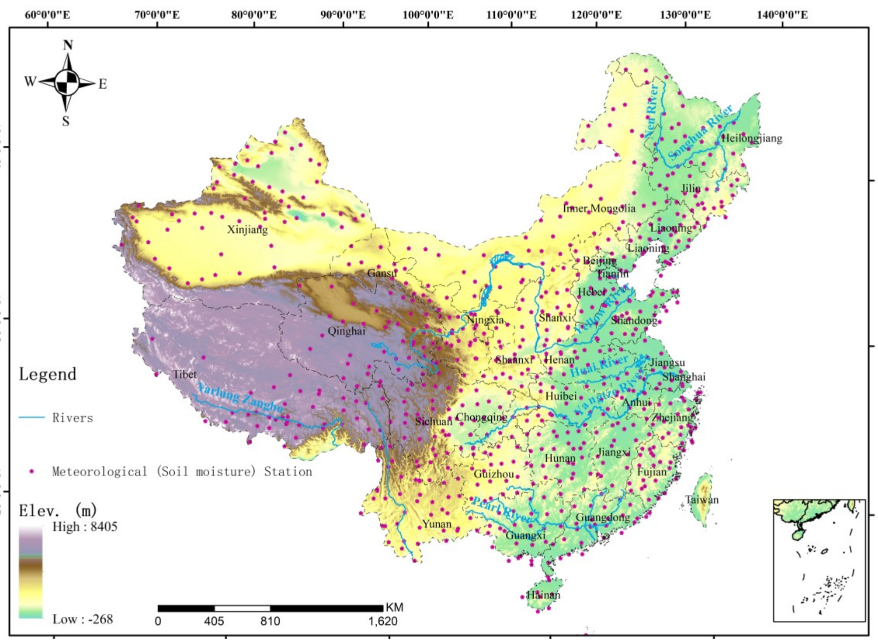


Fig. 1 The provincial-level administration map of China superposed with topographic information, as well as general locations for the 756 basic meteorological stations (http://data.cma.cn/, last access:

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

January 20, 2021) that provide partial benchmark measurements for SSM and LST validation in this study.

We utilized ground soil moisture measurements for validating the downscaled remote sensing SSM product. The ground measurements are derived from 2417 meteorological stations (including 756 basic stations of the National Climate Observatory and 1661 regionally intensified stations) of over China, as partially shown in Fig. 1. The soil moisture measurement devices in these stations, with uniform observation standards, are instrumented under the national project of "Operation Monitoring System of Automatic Soil Moisture Observation Network in China (Wu et al., 2014)", the construction of which has been led by China Meteorological Administration since 2005. Until 2016, all stations have been in operation for automatically observing hourly in situ soil moisture dynamics at eight different depth ranges (0-10 cm, 10-20 cm, 20-30 cm, 30-40 cm, 40-50 cm, 50-60 cm, 70-80 cm, 90-100 cm). In our current study, ground measurements matching the shallowest depth range (0-10 cm) from the initial time of each station until the end of 2019 are employed as validation benchmark of the satellite SSM retrievals. At the temporal dimension, measurements made at 1:00 A.M. and 2:00 A.M are averaged, in order to match the mean satellite transit time of 1:30 A.M. for AMSR descending observations.

Moreover, 0-cm top ground temperatures are simultaneously measured at all these meteorological stations on the daily basis, at the local time windows of 2:00 A.M./P.M. and 10:00 A.M./P.M., respectively. We therefore exploited such measurements recorded at 2:00 A.M. to validate the cloud gap-filled night-time (~1:30 A.M.) LST

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

estimates over the Aqua-MODIS based 1-km pixels containing these stations (see Section 2.2.2). Our primary validation period covers the entire years of 2017, 2018, and 2019.

2.1.4 Ancillary SSM products for comparison

In order to comprehensively demonstrate the validation performance of our proposed SSM product, there is necessity to make an inter-comparison against similar existing datasets. In this regard, we introduced the Level2 SMAP/Sentinel Active-Passive combined SSM product on 1-km earth-fixed grids, i.e., the SPL2SMAP_S_V3 dataset (Das et al., 2020), and used its validation performance against in-situ measurements throughout the years of 2017, 2018, and 2019, as a baseline to better evaluate our proposed SSM product. The SPL2SMAP_S_V3 dataset contains global SSM at resolutions of 3 km and 1 km respectively, which were disaggregated from the SMAP SSM retrievals of 36-km/9-km footprints in conjunction with the high-resolution Sentinel-1 C-band radar backscatter coefficients (Das et al., 2019). To our knowledge, this dataset is possibly the only publicly available product which can provide global remote sensing SSM estimates at the 1-km resolution. The sentinel backscatter coefficient inputs for this product are only those received in the descending orbit scenes (at ~6:00 A.M. of local time), whilst the closest SMAP SSM retrievals from either ascending (at ~6:00 P.M. of local time) or descending orbits are used to spatially match up with the sentinel-1 scene. It is noticed that at the descending observation time the soil moisture vertical profile has approached a hydrostatic balance

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

(Montaldo et al., 2001), thereby providing the optimal chance for soil moisture fusion and validation with observations at different soil depths. Therefore, we only selected the 1-km disaggregated SSM estimates based on descending SMAP SSM retrievals (i.e., the subset with field name of 'disagg_soil_moisture_1 km' in the SPL2SMAP_S_V3 dataset). Meanwhile, the 10-cm-depth in-situ soil moisture measurements observed at 6:00 A.M. were employed as the validation benchmark, in a manner similar to that applied to our proposed SSM product (Section 2.1.3).

2.2 Methodology

The general methodological framework for producing the all-weather daily 1-km SSM product is shown as in Fig. 2, with details described in the following context of this section.

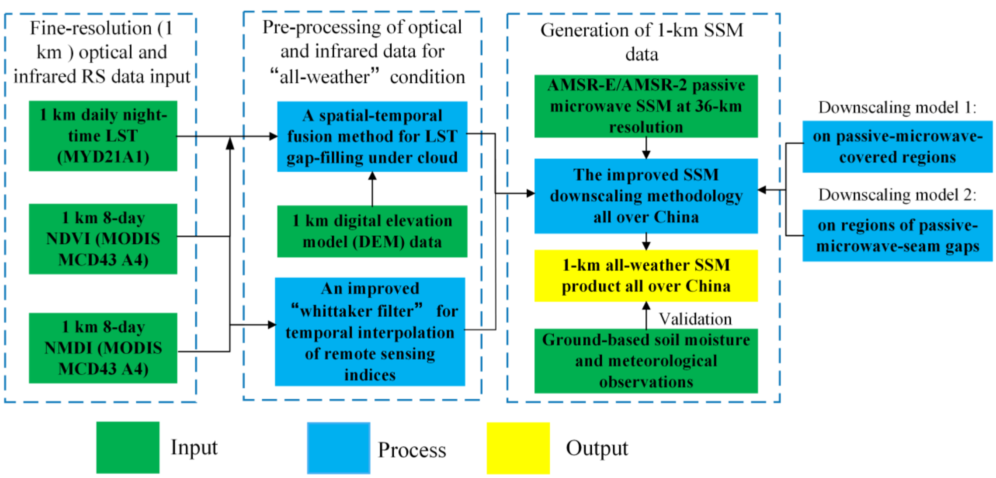


Fig. 2 The overall methodological framework of this study.

2.2.1 Reconstruction of thermal-infrared LST and remote sensing (vegetation) indices under cloud

Reconstruction of missing pixels under cloud in the optical remote sensing input datasets is the prerequisite for achieving the "all-weather" property of the final

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

downscaled SSM output. For reconstructing thermal-infrared LST, we adopted the cloud gap-filling method as proposed by our previous study (Song et al., 2019a). This method, also referred to as a typical "spatio-temporal data fusion" (STDF) method (Dowling et al., 2021), was built using clear-sky LST observations of spatially neighboring pixels observed at proximal dates, with concurrent NDVI and DEM also employed as additional data inputs. The STDF method can be expressed as follows:

$$LST^{*}_{t_1} = a \times LST^{*}_{t_0} + b \times NDVI^{*}_{t_1} + c \times DEM^{*} + d \quad (1)$$

Where the superscript "*" indicates that this variable has been normalized to the range 0 to 1.0 (Song et al., 2019a), based on the maximum and minimum values of that variable found across China (excluding invalid values representing states of snow, ice, and water bodies). Parameters *a, b, c*, and *d* are coefficients fitted between all pixels with clear-sky LST estimates on a specific date t_1 (LST^*_{t1}) and their counterparts on one proximal date, t_0 (LST^*_{t0}). $NDVI^*_{t1}$ indicates the corresponding (normalized) NDVI on the t_1 date calculated using the MCD43A4 daily product. After deriving the coefficients of *a, b, c,* and *d*, Equation (1) was used to fill all cloudy MODIS LST pixels on the t_1 date. For any t_1 date included in the study period, the t_0 date was iterated among all neighboring dates of t_1 meeting the condition | t0- t1|<=30 (from the nearest date to the furthest date). The average of estimated LST values for t_0 was then taken where a cloud gap pixel was filled more than once (based on the iterative t_0 dates). The iteration was stopped when the fraction of pixels with effective LST values on t_1 was equal to or exceeded 0.99.

An important flaw of this STDF method should be noticed with regard to potentially existential bias of the cloud gap-filled LST outputs, because the outputs represent theoretically reconstructed LST under clear sky rather than under the real cloudy condition. Another of our previous studies (Dowling et al., 2021) concerning

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

this STDF method proposed a follow-up step, which incorporated PM-derived surface temperature, to adjust that bias. In our current production pipeline, however, this follow-up step for cloud bias adjustment in LST was not carried out. This is mainly because the gap-filled LST outputs are intended for SSM downscaling. The downscaling techniques as proposed in Section 2.2.1 was developed based on the "universal triangle feature (UTF)" theory (Carlson et al., 1994). In the UTF, clear-sky LST was employed to quantify the land surface evaporation when vegetation cover density was fixed. The degree of land surface wetness was then predicted implicitly through soil evaporation degree and surface soil thermal inertia. Under cloudy conditions, however, the satellite observed LST would be a proxy of not only surface soil property, but also of that related to cloud liquid water and crystals in the atmospheric layers. In comparison, therefore, LST generated by the STDF alone for clear-sky conditions would be a more competent input variable for quantifying surface soil wetness under cloudy conditions. We have made additional evaluations to confirm the validity of this assumption, with the results elaborated in Section Appendix-B of this paper.

