A 3-hour, 1-km surface soil moisture dataset for the contiguous United States from 2015 to 2023

Haoxuan Yang [a*], Jia Yang [a], Tyson E. Ochsner [b], Erik S. Krueger [b], Mengyuan Xu [c],

5 Chris B. Zou [a]

Affiliation

[a] Department of Natural Resource Ecology and Management, Oklahoma State University, Stillwater, OK 74078, USA.

10 [b] Department of Plant and Soil Sciences, Oklahoma State University, Stillwater, OK 74078, USA.

[c] Institute of Agricultural Science and Technology Information, Shanghai Academy of Agricultural Sciences, Shanghai 201403, China.

15

Correspondence to: Haoxuan Yang (yhx965665334@163.com)

20

Abstract.

Surface soil moisture (SSM) is a critical variable for understanding the terrestrial hydrologic cycle, and it influences ecosystem dynamics, agriculture productivity, and water resource management. Although SSM information is widely estimated through satellite-derived and model-assimilated methods, datasets with fine spatio-temporal resolutions remain unavailable at the continental scale, yet are essential for improving weather forecasting, optimizing precision irrigation, and enhancing fire risk assessment. In this study, we developed a new 3-hour, 1-km spatially seamless SSM dataset spanning 2015 to 2023, covering the entire contiguous United States (CONUS), using a spatio-temporal fusion model. This approach effectively combines the distinct advantages of two long-term SSM datasets, namely, the Soil Moisture Active Passive (SMAP) L4 SSM product and the Crop Condition and Soil Moisture Analytics (Crop-CASMA) dataset. The SMAP product provides spatially seamless SSM observations with a 3-hour temporal resolution but at a 9-km spatial resolution, while the Crop-CASMA SSM dataset offers a finer spatial resolution of 1 km but has a daily temporal resolution and contains spatial gaps. To overcome the spatio-temporal mismatch between the two products, we developed a time-series data mining approach known as the highly comparative time-series analysis (HCTSA) method to extract multiple spatially seamless characteristics (e.g., maximum and mean) from the two inter-annual SSM datasets (i.e., SMAP and Crop-CASMA). Then the fusion model was constructed using the extracted 9-km and 1-km characteristics and each scene of the SMAP, in turn. Finally, the 3-hour, 1-km SSM data (named as STF_SSM) were predicted from 2015 to 2023. The comparison with *in-situ* data from multiple SSM observation networks showed that the performance of our STF_SSM dataset is better than the Crop-CASMA and is close to the SMAP L4 product, with mean correlation coefficients (CC) of 0.716 at the daily scale and 0.689 at the 3-hour scale. The STF_SSM dataset in this study is the first long time-series, spatially seamless SSM dataset to realize continuous intra-day 1-km SSM observations every 3 hours across the CONUS, which provides a new insight into the fast changes in soil moisture along with drought and wet spell occurrences, and ecosystem responses. Additionally, this dataset provides a valuable data source for the calibration and validation of land surface models. The STF_SSM dataset is available at https://doi.org/10.6084/m9.figshare.28188011 (Yang et al., 2025).

1. Introduction

Surface soil moisture (SSM) is an important component of global hydrological cycling and serves as a key indicator of drought occurrences **(Souza et al., 2021; Krueger et al., 2024)**, climate change **(Guillod et al., 2015)**, and ecosystem functions **(Green et al., 2019; Liu et al., 2020a)**. To better understand the spatio-temporal changes in SSM, the Soil Moisture Active Passive (SMAP) satellite, launched in 2015, provides SSM data on a global scale at spatial resolutions of 9-km and 36-km using the onboard L-band radiometer **(Entekhabi et al., 2010; Chan et al., 2016)**. SMAP products are generated at four levels of processing. Retrieved from brightness temperature information observed by satellites, SMAP Level 2 (L2) and Level 3 (L3) SSM data are half-orbital and daily composited. However, considering the non-overlapping revisit orbit and snow coverage, spatial gaps are inevitable in SMAP L2 and L3 SSM products. To solve this problem, SMAP Level 4 (L4) SSM product assimilates SMAP L2 and L3 data into a land surface model and provides spatially seamless SSM and root zone soil moisture estimates with a temporal resolution of 3 hours and spatial resolution of 9 km.