Reconstruction of the remote sensing vegetation indices under cloudy conditions, including NDVI and MNDI, was simply based on the modified time series filter of the Whiitaker Smoother (MWS) as developed by Kong et al. (2019). This is reasonable because the dynamic trends of vegetation growth are relatively less volatile compared to LST on the daily basis, and can thus be gap-filled for missing values using a time-series-filtering-like algorithm.

2.2.2 Improved downscaling technique of SSM based on fusion of PM and optical/infrared data

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

The core component of the SSM downscaling methodology is an improved linking model between PM SSM and (fine-resolution) optical remote sensing observations. Since the model origins from our previous study (Song et al., 2021), herein we simply give its mathematical expression as follows:

$$SSM = \frac{a \times \ln(1 - SEE)}{1 - b \times NMDI} + c \tag{2}$$

In Equation (2), *SEE* is a mathematical function of LST and the typical Normalized Difference Vegetation Index (NDVI), with its specific form described in Merlin et al. (2008). NMDI is another remote sensing index calculated as $\frac{R_{infr,860nm}-(R_{sw,1600nm}-R_{sw,2100nm})}{R_{infr,860nm}+(R_{sw,1600nm}-R_{sw,2100nm})}$ (Wang and Qu, 2007). $R_{infr,860nm}$, $R_{infr,1600nm}$, and $R_{infr,2100nm}$ represent land surface reflectance signals derived from three different MODIS-MCD43A4 based near-infrared/shortwave-infrared bands, with their wavelengths centering at 860 nm, 1600 nm, and 2100 nm respectively. The parameters *a, b,* and, *c* are empirical coefficients that represent background information of local soil texture and vegetation types. In Song et al. (2021), these coefficients have been fitted and calibrated based on multi-temporal observations at the PM pixel scale. In our current study, however, we have discovered that coupling of multiphase observations at both the spatial and the temporal dimensions can lead to more optimal solution of the coefficients, as they can produce downscaled SSM images with notably declined effect of 'mosaic' against the original PM 36-km pixels. Therefore, the modified optimal cost function χ^2 for deriving these coefficients is re-defined as follows:

$$\chi^2 = \sum_{d=-5}^{5} \sum_{i=0}^{N=7\times7} w_i \times (SSM_{ob,i,d} - SSM_{\mathrm{mod},i,d})^2 \tag{3}$$

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Through the cost function, the spatial extent of each 36-km pixel P_0 on any arbitrary date D_0 obtains a unique set of coefficients. As shown by Equation (3), all pixels within the N=7×7 spatial window centered at P_0 ranging from -5*th* day to 5*th* day relative to the date of D_0 were exploited. SSM_{ob} and SSM_{mod} denote the AMSR NN-SM 36-km SSM observations as well as SSM observations modelled by Equation (2) based on upscaled optical datasets, respectively. w_i is a weight coefficient used to ensure that neighboring observations near the centering pixel P_0 play more dominating roles as compared with the far-end pixels in the cost function, considering the "Tobler's First Law of Geography (Sui, 2004)" . w_i is calculated using an adaptive bi-square function:

$$\begin{aligned} &w_i = [1-(\frac{dis_i}{b})^2]^2, dis_i < b \\ &w_i = 0, dis_i >= b \end{aligned} \tag{4}$$

where dis_i indicates the distance between the i-*th* pixel and the centering pixel P_0. b is named as the adaptive kernel bandwidth of the bi-square function (Duan and Li, 2016), and is optimized as 200 km through using a cross validation method as recommended by Brunsdon et al. (1996).

With the linking model obtained, we can subsequently utilize the spatial downscaling relationship function to produce 1-km high resolution SSM. The downscaling relationship function is constructed by transforming the linking model into its Taylor expansion formula and preserving all components with respect to the input optical variables of the linking model at first and second orders. This relationship is inspired from Malbéteau et al. (2016) and Merlin et al. (2010), and is mathematically described below:

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

$$SSM_{1\text{-}km}=SSM_{36km}+\left(\frac{\partial SSM}{\partial SEE}\right)_{36km}\times(SSE_{1km}-<SSE>_{36km})+0.5\times\left(\frac{\partial^2 SSM}{\partial SEE^2}\right)\times(SSE_{1km}-<SSE>_{36km})^2+\left(\frac{\partial SSM}{\partial NMDI}\right)_{36km}\times(NMDI_{1km}-<NMDI>_{36km})+0.5\times\left(\frac{\partial^2 SSM}{\partial NMDI^2}\right)\times(NMDI_{1km}-<NMDI>_{36km})^2 \quad (5)$$

In the above relationship, <> denotes the operator of spatial averaging disaggregation for the 1-km optical remote sensing input variables at the corresponding 36-km pixel, $\frac{\partial SSM}{\partial SEE}\left(\frac{\partial^2 SSM}{\partial SEE^2}\right)$ and $\frac{\partial SSM}{\partial NMDI}\left(\frac{\partial^2 SSM}{\partial NMDI^2}\right)$ respectively denoting the first-(second-) order partial derivative of the linking model described in Equation (2).

It should be noticed that there exist middle-/low-latitude gap regions between seams of neighboring daily AMSR-E(-2) swaths, indicating that SSM_{36km} in Equation (5) is not always available on the daily basis (Song and Zhang, 2021a). For such PM-seam gaps on a particular date t_0, the corresponding $SSM_{36km,t0}$ in Equation (5) is substituted by $0.5\times(SSM_{36km,t0+1}+SSM_{36km,t0-1})+\Delta SSM_{36km,t0}$. Herein $SSM_{36km,t0-1}$ and $SSM_{36km,t0+1}$ respectively denote the SSM estimate before and after the date of t_0. $\Delta SSM_{36km,t0}$ is a component for correcting inter-day bias, with the following expression:

$$\Delta SSM_{36km,t0}=SSM\left(SEE_{36km,t0},\ NMDI_{36km,t0}\right)-0.5\times\left(SSM\left(SEE_{36km,t0-1},\ NMDI_{36km,t0-1}\right)+SSM\left(SEE_{36km,t0+1},\ NMDI_{36km,t0+1}\right)\right) \quad (6)$$

In the above equation, $SSM(SEE_{36km},\ NMDI_{36km})$ denotes SSM that is directly modelled based on Equation (1) using 36-km SEE and NMDI. The 36-km SEE and NMDI are obtained via averaging the variables spatially from their native resolution at 1-km. If all $SSM_{36\text{-}km}$ during the three consecutive days (t_0-1, t_0, and t_0+1) are missing due to other extreme conditions like snow, ice, or surface dominated by substantially large water bodies, the downscaling process cannot be fulfilled and all 1-km sub-pixels with the $SSM_{36\text{-}km}$ have to be set as null values.

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

2.2.3 Evaluation metrics

We employed the classic metrics of 'Root Mean Square Difference (RMSD)' and correlation coefficient (r-value) for evaluating satellite-based (SSM and LST) estimates against ground measurements. Herein RMSD is not referred to as 'Root Mean Square Error (RMSE)', although the latter term shares the same definition and has been used more commonly in previous studies. This is because the ground benchmark data may also present measurement uncertainties in practice. For SSM evaluation, the unbiased RMSD, or ubRMSD (Entekhabi et al., 2010; Molero et al., 2016), is calculated instead of RMSD in order to better investigate the time series similarity between satellite and ground soil moisture datasets by eliminating the systematic bias caused by spatial scale mismatch between them.

The above-mentioned classic metrics are primarily suitable to evaluate the absolute reliability of an independent remote sensing product. However, we also require another metric for characterizing the relative improvement of the downscaled SSM estimates against the original PM observations on capturing local soil moisture dynamics. For this purpose, we employed the "gain metric" of G_{down}, which was developed particularly by Merlin et al. (2015) for assessment of soil moisture downscaling methodology. G_{down} is a comprehensive indicator for evaluating gains of the downscaled SSM against the original coarse-resolution PM data in terms of their mean bias, bias in variance (slope), and time series correlation with ground benchmark. It has a valid domain between -1 and 1, with positive (negative) value indicating improved (deteriorated) spatial representativeness of the downscaled SSM against the

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

original PM data. Detailed definition and introduction of G_{down} are given in Equation (8) and Section 3.3 of Merlin et al. (2015).

3. Results

3.1 Evaluation on reconstructed thermal-infrared LST under cloud

The meteorological-station-based validation of reconstructed 1-km thermal-infrared LST under cloud were preliminarily fulfilled, to ensure the high quality of input dataset variables for SSM downscaling. Since negative effects might be brought to this validation campaign by the potentially existing heterogeneity of the validated 1-km thermal-infrared remote sensing pixels, we firstly analyzed correlations between estimated and benchmark datasets at each station, only based on satellite remote sensing observations obtained under clear sky. Stations that have their correlation coefficients (r_{clr}) lower than 0.9 herein have to be screened out because there exist higher chances of cross-scale spatial mismatch within and around these stations in terms of the land surface thermal properties. Among all 2417 stations (see Section 2.1.3) where 0-cm in-situ top-ground temperature measurements were available, we finally preserved 2107 stations characterized by r_{clr} >0.9. In the subsequent step, remote sensing LST under cloud and under clear-sky conditions were respectively validated at these stations, with the results revealed in Fig. 3. It is manifested through Fig. 3-(a) and -(b) that very close performances have been achieved between the clear-sky and the cloudy scenarios, especially considering their almost equally high validating correlations between 0.94-

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

0.95. For each independent station, we calculated the "RMSD difference (RMSD_diff)" between the two scenarios, based on the formula of "$RMSD_{clr}$- $RMSD_{cld}$ (the subscripts of *'clr' and 'cld'* denote clear-sky and cloudy conditions separately)". The statistical distribution of this RMSD difference with regard to different stations is shown in Fig. 3-(c). Apparently, 1942 stations all over the country have obtained an RMSD difference value below 2.6 K, and the mean RMSD difference is only about 1.9 K. All above results have indicated small uncertainty of our night-time LST reconstruction algorithm proposed for cloudy conditions. The corresponsive uncertainty that could be propagated to downscaled SSM in this stage is analyzed below in Section 3.2.

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

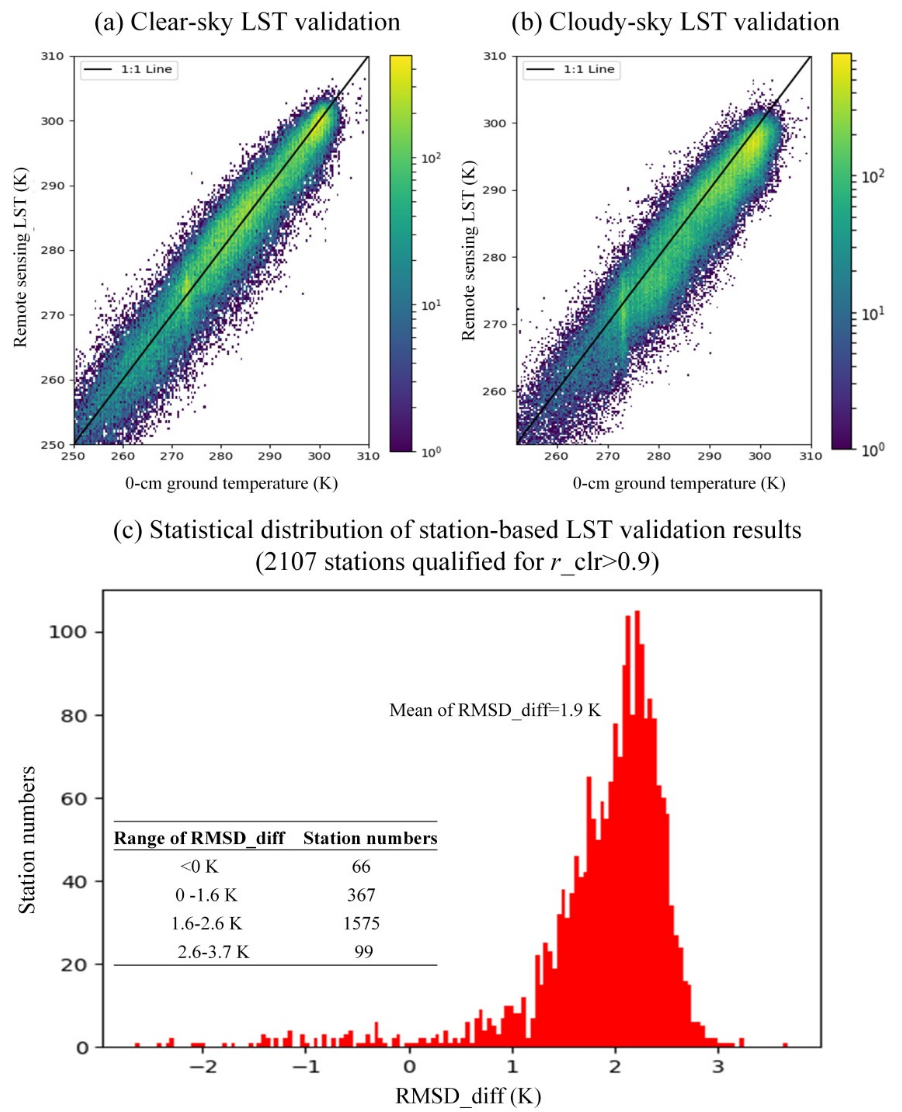


Fig. 3 validation results of the cloud gap-filled LST in China. (a) Density plot of thermal infrared LST under clear-sky condition compared to the 0-cm ground temperature measurements for all stations. (b) Same to (a) but for thermal infrared LST under cloudy conditions. (c) Statistical distribution of difference between RMSD of clear-sky LST and RMSD of gap-filled LST under cloudy condition with regard to different meteorological stations over the study region.

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

3.2 Evaluation on the final 1-km SSM product

The overall validation results of the finally downscaled 1-km SSM product is demonstrated in Fig. 4. Fig. 4-(a) shows that about 85% (N: 1833) of the total 2154 stations (the remaining 263 stations are located in pixels with no effective PM observations and are thus removed) have obtained significantly positive downscaling gains (G_{down}>0.03). This hints that the 1-km SSM product can better capture the dynamic behaviors of local ground soil moisture data than the original 36-km PM NN-SM data, revealing higher spatial representativeness of the downscaled SSM data product over the country. According to Fig. 4-(b), the mean ubRMSD of all stations is about 0.054 vol/vol, while 90% of those stations have the number lower than 0.088 vol/vol. In addition, we made another analysis concerning the possible influence of land cover types on SSM downscaling performance in Fig. 4-(c). The spatial information of land cover types was derived from the MODIS MCD12Q1 (10.5067/MODIS/MCD12Q1.006) IGBP-based land use image in 2019. For stations that experienced land use change throughout the years of the study period, the ubRMSD is only reported for data in the year of 2019. Clearly, better accuracies are observed mainly in grassland, cropland and bare soil surface, whilst relatively poorer performances (with averages of ubRMSD higher than 0.06 vol/vol) are seen in urban regions, (woody) savanna, and crop-to-natural-vegetation mosaic areas. Such a relative performance across land covers is logical because all the land cover types with their average ubRMSD higher than 0.06 vol/vol are characterized by lower hydrologic homogeneity in terms of their definition, e.g. savanna, which is a mixture of grass and tall trees, and urban areas, which are composed of impervious underlying surface.

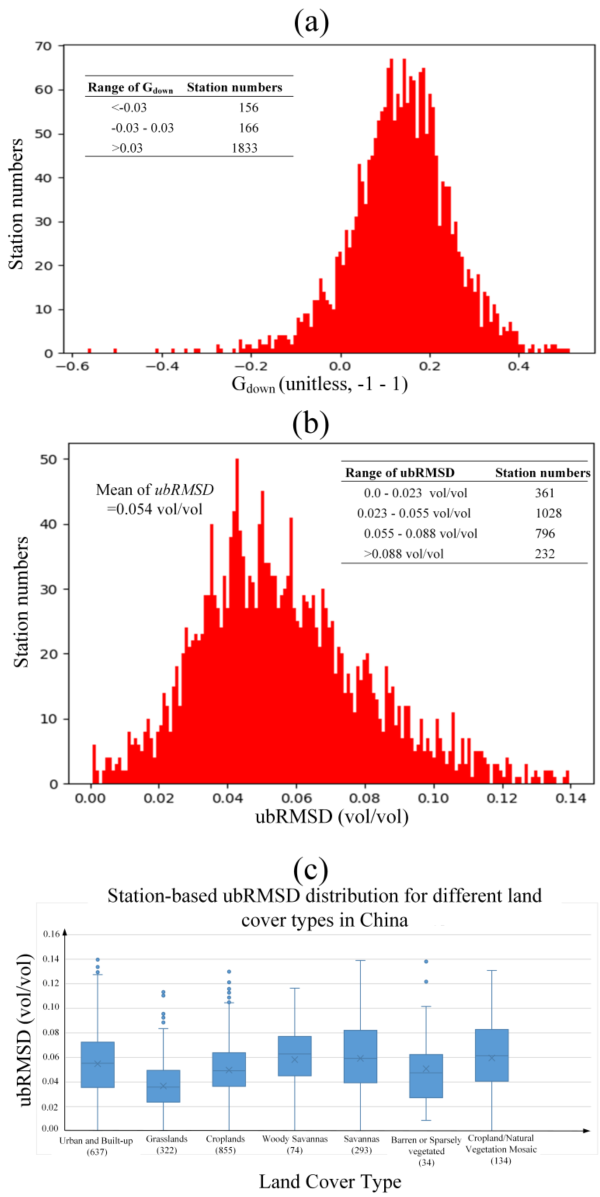


Fig. 4 General validation results of the currently developed SSM product. (a) G_{down} distribution for different stations over China. (b) ubRMSD distribution for different stations over China. (c) ubRMSD