Numerous studies have investigated the performance of satellite-derived SSM based on the triple collocation analysis **(Chen et al., 2018)**, information theory **(Kumar et al., 2018)**, and ground-based *in-situ* data **(Kim et al., 2018)**. Results showed that the SMAP data typically have better performance at a global scale compared with other satellite-derived SSM products, such as the Soil Moisture and Ocean Salinity (SMOS), the Advanced Scatterometer (ASCAT), and Advanced Microwave Scanning Radiometer 2 (AMSR2). For example, **Ma et al. (2019)** and **Min et al. (2023)** demonstrated that SMAP data outperformed the SMOS, AMSR2, and the Climate Change Initiative of the European Space Agency (CCI) SSM product in terms of capturing temporal dynamics. **Montzka et al. (2017)** evaluated different SSM products in six regions and found that the SMAP data had greater accuracy than SMOS, AMSR2, and ASCAT. Additionally, SMAP data outperformed other satellite-derived SSM products in the Gilgel Abay watershed of Ethiopia **(Alaminie et al., 2024)**, the Genhe area in China **(Cui et al., 2017)**, and the Huai River basin of China **(Wang et al., 2021)**.

Although the SMAP data are stable and reliable, potential applications are constrained by their coarse spatial resolution. To address this issue, several downscaling strategies have been

applied by integrating optical/thermal-infrared data **(Peng et al., 2017; Sabaghy et al., 2020; Abbaszadeh et al., 2021; Meng et al., 2024)**. In general, the universal triangle feature **(Carlson et al., 1990, 1995)** and trapezoidal feature spaces **(Moran et al., 1994; Merlin et al., 2012)** provide the theoretical basis for most downscaling studies. The universal triangle feature primarily leverages land surface temperature (LST) and normalized difference vegetation index (NDVI) from optical/thermal-infrared data, e.g., the Moderate Resolution Imaging Spectroradiometer (MODIS), to capture spatio-temporal variations in SSM, highlighting that SSM is closely related with LST and NDVI. **(Carlson, 2007)**. Compared to the universal triangle feature, the trapezoidal feature considers the influence of the fraction of water-stressed vegetation **(Djamai et al., 2016)**. By incorporating fine-resolution LST and NDVI at various fractional vegetation cover conditions, the effects of evaporation (e.g., soil evaporative efficiency) can be quantified, enabling the development of SSM products at high spatial resolution **(Merlin et al., 2012; Kim and Hogue, 2012; Molero et al., 2016)**. Therefore, fine-resolution LST and NDVI are often employed as auxiliary data for SSM downscaling.

Commonly, downscaling approaches are based on geostatistical models, which consider the spatial variations of LST and NDVI within and outside of the SSM pixel **(Song et al., 2019)**. For example, a regression kriging-based model and its modified version were used to disaggregate the coarse pixels in SSM data using LST and NDVI data **(Jin et al., 2018; Wen et al., 2020; Jin et al., 2021; Yang et al., 2024)**. Robust ensemble learning approaches have also been used to downscale SSM **(Zhao et al., 2018; Wei et al., 2019; Karthikeyan and Mishra, 2021)**. For example, by integrating multiple decision tree models, a random forest model was employed to downscale SMAP SSM data from 36 km to 1 km **(Hu et al., 2020)**. This approach can include fine-resolution auxiliary information, such as topography, location, and soil texture **(Abbaszadeh et al., 2019; Liu et al., 2020b; Guevara et al., 2021; Wang et al., 2022)**. Deep learning is another popular downscaling approach, as the strong fitting capability effectively characterizes the SSM using LST, NDVI, and other auxiliary information **(Xu et al., 2022; Zhao et al., 2022; Xu et al., 2024)**.

Using the aforementioned methods, many high-resolution SSM datasets have been developed **(Vergopolan et al., 2021; Han et al., 2023; Brocca et al., 2024)**. However, the optical/thermal-infrared auxiliary data are usually disturbed by atmospheric conditions, such

as clouds and haze **(Ma et al., 2022a, b)**, resulting in difficult disaggregation of coarse SSM pixels under clouds or haze. To mitigate these issues, multiple-day composited optical/thermal-infrared data are often used as the auxiliary variables for producing the SSM dataset **(Li et al., 2022; Zheng et al., 2023)**. Meanwhile, reconstruction of the missing optical/thermal-infrared data is a reliable choice, which is then used for SSM generation **(Long et al., 2019; Abowarda et al., 2021; Song et al., 2022)**. For example, **Zhao et al. (2021)** reconstructed the seamless LST data and used them to generate the SSM dataset at a 1-km spatial resolution. In addition, integrating other high-resolution SSM products is also an appropriate method, which can avoid the influence of the optical/thermal-infrared auxiliary data **(Jiang et al., 2019; Yang et al., 2022; Jiang et al., 2024)**. In addition, synthetic aperture radar (SAR) data are also beneficial for generating high-resolution SSM datasets. Since microwave signals can penetrate the cover of clouds or haze, the SSM estimation can avoid the influence of weather factors. However, producing a large-scale, fine temporal resolution SSM product is limited by the coarse revisit period and narrow swath width of SAR data **(Wang et al., 2023; Zhu et al., 2023; Fan et al., 2025)**.

High-resolution SSM datasets have been developed based on the original SMAP SSM product. For instance, the National Aeronautics and Space Administration (NASA) combined data from the Sentinel-1 satellites' synthetic aperture radar with SMAP's passive radiometer to produce the SSM product (SPL2SMAP_S), which offers a spatial resolution of 3 km **(Jagdhuber et al., 2019)**. However, differences in the revisit orbits of SMAP and Sentinel-1, coupled with the narrower swath width of Sentinel-1 compared to SMAP, restrict the spatial coverage of the SPL2SMAP_S product **(Das et al., 2019; Kim et al., 2021)**. In addition, the United States Department of Agriculture's National Agricultural Statistics Service (USDA-NASS) has developed a daily, 1-km resolution SSM product within the Crop Condition and Soil Moisture Analytics (Crop-CASMA) system **(Colliander et al., 2019; Zhang et al., 2022)**. This product disaggregates the 9-km satellite-derived SMAP SSM data by incorporating auxiliary 1-km data from MODIS **(Liu et al., 2021, 2022)**. Although the Crop-CASMA SSM product provides sufficient spatial details, it retains spatial gaps inherent to the daily SMAP SSM data, limiting its overall spatial continuity. Similarly, the lack of spatial information also restricts the application of other high-resolution SSM datasets **(Fang et al., 2022; Lakshmi**

and Fang, 2023; Yang et al., 2024).

Finer spatial and temporal resolution have become increasingly important for SSM datasets to facilitate more accurate monitoring of dynamic soil moisture variations. For long time-series and large-scale SSM datasets, a 1-km spatial resolution is commonly adopted, as daily auxiliary data at 1-km resolution can be extracted from MODIS. However, the commonly used 1-km SSM datasets at a large scale have a daily or even coarser temporal resolution, limiting their capacity to depict the intra-day SSM variations. This highlights the challenge of achieving both high temporal and spatial resolution in SSM datasets simultaneously.

In this work, we generated a 3-hour, 1-km SSM dataset (denoted as STF_SSM) for the contiguous United States (CONUS) from 2015 to 2023. Based on an advanced and efficient spatio-temporal fusion model, the advantages of high observation frequency (3-hour) in the SMAP L4 SSM product and satisfactory spatial details (1-km) in the Crop-CASMA SSM dataset were integrated into a new SSM dataset. A time-series data mining approach was employed to extract multiple spatially seamless characteristics from the SMAP L4 and Crop-CASMA SSM datasets, effectively addressing the spatio-temporal mismatches between the two input datasets within the fusion model. To evaluate the performance of the STF_SSM dataset, ground-based *in-situ* measurements were used for validation at both 3-hour and daily scales. The generated STF_SSM dataset facilitates intra-day SSM observations, providing a valuable resource for the related studies.

2. Data and methods

2.1 Data

2.1.1 SMAP L4 SSM product

170 SMAP L4 SSM data were downloaded from https://nsidc.org. The temporal and spatial resolutions of the SMAP L4 SSM product are 3-hour and 9-km, respectively. The SMAP L4 SSM product has a spatially complete coverage at a global scale. Validation studies showed that the SMAP L4 product provides more accurate and stable performance than SMAP L3 across all seasons **(Tavakol et al., 2019)**. In this work, we used the latest version 7 SMAP L4
175 geophysical dataset.