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

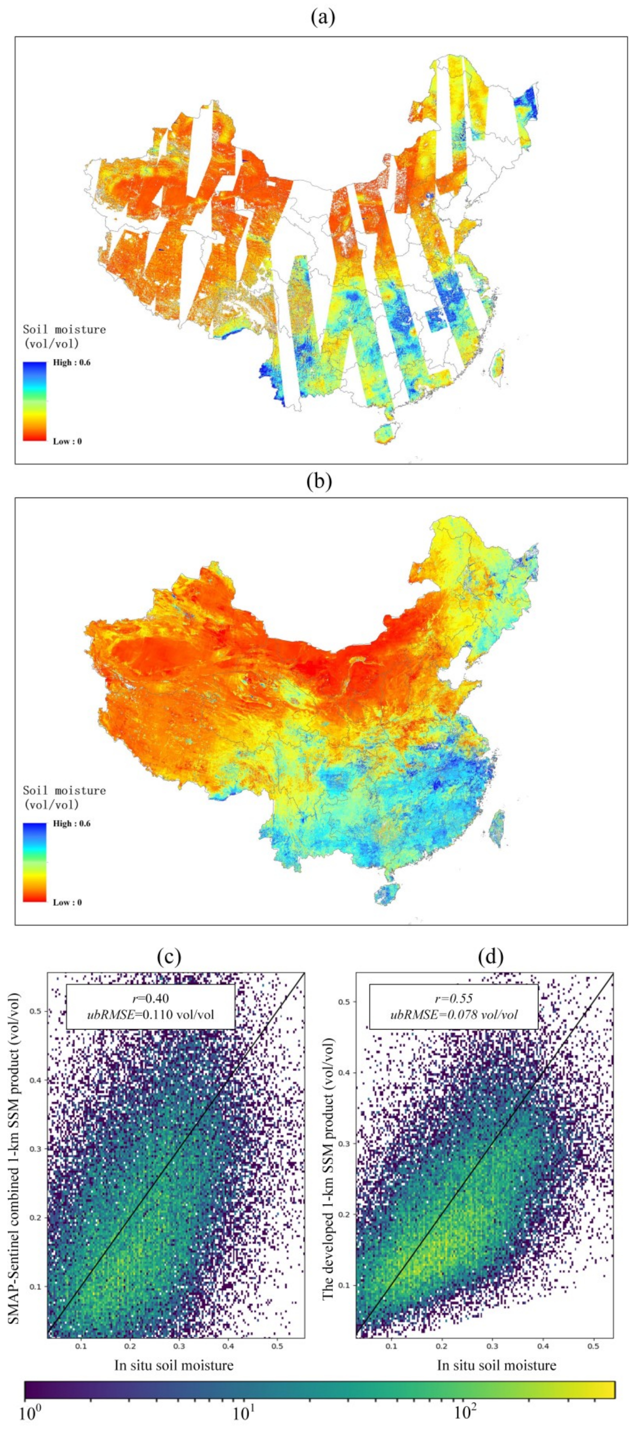
Open Access Earth System Science Data Discussions

statistics reported for different land covers. The numbers in the parentheses of the x-axis labels represent the amount of meteorological stations corresponding to that specific land cover type.

In Fig. 5-(b) we employed the downscaled SSM image on April 9, 2018, as an example to demonstrate the spatial features of the developed product. Meanwhile, we also show the map of SMAP/Sentinel combined SSM (SPL2SMAP_S_V3) obtained from April 6 to April 11, 2018 in Fig. 5-(a), as a contemporaneous comparison reference. Clearly, the SPL2SMAP_S_V3 map has a much lower coverage percentage over the study region compared with the map of the currently developed product on one single date, even though the former was generated based on multi-date images. Both maps show similar spatial texture depicting the relatively dry climate in northwestern China compared with the humid climate in the Middle-lower Yangtze River Plain. Nevertheless, there also exist cases where the details in texture differ prominently, like that in the far northeastern end of the country. For the sake of further analysis on this point, results of the quantitative comparison as proposed in Section 2.1.4, is demonstrated in Fig. 5-(c) and Fig. 5-(d). The currently developed SSM product obtained a 0.078 vol/vol ubRMSD and a correlation coefficient of 0.55 against the in-situ soil moisture measurements, converging more apparently to the 1:1 line when compared with validation result of the SPL2SMAP_S_V3 dataset. As with the area of China, therefore, the currently developed product is superior to the global SMAP/Sentinel combined SSM in terms of both coverage percentage and estimate accuracy.

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Earth System Science Data Discussions



494

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Fig. 5 Comparison results between the currently developed 1-km SSM product and the SMAP/Sentinel combined 1-km SSM (SPL2SMAP_S_V3). (a) SPL2SMAP_S_V3 SSM images over China at about 6:00 a.m. systhesized by 6 continous dates from April 6, 2018 to April 11, 2018. (b) The SSM image at 1:30 a.m. of April 9, 2018 from the currently developed product. (c) Validation results of the SPL2SMAP_S_V3 product against in-situ soil moisture measurements over China for years of 2017, 2018, and 2019. The black solid line is the 1:1 line. (d) Same to (c) but for validaiton of the currently developed SSM product.

In Fig. 6, we display the cumulative distribution frequency of coverage percentages of the downscaled SSM product and of the original PM NN-SM product for each season. We should be noted that in this statistical scheme, pixels identified as static water body by the MODIS MCD12Q1 land cover type product were not considered in the denominator of the coverage percentage. Besides, the gap time between the respective on-orbit period of AMSR-E and of AMSR-2 (from October 2011 to June 2012, during which there are no effective observations from the PM NN-SM product) were also excluded. It is apparent that in Fig. 6-(b) and -(c), almost all downscaled daily SSM images over the 16-17 years have achieved a coverage percentage close to 100% (at least above 95%). In comparison, the majority of the PM NN-SM daily images have their coverage percentages below 80% over the study region, primarily due to the PM-seam gaps particularly existing in low latitudes (see Section 2.2.2). In Fig. 6-(a) and -(d), the percentages of effective pixels in both the PM and the downscaled SSM images are far lower than their counterparts in the other two subfigures. This is mainly ascribed to extreme meteorological conditions including snow, ice, and frozen soils that are typically persistent throughout most of these specified months in the northwestern regions of China. Such conditions can impede reliable estimates of SSM based on all satellite remote sensing techniques in the current time.

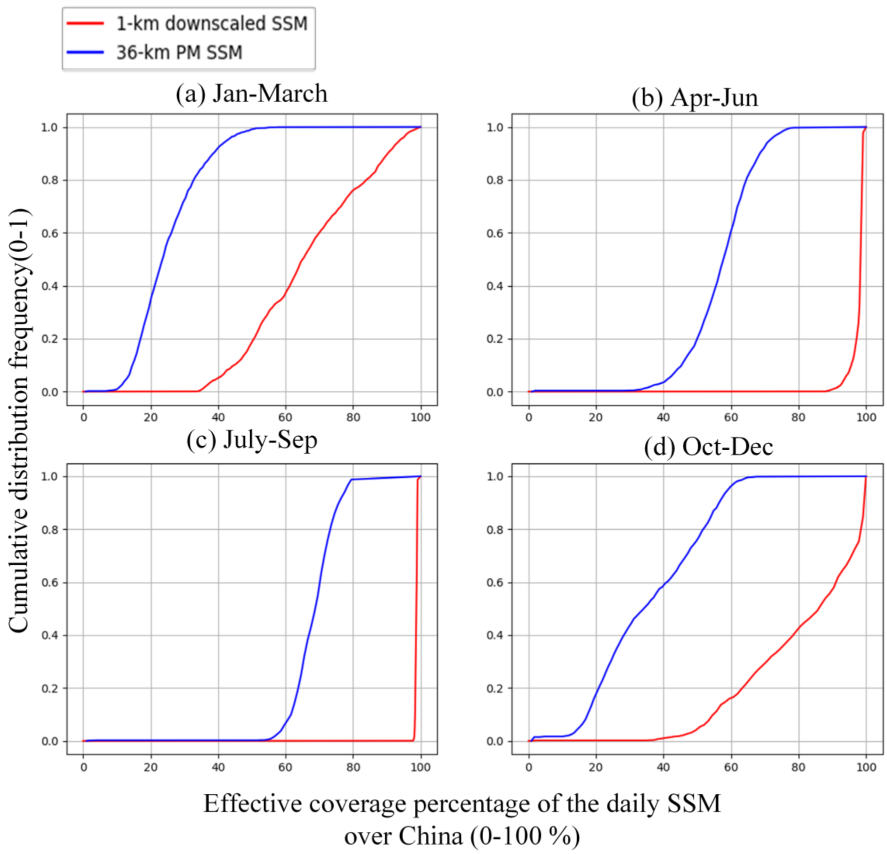
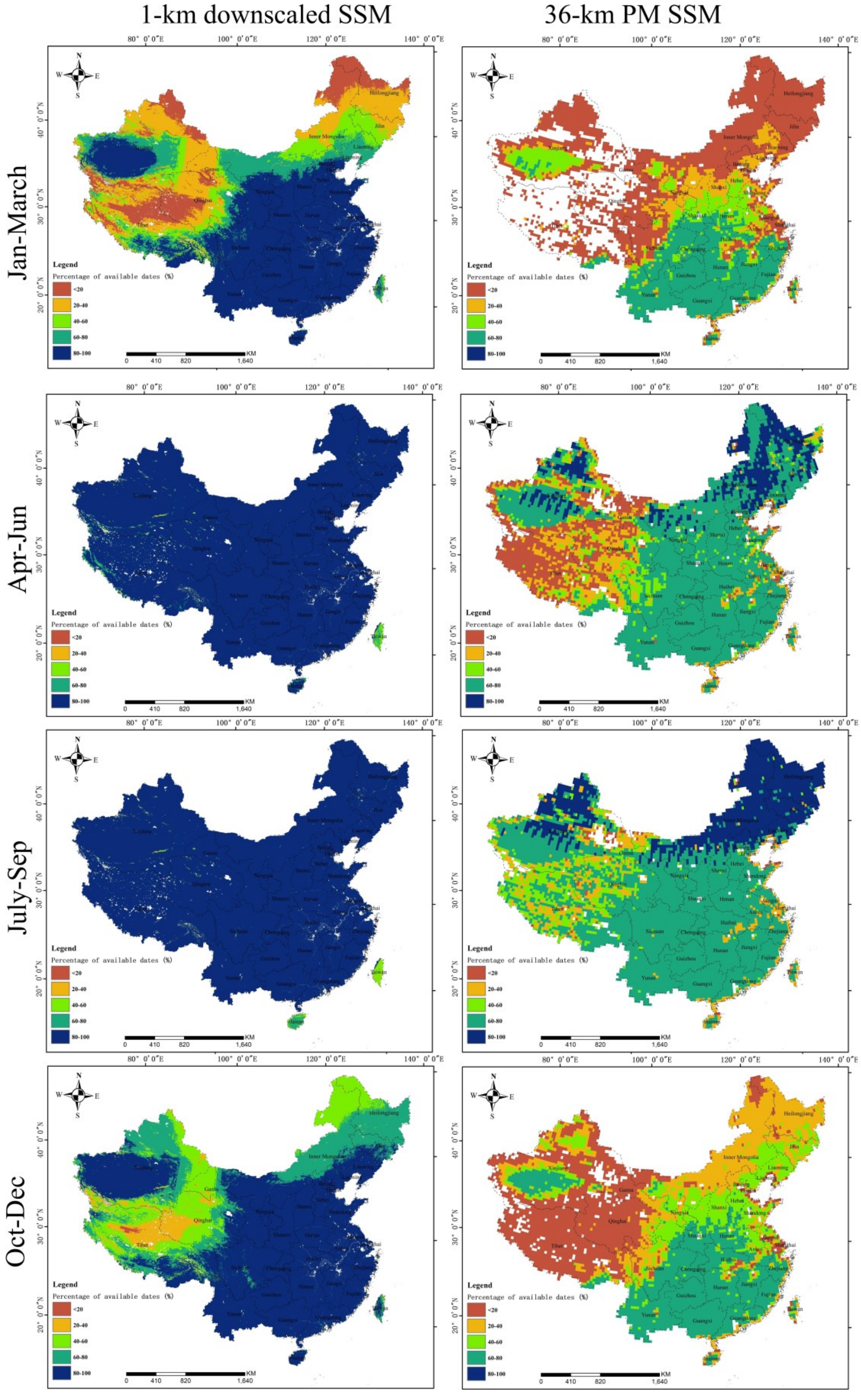


Fig. 6 Cumulative distribution frequency of our proposed SSM product against the original 36-km SSM product for different seasons. The period between October 2011 and June 2012 is excluded in the current statistics.