2.1.2 Crop-CASMA SSM data

The Crop-CASMA system integrates crucial vegetation and soil moisture data for the CONUS (such as SSM, root-zone soil moisture, and NDVI). These data are continuously updated and can be freely accessed from the USDA-NASS website at
180 https://nassgeo.csiss.gmu.edu/CropCASMA/. The system supports direct download, analysis, and visualization. In this study, the Crop-CASMA SSM data are derived from the SMAP Thermal Hydraulic disaggregation of Soil Moisture (SMAP THySM) dataset, which can provide 1-km daily SSM data and have two days of latency **(Liu et al., 2021, 2022; Zhang et al., 2022)**.

185 #### 2.1.3 *In-situ* data

In-situ data are measured and recorded by ground-based sensors at different depths, which has often been used as the reference for validation of satellite-derived SSM datasets **(Dorigo et al., 2015)**. In this study, *in-situ* data from 2015 to 2023 were obtained from the International Soil Moisture Network (https://ismn.earth/en/) and the Oklahoma Mesonet
190 (https://www.mesonet.org/), with measurements taken at a depth of 5 cm **(McPherson et al., 2007)**. Noted that the available *in-situ* data were further filtered to ensure that only one site was included per SSM pixel. Detailed information and locations of the *in-situ* networks were presented in Table 1 and Figure 1.

Table 1. Details of the selected *in-situ* data and soil moisture observation network for validation.

| Network	Site number	Sensor
ARM	14	Hydraprobe II Sdi-12 E
CW3E	16	CS616/ Stevens-Hydra-Probe
FLUXNET-AMERIFLUX	2	CS655/ ThetaProbe-ML2X
MESONET	123	Campbell Scientific 229-L
PBO_H2O	130	GPS
SCAN	183	Hydraprobe-Analog-(2.5-Volt)/ Hydraprobe-Digital-Sdi-12-(2.5-Volt)
SNOTEL	261	Hydraprobe-Analog-(2.5-Volt)/ Hydraprobe-Analog-(5.0-Volt)
TxSON	15	CS655
USCRN	91	Stevens-Hydraprobe-II-Sdi-12
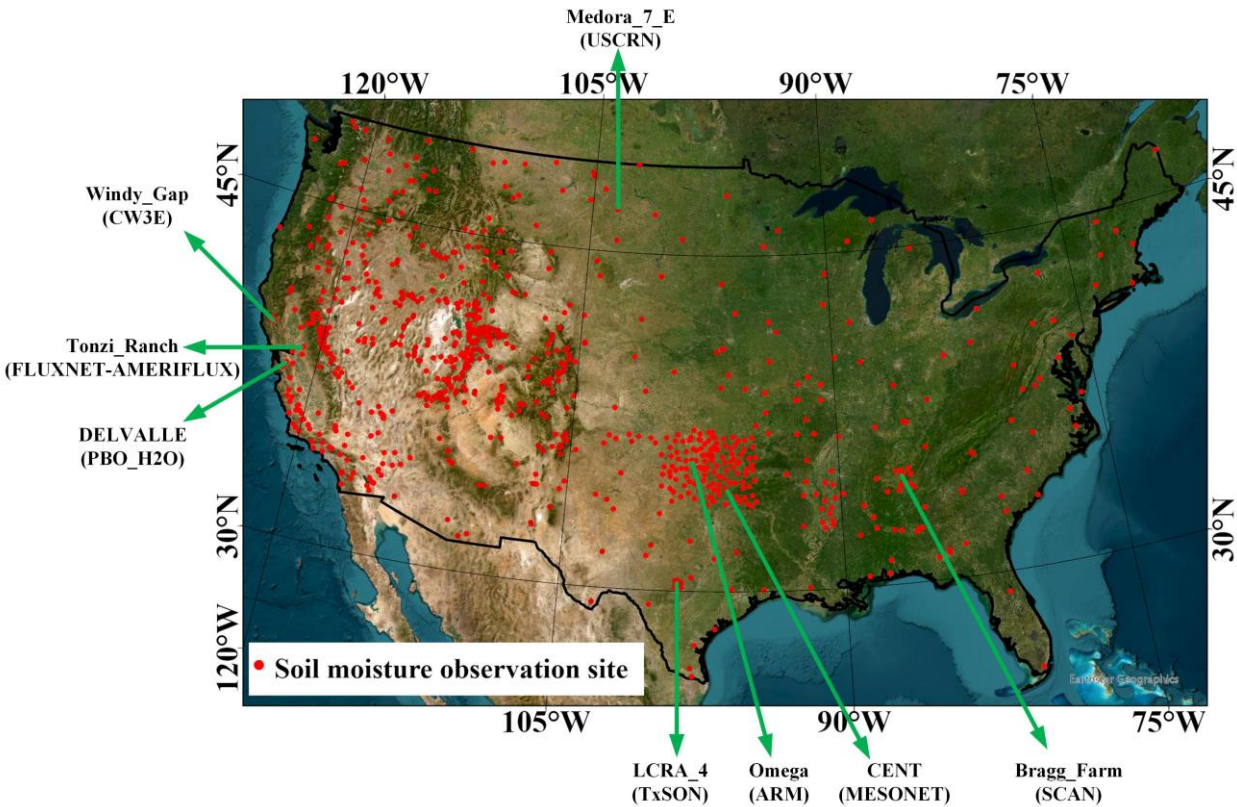
Figure 1. Spatial distribution of the *in-situ* soil moisture observation sites used in this study for soil moisture validation. Each red point refers to one site. The eight marked sites show temporal variations of Surface Soil Moisture (SSM) in Figures 6 and 7, respectively. The basemap is from Esri, Earthstar Geographics, and the GIS User Community.

2.1.4 Land cover and terrain data

In this study, we validated the performance of our developed datasets for different land cover types and different terrain conditions. Land cover type data were from the National Land Cover Database (NLCD) product available at https://www.mrlc.gov/, which includes annual land cover type at a spatial resolution of 30 m **(Homer et al., 2020; Jin et al., 2023)**. Additionally, 30-m digital elevation models (DEM) from the NASA Shuttle Radar Topography Mission project were utilized to describe the terrain information in the CONUS **(Rabus et al., 2003)**.

2.2 Method

2.2.1 Characteristic extraction

Because of the spatial gaps in the Crop-CASMA SSM data, it is difficult to directly exploit the daily Crop-CASMA SSM scene for the construction of the spatio-temporal fusion model (see next section). To deal with this problem, a time-series mining approach, i.e., highly comparative time-series analysis (HCTSA), was adopted to extract four spatially seamless 1-km characteristics (including maximum, minimum, mean and median) in each pixel from the Crop-CASMA SSM time-series data **(Fulcher et al., 2013; Fulcher and Jones, 2017)**. Additionally, to match the extracted characteristics from the Crop-CASMA SSM data, the same characteristics were also extracted at 9-km resolution from the corresponding SMAP L4 time-series data.

Even though the location of spatial gaps in the Crop-CASMA SSM time-series are varying over time, the extracted HCTSA-based characteristics are not affected **(Yang and Wang, 2023).** Typically, some factors that significantly influence SSM (e.g., precipitation, vegetation, and temperature) exhibit periodic changes in the year, indicating that inter-annual fluctuations in SSM tend to follow a periodic pattern. Thus, we selected a one-year temporal span for extracting these characteristics of maximum, minimum, mean, and median SSM. The extracted characteristics were then utilized to generate the corresponding STF_SSM scene.

2.2.2 Spatio-temporal fusion model

In this study, the virtual image pair-based spatio-temporal fusion (VIPSTF) model was

employed to generate the 3-hour, 1-km STF_SSM dataset, due to its stable performance, superior computational efficiency, and flexible usage **(Wang et al., 2020; Yang et al., 2023)**. The spatial weighting version of the VIPSTF model was adopted in this study, because of the reliable accuracy. The operation of the VIPSTF model requires at least one or multiple known image pairs at different spatial resolutions (a data pair is defined as one coarse and one fine resolution characteristic extracted from the same year). Here, the extracted 1-km and 9-km HCTSA-based characteristics (i.e., maximum, minimum, mean, and median of SSM time-series) from the SMAP L4 and Crop-CASMA SSM time-series (i.e., four image pairs) were blended using the VIPSTF model. Specifically, each STF_SSM scene is produced as follows:

$$\widehat{\textbf{STF\_SSM}}_t = \textbf{STF\_SSM}_{\text{VIP}} + \Delta\textbf{STF\_SSM}_t \, , \tag{1}$$

where $\widehat{\textbf{STF\_SSM}}_t$ is the generated 3-hour, 1-km SSM scene at time t, $\Delta\textbf{STF\_SSM}_t$ refers to the increment data of the model at a 1-km spatial resolution at time t. The virtual $\textbf{STF\_SSM}_{\text{VIP}}$ scene was predicted using a linear combination of the four extracted 1-km characteristics from the Crop-CASMA SSM time-series data in Eq. (2):

$$\textbf{STF\_SSM}_{\text{VIP}} = \sum_{i=1}^{n} a_i \textbf{F\_C}_i + b \, , \tag{2}$$

where a_i is the coefficient for the i-th extracted fine characteristic $\textbf{F\_C}_i$ at a 1-km spatial resolution. n is the number of the characteristics ($n = 4$ in this study), and b denotes a constant. Based on the assumption of scale invariance **(Wang and Atkinson, 2018)**, the optimal coefficients a_i and b for each 3-hour SMAP scene were calculated in a linear regression, as follows:

$$\textbf{SMAP}_t = \sum_{i=1}^{n} a_i \textbf{C\_C}_i + b + \Delta\textbf{SMAP}_t \, , \tag{3}$$

where $\textbf{SMAP}_t$ refers to the known SMAP L4 SSM scene at time t, and $\textbf{C\_C}_i$ is the i-th extracted coarse characteristic at a 9-km spatial resolution from the SMAP L4 SSM time-series data. $\Delta\textbf{SMAP}_t$ represents the residual data from the regression at time t. Moreover, the increment data (i.e., $\Delta\textbf{STF\_SSM}_t$) in Eq. (1) can be disaggregated by the following spatial weighting scheme:

$$\Delta\mathbf{STF\_SSM}_t(x_0,y_0)= \sum_{k=1}^{s} w_j\Delta\mathbf{SMAP}_t(x_k,y_k) \,, \tag{4}$$

In Eq. (4), (x_k,y_k) denotes the spatial distribution of the k-th similar pixel surrounding the center pixel (x_0,y_0). The number of surrounding similar pixels is represented by s. Additionally, w_j represents a weight calculated based on the distance between the center pixel and the k-th surrounding similar pixels. $\Delta\mathbf{SMAP}_t$ was interpolated from 9 km to 1 km using the bicubic method.

2.2.3 Data generation

Based on the VIPSTF model, the 3-hour, 1-km STF_SSM dataset was generated from 2015-04-01 to 2023-12-31. The generation flowchart is depicted in Figure 2. The specific process steps for producing the STF_SSM dataset are described as follows:

(1) For each year, 1-km and 9-km spatially seamless characteristics (i.e., maximum, minimum, mean, and median) were extracted from the Crop-CASMA and SMAP L4 SSM time-series using the HCTSA method, respectively.

(2) To generate a STF_SSM scene at time t, a VIPSTF model was constructed using the extracted characteristics in step (1) and a SMAP L4 SSM scene at time t within the year. Then, the STF_SSM scene at time t was generated.

(3) The aforementioned steps were repeated each 3-hour period to produce the 3-hour, 1-km STF_SSM data.