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Earth System Science Data
Open Access
Discussions



526

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Fig. 7 Spatial distributions on percentage of day numbers with available estimates for the currently developed 1-km SSM product and the original 36-km PM data during 2003-2019. The four different periods (i.e., January-March, April-June, July-September, October-December) of a year are treated respectively. The period between October 2011 and June 2012 is excluded.

The techniques behind coverage improvement of the downscaled SSM (against PM and optical data inputs) can be categorized into two classes, i.e. cloud gap-filling of the input optical datasets (see Section 2.2.1), as well as the filling of downscaled SSM in PM-seam gaps (see Section 2.2.2). Table 2 reports the specific validation results (using averages of all stations) of downscaled SSM in these coverage-improved conditions, relative to that generated without using any coverage improvement technique, in order to evaluate the propagated effect of such techniques on the final product. The very limited difference for ubRMSD values (0.053 vol/vol versus 0.056 vol/vol) between cloudy and clear-sky conditions suggest that the cloud gap-filling techniques are generally compatible with SSM downscaling. To a certain extent, our pre-assumption that the theoretically hypothesized ‘clear-sky’ LST reconstruction is proved suitable for quantifying soil wetness variation. The downscaled SSM estimated for regions of PM-seam gaps have a slightly worse (but still acceptable) accuracy, considering its ubRMSD of 0.059 vol/vol compared to the 0.052 vol/vol ubRMSD of the PM-observed 1-km pixels. In summary of Fig. 6 and Table 2, the currently developed product has achieved a substantially improved spatial coverage against the original remote sensing input datasets, whilst successfully preserved the SSM downscaling accuracy of the observation-covered pixels at the same time.

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Table 2 Comparisons between validation results for pixels under coverage-improved regions and for pixels under remote-sensing-observation-covered regions.

Evaluation metric*	**Comparison between cloudy and clear-sky conditions**		**Comparison between passive microwave (PM) observed regions and regions of PM-seam gaps**	
	Clear-sky condition	Cloudy condition	PM-observed regions	PM-seam gaps
ubRMSD (vol/vol)	0.053	0.056	0.052	0.059
Correlation coefficient	0.49	0.47	0.49	0.44

*All evaluation metrics in this column indicate the average of all available stations

4. Discussion

4.1 Uncertainty on SSM evaluation between satellite- and ground- scales

In this study, we made evaluations on remote sensing SSM products at different spatial resolutions, using measurements from 2000+ stations provided by the national-level soil moisture observation network of China as standard benchmark. Through the evaluations, a ubRMSD of 0.074 vol/vol is reported for the original 36-km NN-SM SSM product (Fig.A1-b). We notice that this result is considerably poorer if compared with another previous evaluation campaign targeting at the same product (Yao et al., 2021), which achieved a global RMSE (RMSD) of 0.029 vol/vol. However, this difference is not unexpected because the two campaigns were carried out in different regions of the world. Also, that particular study (Yao et al., 2021) was conducted based on completely different ground soil moisture observations provided by the International

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Soil Moisture Network (ISMN) (Dorigo et al., 2021). Compared to the observation network employed in this study, the observation sites of ISMN are more intensively distributed as an "integrated soil moisture station" so as to provide spatially average soil moisture within a grid of tens of kilometers. In this regard, we admit that the ISMN is generally more professional in evaluating satellite PM-based SSM retrievals at a coarser resolution. But on the other hand, only a few (⩽4) of such "integrated stations" have been set up sporadically within China, making the ISMN data much less representative of our study region compared with the national-level soil moisture network of China exploited by our current study.

Although the higher RMSD of the national-level soil moisture network of China may indicate larger measurement uncertainty than the ISMN, the negative influence that might be imposed on our study purpose should be inconsequential. This is because we focus more on the relative validation performance of different SSM products, rather than on the absolute value of any evaluation metric including ubRMSD and correlation coefficient calculated against ground measurements. Specifically, the 1-km downscaled SSM obtained an average ubRMSD of about 0.054 vol/vol among different stations according to Fig. 4-(b). Besides, result of the evaluation in Fig. 5-(d) based on combination of multi-station ground measurements shows a global ubRMSD of 0.078 vol/vol for this product. Overall, the above-mentioned results can be identified as at least comparable to the global (multi-station based) ubRMSD of 0.074 vol/vol of the original NN-SM data as they are evaluated against the same benchmark. Therefore, conclusion is safely drawn that the currently developed product preserves the retrieval

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

accuracy of the coarse-resolution NN-SM data, whilst improving the spatial representativeness of the latter product substantially according to the mostly positive G_{down} values in Fig. 4-(a).

Moreover, one may also argue that the *r*-value of 0.55 for the currently developed product in Fig. 5-(d) is not sufficiently high compared with several previous studies (Wei et al., 2019; Sabaghy et al., 2020) obtaining *r*-values above 0.7 for temporal analysis of satellite remote sensing soil moisture. However, we should be noticed that these previous studies have conducted analyses respectively at the temporal and the spatial dimensions. Based on their results, the spatial analysis typically derived lower *r*-values (<0.4) compared to that at the temporal dimension. This is probably because the heterogeneity degree of remote sensing pixels can vary significantly across different sites. Since the evaluation in Fig. 5-(d) was deployed at the 'spatio-temporal' integrated dimensions, such an *r*-value is expected. This is also close to the global *r*-value of 0.6 for validation of the coarse-resolution NN-SM product as reported in Yao et al. (2021).

4.2 Major novelty, unique profit, and future prospect of the developed product

Compared with the widely known active/passive microwave combined SSM product (e.g. the SPL2SMAP_S_V3) and other PM/optical-data combined counterparts which were also published recently but at the monthly scale (Meng et al., 2021), the major novelty of the currently developed product mainly lies in the fact that it has achieved progress on all of the three crucial dimensions of satellite remote sensing, including the temporal revisit cycle (daily), the spatial resolution (1-km), and the quasi-complete coverage under all-weather conditions. To our knowledge, this has rarely been achieved by previously developed satellite soil moisture product at regional scales. For realization of the above-mentioned progresses, we have fused the SSM downscaling framework with other techniques including cloud gap-filling of thermal infrared LST, MWS-based temporal filtering of vegetation indices, as well as reconstruction of seams between neighboring PM swaths in low latitudes. The final SSM estimates under cloudy conditions and intersected with the PM-seam gaps were specially validated against the rest estimates under clear sky and in the regions covered by PM observations, respectively (Table 2). The comparable performances among all treatment groups herein confirm that the accuracy of the product is stable and consistent among all weather conditions.

With improvement achieved at the three dimensions, unique profit of the currently developed product can be taken by subsequent studies and various industrial applications. For example, the capability of this product can be investigated on capturing the short-term anomaly of local hydrological signals as well as improved monitoring on drought disasters, which used to be investigated mainly at a coarser resolution by PM SSM (Scaini et al., 2015; Champagne et al., 2011; Albergel et al., 2012). For another, taking advantage of its all-weather daily time series, the product can be utilized together with precipitation data to isolate and quantify the anthropic influence on regional water resources from the natural hydrological dynamics. Examples of such anthropic signals include agricultural irrigation activities, as well as

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

finer-scale information on agricultural crops which was previously interpreted based on PM-driven techniques (Song et al., 2018). In addition, we should realize the important role of soil moisture as a constraint for accurate estimation of surface evapotranspiration and runoff (Zhang et al., 2020; Zhang et al., 2019). Therefore, the profit of this product can be further enhanced if coupled with land-atmosphere coupled models to produce new insights into water-cycle processes of earth surface at a finer spatio-temporal scale.

In the future, the methodological framework proposed in this paper is prospective to be universally applied in other regions of the world to serve for better monitoring of the global surface wetness in the following studies. If applied in continental and global scales, however, the current process for gap-filling of PM seams may require further attention and improvement. In this study, SSM in regions intersected with PM-seam gaps were estimated using TB observations from PM swaths at neighboring dates (see Equation-5). Although the errors in the PM-seam gaps over China as reported by Table 2 are only slightly larger compared to the PM-covered regions, they cannot be ignorable completely and may leave extra concern on the universality of this technique, especially in the low latitudinal tropical regions where the effect of PM-seam gap is more apparent than in our study area. Besides, another imperfection of this data product lies in the gap period between AMSR-E and AMSR-2. Considering the different systematic error patterns of various PM SSM products, we did not generate downscaled SSM based on other PM products (e.g. the SMOS SSM product) during this period but just left the period as null values. We suggest a more rigorous and universal inter-calibration framework on different PM SSM products to be developed in the future for a long-term consistent 1-km downscaled SSM dataset.