Finally, a total of 25,567 STF_SSM scenes were produced, accounting for approximately 1.78 TB. Each STF_SSM scene requires approximately 73.0 MB of storage space. The Pete High-Performance Computing (HPC) facility at Oklahoma State University was employed for data generation.
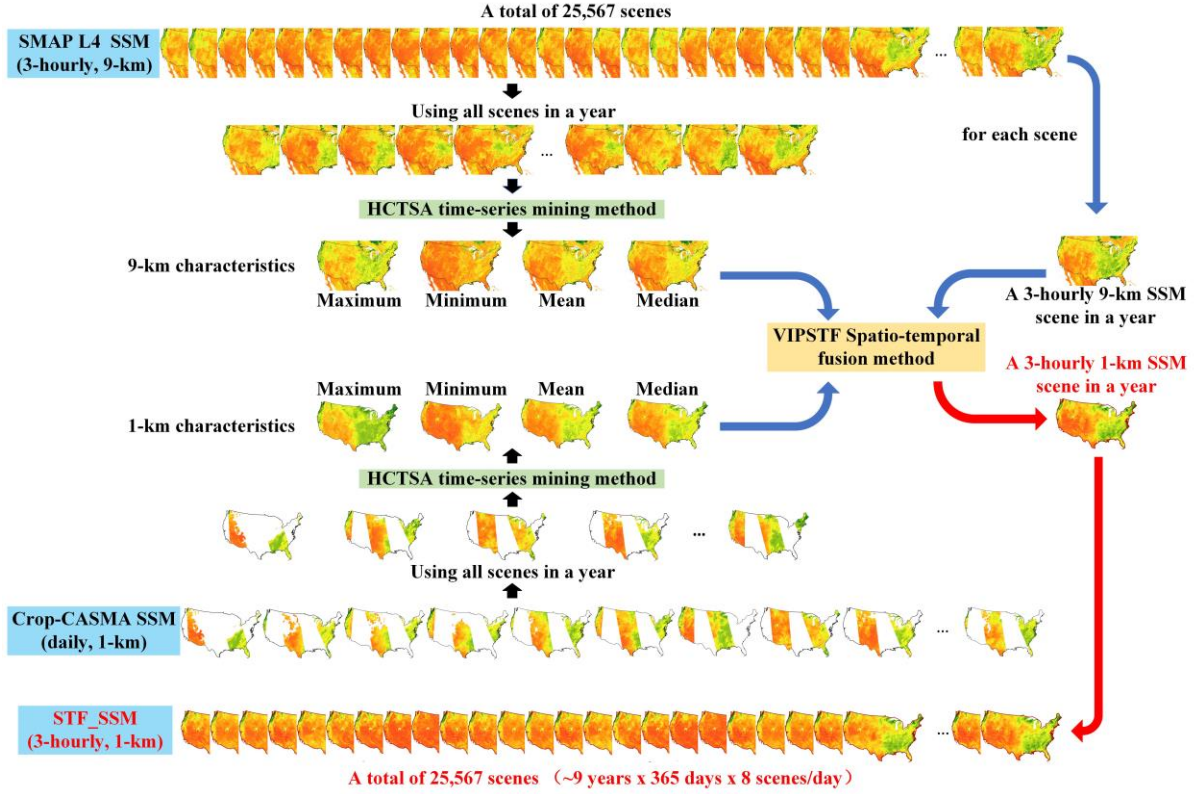
Figure 2. Flowchart for generating the 3-hour, 1-km STF_SSM dataset from 2015 to 2023.

2.3 Validation

In this paper, we divided the soil moisture observations (Figure 1) into two groups (3-hour group and daily group) for the validation of our generated data. Considering that the time of SMAP L4 and STF_SSM data is not on the hour (at 1:30, 4:30, 7:30, 10:30, 13:30, 16:30, 19:30, and 22:30), the average SSM value between two adjacent integer times was used to represent the SSM value at the specific time. For example, when validating the SMAP L4 and STF_SSM scenes at 4:30, the mean value of the 4:00 and 5:00 *in-situ* data was used to represent the soil moisture at 4:30. For daily scale validation, the hourly *in-situ* data were averaged to obtain daily values. Similarly, the 3-hour SMAP L4 and STF_SSM data were composited into daily scenes. To assess accuracy, five widely used statistical metrics were adopted, that is, the correlation coefficient (CC), root mean square error (RMSE), bias (Bias), unbiased root mean square error (ubRMSE), and Kling-Gutpa efficiency (KGE).