5. Conclusions

This paper describes the main technical procedures of a recently developed remote sensing surface soil moisture (SSM) product over China covering the recent ten years and more. Based on combination of passive microwave SSM downscaling theory and other related remote sensing techniques, the product achieves multi-dimensional distinctive features including 1-km resolution, daily revisit cycle, and quasi-complete all-weather coverage. These were rarely satisfied completely by other existing remote sensing SSM product at regional scales. Validations were conducted against measurements from 2000+ automatic soil moisture observation stations over China. Overall, an average ubRMSE of 0.054 vol/vol across different stations is reported for the currently developed product. The mostly positive G_{down} values show this product has significantly improved spatial representativeness against the 36-km PM SSM data (a major source for downscaling). Meanwhile, it generally preserves the retrieval accuracy of the 36-km data product. Moreover, additional validation results show that the currently developed product surpasses the widely used SMAP-sentinel combined global 1-km SSM product, with a correlation coefficient of 0.55 achieved against that of 0.40 for the latter product. The methodological framework for product generation is promising to be applied at the continental and global scales in the future, and the product

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

is potential to benefit various research/industrial fields related to hydrological processes and water resource management.

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Appendix

A. Evaluation on different PM SSM products

We have made evaluations on the various AMSR-based SSM products (as shown in Table 1) covering the recent 10 years or longer, based on our soil moisture observation network all over China. The L-band based SMAP SSM dataset was also evaluated as a reference. The evaluation period covers the three years of 2017, 2018, and 2019. All AMSR-based 25-km grids were re-set to the SMAP 36-km grid system using the nearest resampling method. Only grids that contain equal or more than 4 soil moisture measurement stations were employed, in which, the grid-based PM SSM estimate was compared with average of measurements from all interior stations. Finally, 53 grids were selected, as shown by the green color in Fig.A1-(g). For AMSR-based products, only the mid-night descending datasets were evaluated, whist for the SMAP product, our evaluation only focused on its descending mode in the early morning.

As manifested by Fig.A1-(a) to -(f), the selected SSM product in the current study, i.e., the NN-SM product has an unbiased RMSD of 0.074 vol/vol and a correlation coefficient of 0.49. This obviously outperforms the other three traditional AMSR-based SSM products (i.e. JAXA-AMSR, LPRM-AMSR, and UMT-AMSR products) and is only inferior to the SMAP SSM retrievals, whilst the later only covers the latest period since 2015. As far as CCI data are concerned, it has a similar performance against the selected NN-SM in general. Nevertheless, the region marked by red circle in Fig.A1-(c) indicates that CCI estimates have a considerably larger proportion of overestimated

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

anomalies. But overall, the primary reason that we have abandoned CCI but selected NN-SM is because the latter can provide a higher coverage fraction of valid pixels in our study region, as has been stated in Section 2.1.1.

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Open Access Earth System Science Data Discussions

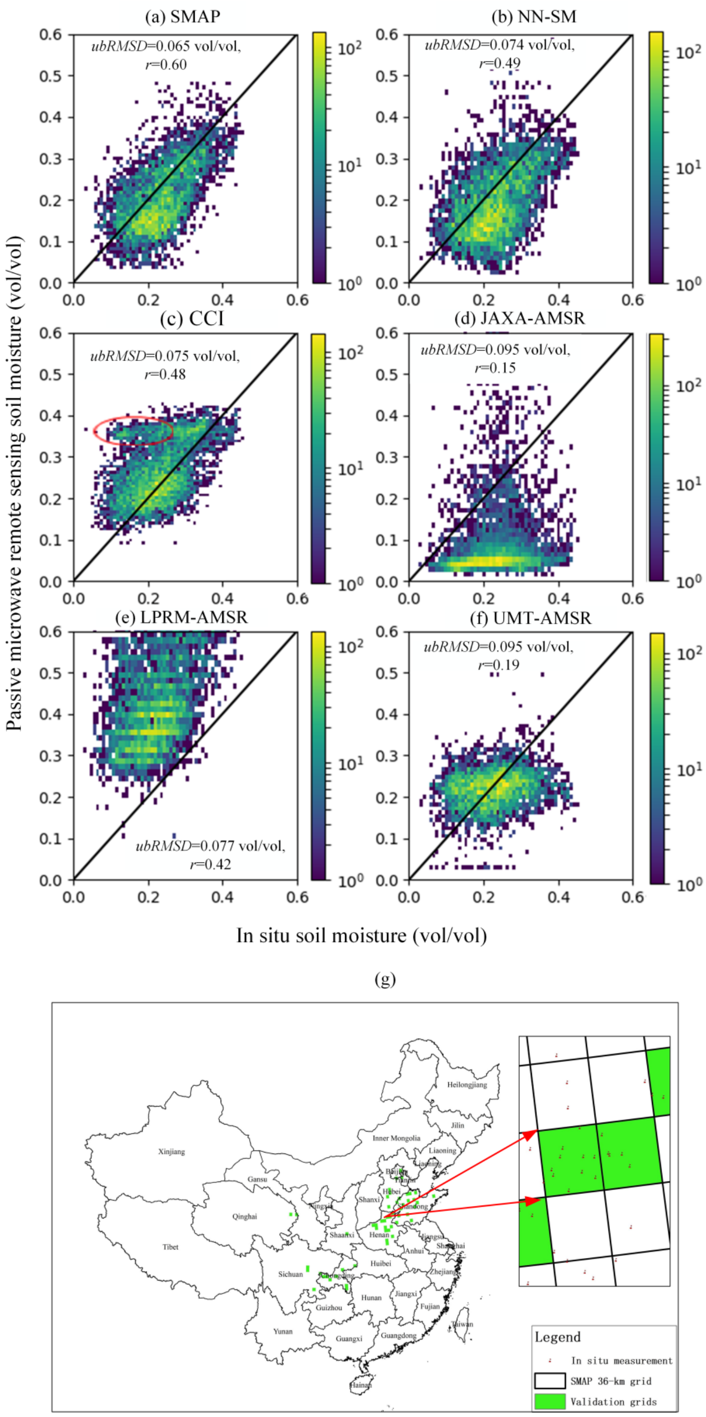


Fig. A1 (a)-(f) Comparison of different PM SSM products (as reported in Table 1) against the in situ SSM measurements in China. (g) Locations of the 36-km EASE-GRID-projection based pixels used for this comparison campaign.

B. Evaluation on the influence of bias adjustment for reconstructed 'clear-sky' LST under cloud

In Section 2.2.2, we have emphasized that the gap-filled LST for cloudy pixels reflects the theoretical surface temperature of that pixel under a hypothetical clear-sky condition. As this cloud gap-filled LST would suffer from a possible bias against the real surface temperature under cloud (Dowling et al., 2021), we made an additional experiment regarding to further improvement of this cloud gap-filled LST. The follow-up step for bias adjustment of this hypothetical clear-sky LST (but actually under cloudy conditions), as expounded in Section 4.2 of Dowling et al. (2021), was conducted herein using remote sensing and in situ LST data over China but only in 2018. We illustrate the validation results for bias adjusted and non-bias adjusted LST under cloudy conditions in Fig. A2-(b) and -(c), respectively. Similar to Fig. 3, validation results for clear-sky LST of that year are also displayed (Fig. A2-(a)) for comparison. The results generally show that the follow-up step is effective in reducing the bias of the originally gap-filled 'clear-sky LST' under cloudy conditions (from -1.7 K to 0.4 K).

In the subsequent step, we substituted the original non-bias adjusted LST under cloudy conditions with its bias adjusted counterpart, and used the latter as the input for SSM downscaling. The general validation results of the downscaled SSM are illustrated

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

in Fig. A3 (similar to that presented in Fig. 4-a and -b). Contrary to the above-analyzed Fig. A2, the bias adjusted cloudy LST with better gap-filling accuracies, however, obtained inferior performance in SSM downscaling. This final validation result, to some degree, confirms our assumption in Section 2.2.2 that the reconstructed cloudy LST but for the hypothesized clear-sky condition is the better proxy of surface moisture dynamics. But overall, as all LST estimates discussed herein are for the midnight scenario (when the energy interaction between atmosphere and land surface is relatively weak), the RMSD difference for different weather conditions in Fig.A2 is expectedly marginal. As a consequence, the difference in ubRMSD of SSM in Fig.A3 can hardly be identified as 'very significant'. Therefore, we encourage further tests on this conclusion in specific future studies to confirm its universality, especially for situation of the 'morning to noon' time window.

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

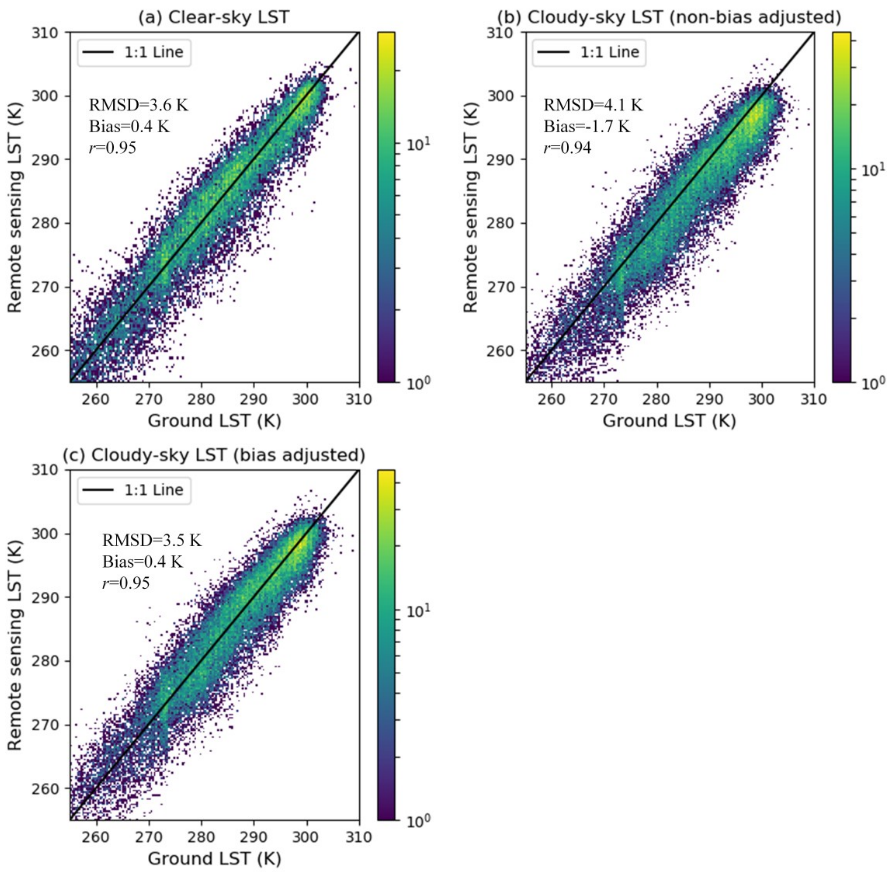


Fig. A2 Validation of the clear sky LST (a), reconstructed LST under cloud but with no passive-microwave based bias adjustment (b), as well as the reconstructed LST under cloud with passive-microwave based bias adjustment (c) respectively, based on the 0-cm ground temperature measurements at meteorological stations.

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Open Access Earth System Science Data Discussions

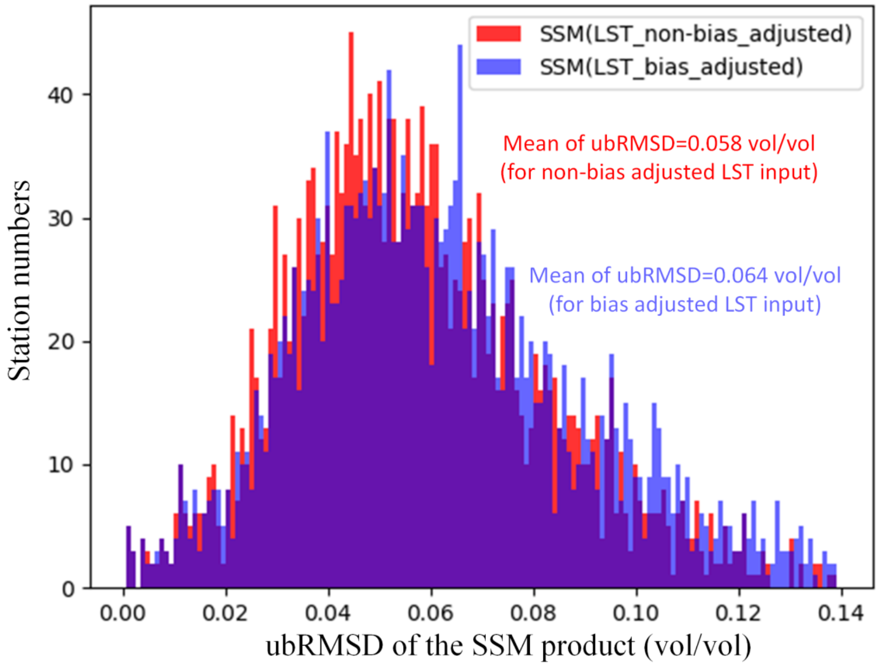


Fig. A3 The statistical distribution of ubRMSD at different stations for SSM estimates driven by two respective kinds of cloudy LST inputs.

Author contributions

Peilin Song and Yongqiang Zhang designed the research and developed the whole methodological framework; Peilin Song and Yongqiang Zhang supervised the processing line of the 1-km SSM product; Jianping Guo and Bingtong provide in situ soil moisture data for validation; Peilin Song wrote the original draft of the manuscript; Yongqiang Zhang, Jiancheng Shi, and Tianjie Zhao revised the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Data availability

The published SSM dataset is available under the Creative Commons Attribution 4.0 International License at the following link: http://dx.doi.org/10.11888/Hydro.tpdc.271762 (Song and Zhang, 2021b). This dataset covers all of China's terrestrial area at a daily revisit frequency (about 1:30 A.M. at local time) and a 1km spatial resolution from January 2003 to October 2011 and from July 2012 to December 2019.

Acknowledgement

The authors would like to thank the National Aeronautics and Space Administration (NASA) for providing all MODIS and DEM data sets free of charge.

Financial support

This study was jointly supported by the National Natural Science Foundation of China (Grant No. 42001304), the Open Fund of State Key Laboratory of Remote Sensing Science (Grant No. OFSLRSS202117), CAS Pioneer Talents Program, CAS-CSIRO International Cooperation Program, and the International Partnership Program of Chinese Academy of Sciences (Grant No. 183311KYSB20200015).

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Reference

Albergel, C., de Rosnay, P., Gruhier, C., Munoz-Sabater, J., Hasenauer, S., Isaksen, L., . . . Wagner, W.: Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations, Remote Sens. Environ., 118, 215-226, 10.1016/j.rse.2011.11.017, 2012.

Brunsdon, C., Fotheringham, A. S., and Charlton, M. E.: Geographically weighted regression: A method for exploring spatial nonstationarity, Geogr. Anal., 28, 281-298, 1996.

Busch, F. A., Niemann, J. D., and Coleman, M.: Evaluation of an empirical orthogonal function-based method to downscale soil moisture patterns based on topographical attributes, Hydrological Processes, 26, 2696-2709, 2012.

Carlson, T. N., Gillies, R. R., and Perry, E. M.: A method to make use of thermal infrared temperature and NDVI measurements to infer surface soil water content and fractional vegetation cover, Remote sensing reviews, 9, 161-173, 1994.

Champagne, C., McNairn, H., and Berg, A. A.: Monitoring agricultural soil moisture extremes in Canada using passive microwave remote sensing, Remote Sens. Environ., 115, 2434-2444, 2011.

Chauhan, N. S., Miller, S., and Ardanuy, P.: Spaceborne soil moisture estimation at high resolution: a microwave-optical/IR synergistic approach, Int. J. Remote Sens., 24, 4599-4622, http://doi.org/10.1080/0143116031000156837, 2003.

Choi, M. and Hur, Y.: A microwave-optical/infrared disaggregation for improving spatial representation of soil moisture using AMSR-E and MODIS products, Remote Sens. Environ., 124, 259-269, http://doi.org/10.1016/j.rse.2012.05.009, 2012.

Das, N., Entekhabi, D., Dunbar, R. S., Kim, S., Yueh, S., Colliander, A., . . . Cosh, M.: SMAP/Sentinel-1 L2 Radiometer/Radar 30-Second Scene 3 km EASE-Grid Soil Moisture, Version 3 [dataset], https://doi.org/10.5067/ASB0EQO2LYJV, 2020.

Das, N. N., Entekhabi, D., Dunbar, R. S., Chaubell, M. J., Colliander, A., Yueh, S., . . . Thibeault, M.: The SMAP and Copernicus Sentinel 1A/B microwave active-passive high resolution surface soil moisture product, Remote Sens. Environ., 233, 111380, https://doi.org/10.1016/j.rse.2019.111380, 2019.

Dorigo, W., Himmelbauer, I., Aberer, D., Schremmer, L., Petrakovic, I., Zappa, L., . . . Sabia, R.: The International Soil Moisture Network: serving Earth system science for over a decade, Hydrol. Earth Syst. Sci., 25, 5749-5804, 10.5194/hess-25-5749-2021, 2021.

Dowling, T. P. F., Song, P., Jong, M. C. D., Merbold, L., Wooster, M. J., Huang, J., and Zhang, Y.: An Improved Cloud Gap-Filling Method for Longwave Infrared Land Surface Temperatures through Introducing Passive Microwave Techniques, Remote Sens., 13, 3522, 2021.

Du, J. Y., Kimball, J. S., and Jones, L. A.: Passive microwave remote sensing of soil moisture based on dynamic vegetation scattering properties for AMSR-E, IEEE Trans. Geosci. Remote Sens, 54, 597-608, 2016.

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Duan, S. and Li, Z.: Spatial Downscaling of MODIS Land Surface Temperatures Using Geographically Weighted Regression: Case Study in Northern China, IEEE Trans. Geosci. Remote Sens, 54, 6458-6469, http://doi.org/10.1109/TGRS.2016.2585198, 2016.

Entekhabi, D., Reichle, R. H., Koster, R. D., and Crow, W. T.: Performance Metrics for Soil Moisture Retrievals and Application Requirements, J. Hydrometeorol., 11, 832-840, 10.1175/2010jhm1223.1, 2010.

Entekhabi, D., Das, N., Kim, S., Jagdhuber, T., Piles, M., Yueh, S., . . . Martínez-Fernández, J.: High-Resolution Enhanced Product based on SMAP Active-Passive Approach and Sentinel 1A Radar Data, AGU Fall Meeting Abstracts, H24C-08,

Fang, B. and Lakshmi, V.: Passive Microwave Soil Moisture Downscaling Using Vegetation and Surface Temperatures, Vadose Zone J, 12, 1712-1717, 2013.

Fang, B., Lakshmi, V., Bindlish, R., and Jackson, T.: AMSR2 Soil Moisture Downscaling Using Temperature and Vegetation Data, Remote Sens., 10, 2018.

Fang, B., Lakshmi, V., Bindlish, R., Jackson, T. J., Cosh, M., and Basara, J.: Passive Microwave Soil Moisture Downscaling Using Vegetation Index and Skin Surface Temperature, 2013.

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

Im, J., Park, S., Rhee, J., Baik, J., and Choi, M.: Downscaling of AMSR-E soil moisture with MODIS products using machine learning approaches, Environ Earth Sci, 75, 1-19, http://doi.org/10.1007/s12665-016-5917-6, 2016.

Jones, L. A., Kimball, J. S., Podest, E., McDonald, K. C., Chan, S. K., and Njoku, E. G.: A method for deriving land surface moisture, vegetation optical depth, and open water fraction from AMSR-E, IEEE IGARSS 2009., Cape Town, South Africa, 2009, III-916-III-919, http://doi.org/10.1109/IGARSS.2009.5417921,

Kim, J. and Hogue, T. S.: Improving spatial soil moisture representation through integration of AMSR-E and MODIS products, IEEE Trans. Geosci. Remote Sens, 50, 446-460, http://doi.org/10.1109/TGRS.2011.2161318, 2012.

Koike, T., Nakamura, Y., Kaihotsu, I., Davva, G., Matsuura, N., Tamagawa, K., and Fujii, H.: Development of an Advanced Microwave Scanning Radiometer (AMSR-E) algorithm of soil moisture and vegetation water content (written in Japanese), Annual Journal of Hydraulic Engineering, 48, 217-222 2004.

Kong, D., Zhang, Y., Gu, X., and Wang, D.: A robust method for reconstructing global MODIS EVI time series on the Google Earth Engine, Isprs J Photogramm, 155, 13-24, 2019.

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Malbéteau, Y., Merlin, O., Molero, B., Rüdiger, C., and Bacon, S.: DisPATCh as a tool to evaluate coarse-scale remotely sensed soil moisture using localized in situ measurements: Application to SMOS and AMSR-E data in Southeastern Australia, Int J Appl Earth Obs, 45, 221-234, https://doi.org/10.1016/j.jag.2015.10.002, 2016.

Meesters, A. G. C. A., De Jeu, R. A. M., and Owe, M.: Analytical derivation of the vegetation optical depth from the microwave polarization difference index, IEEE Geosci. Remote Sens. Lett., 2, 121-123, 2005.

Mendoza, P. A., Mizukami, N., Ikeda, K., Clark, M. P., Gutmann, E. D., Arnold, J. R., . . . Rajagopalan, B.: Effects of different regional climate model resolution and forcing scales on projected hydrologic changes, J. Hydrol., 541, 1003-1019, https://doi.org/10.1016/j.jhydrol.2016.08.010, 2016.

Meng, X. J., Mao, K. B. A., Meng, F., Shi, J. C., Zeng, J. Y., Shen, X. Y., . . . Guo, Z. H.: A fine-resolution soil moisture dataset for China in 2002-2018, Earth Syst. Sci. Data, 13, 3239-3261, 10.5194/essd-13-3239-2021, 2021.

Merlin, O., Al Bitar, A., Walker, J. P., and Kerr, Y.: An improved algorithm for disaggregating microwave-derived soil moisture based on red, near-infrared and thermal-infrared data, Remote Sens. Environ., 114, 2305-2316, http://doi.org/10.1016/j.rse.2010.05.007, 2010.

Merlin, O., Walker, J. P., Chehbouni, A., and Kerr, Y.: Towards deterministic downscaling of SMOS soil moisture using MODIS derived soil evaporative efficiency, Remote Sens. Environ., 112, 3935-3946, http://doi.org/10.1016/j.se.2008.06.012, 2008.

Merlin, O., Chehbouni, A. G., Kerr, Y. H., Njoku, E. G., and 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, http://doi.org/10.1109/TGRS.2005.853192, 2005.

Merlin, O., Escorihuela, M. J., Mayoral, M. A., Hagolle, O., Al Bitar, A., and Kerr, Y.: Self-calibrated evaporation-based disaggregation of SMOS soil moisture: An evaluation study at 3 km and 100 m resolution in Catalunya, Spain, Remote Sens. Environ., 130, 25-38, 10.1016/j.rse.2012.11.008, 2013.

Merlin, O., Malbeteau, Y., Notfi, Y., Bacon, S., Er-Raki, S., Khabba, S., and Jarlan, L.: Performance Metrics for Soil Moisture Downscaling Methods: Application to DISPATCH Data in Central Morocco, Remote Sens., 7, 3783-3807, http://doi.org/10.3390/rs70403783, 2015.

Molero, B., Merlin, O., Malbéteau, Y., Al Bitar, A., Cabot, F., Stefan, V., . . . Jackson, T. J.: SMOS disaggregated soil moisture product at 1km resolution: Processor overview and first validation results, Remote Sens. Environ., 180, 361-376, http://doi.org/10.1016/j.rse.2016.02.045, 2016.

Montaldo, N., Albertson, J. D., Mancini, M., and Kiely, G.: Robust simulation of root zone soil moisture with assimilation of surface soil moisture data, Water Resour Res, 37, 2889-2900, 10.1029/2000WR000209, 2001.

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Owe, M., de Jeu, R., and Walker, J.: A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index, IEEE Trans. Geosci. Remote Sens, 39, 1643-1654, 2001.

Peng, J., Loew, A., Zhang, S. Q., Wang, J., and Niesel, J.: Spatial downscaling of satellite soil moisture data using a vegetation temperature condition index, IEEE Trans. Geosci. Remote Sens, 54, 558-566, http://doi.org/10.1109/TGRS.2015.2462074, 2016.

Piles, M., Entekhabi, D., and Camps, A.: A Change Detection Algorithm for Retrieving High-Resolution Soil Moisture From SMAP Radar and Radiometer Observations, IEEE Trans. Geosci. Remote Sens, 47, 4125-4131, 10.1109/TGRS.2009.2022088, 2009.

Sabaghy, S., Walker, J. P., Renzullo, L. J., Akbar, R., Chan, S., Chaubell, J., . . . Yueh, S.: Comprehensive analysis of alternative downscaled soil moisture products, Remote Sens. Environ., 239, 111586, https://doi.org/10.1016/j.rse.2019.111586, 2020.

Sanchez-Ruiz, S., Piles, M., Sanchez, N., Martinez-Fernandez, J., Vall-llossera, M., and Camps, A.: Combining SMOS with visible and near/shortwave/thermal infrared satellite data for high resolution soil moisture estimates, J. Hydrol., 516, 273-283, 10.1016/j.jhydrol.2013.12.047, 2014.

Scaini, A., Sanchez, N., Vicente-Serrano, S. M., and Martinez-Fernandez, J.: SMOS-derived soil moisture anomalies and drought indices: a comparative analysis using in situ measurements, Hydrological Processes, 29, 373-383, 10.1002/hyp.10150, 2015.

Song, P. and Zhang, Y.: An improved non-linear inter-calibration method on different radiometers for enhancing coverage of daily LST estimates in low latitudes, Remote Sens. Environ., 264, 112626, https://doi.org/10.1016/j.rse.2021.112626, 2021a.

Song, P. and Zhang, Y.: Daily all weather surface soil moisture data set with 1 km resolution in China (2003-2019), National Tibetan Plateau Data Center [dataset], 10.11888/Hydro.tpdc.271762, 2021b.

Song, P., Huang, J., and Mansaray, L. R.: An improved surface soil moisture downscaling approach over cloudy areas based on geographically weighted regression, Agr Forest Meteorol, 275, 146-158, 10.1016/j.agrformet.2019.05.022, 2019a.

Song, P., Zhang, Y., and Tian, J.: Improving Surface Soil Moisture Estimates in Humid Regions by an Enhanced Remote Sensing Technique, Geophys Res Lett, 48, e2020GL091459, https://doi.org/10.1029/2020GL091459, 2021.

Song, P., Mansaray, L. R., Huang, J., and Huang, W.: Mapping paddy rice agriculture over China using AMSR-E time series data, Isprs J Photogramm, 144, 469-482, 10.1016/j.isprsjprs.2018.08.015, 2018.

Song, P., Huang, J., Mansaray, L. R., Wen, H., Wu, H., Liu, Z., and Wang, X.: An Improved Soil Moisture Retrieval Algorithm Based on the Land Parameter Retrieval Model for Water-Land Mixed Pixels Using AMSR-E Data, IEEE Trans. Geosci. Remote Sens, 1-15, 10.1109/TGRS.2019.2915346, 2019b.

https://doi.org/10.5194/essd-2021-428
Preprint. Discussion started: 10 January 2022

Sui, D. Z.: Tobler's First Law of Geography: A Big Idea for a Small World?, Annals of the Association of American Geographers, 94, 269-277, https://doi.org/10.1111/j.1467-8306.2004.09402003.x, 2004.

Wang, L. and Qu, J. J.: NMDI: A normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing, Geophys Res Lett, 34, L20405, 10.1029/2007GL031021, 2007.

Wei, Z., Meng, Y., Zhang, W., Peng, J., and Meng, L.: Downscaling SMAP soil moisture estimation with gradient boosting decision tree regression over the Tibetan Plateau, Remote Sens. Environ., 225, 30-44, 2019.

Wu, D., Liang, H., Cao, T., Yang, D., Zhou, W., and Wu, X.: Construction of operation monitoring system of automatic soil moisture observation network in China, Meteorological Science and Technology, 42, 278-282, 2014

Yao, P., Lu, H., Shi, J., Zhao, T., Yang, K., Cosh, M. H., . . . Entekhabi, D.: A long term global daily soil moisture dataset derived from AMSR-E and AMSR2 (2002–2019), Scientific Data, 8, 143, 10.1038/s41597-021-00925-8, 2021.

Zhang, Y., Kong, D., Gan, R., Chiew, F. H. S., Mcvicar, T. R., Zhang, Q., and Yang, Y.: Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002-2017, Remote Sens. Environ., 222, 165-182, 2019.

Zhang, Y. Q., Chiew, F. H. S., Liu, C. M., Tang, Q. H., Xia, J., Tian, J., . . . Li, C. C.: Can Remotely Sensed Actual Evapotranspiration Facilitate Hydrological Prediction in Ungauged Regions Without Runoff Calibration?, Water Resour Res, 56, 2020.

Zheng, J. Y., Lu, H. S., Crow, W. T., Zhao, T. J., Merlin, O., Rodriguez-Fernandez, N., . . . Gou, Q. Q.: Soil moisture downscaling using multiple modes of the DISPATCH algorithm in a semi-humid/humid region, Int J Appl Earth Obs, 104, 10.1016/j.jag.2021.102530, 2021.