12

3. Results

3.1 Spatial pattern of the developed SSM dataset

The spatial pattern of the Crop-CASMA SSM, the SMAP L4 SSM, and the generated STF_SSM datasets are shown in Figure 3 at four randomly selected time points (i.e., 2015-04-01, 2017-06-08, 2019-08-16, and 2021-10-25). The 1-km Crop-CASMA SSM dataset has spatial gaps and does not have wall-to-wall data covering the entire CONUS. In contrast, both the SMAP L4 and STF_SSM datasets can provide spatially seamless observations. It is noted that the SMAP L4 SSM scenes contain some abnormal pixels with extremely high SSM values (SSM values of 0.6 and higher), especially in the northern part of the CONUS (e.g., some pixels around the Great Lakes in Figure 3).
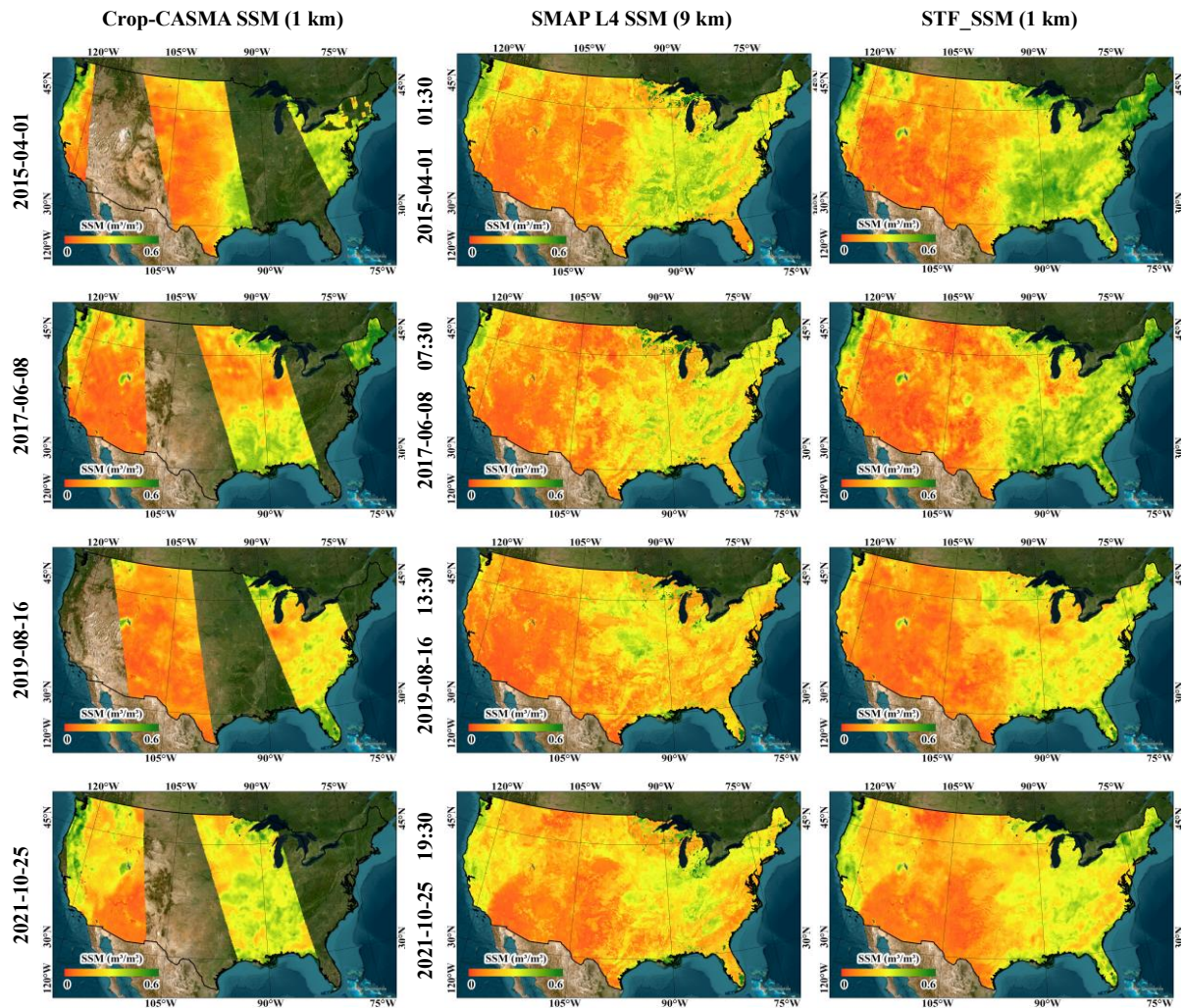
Figure 3. Spatial pattern of Surface Soil Moisture (SSM) in the Crop-CASMA SSM dataset (left), SMAP L4 SSM product (middle), and the STF_SSM dataset (right) on 2015-04-01 (01:30), 2017-06-08 (07:30), 2019-08-16 (13:30), and 2021-10-25 (19:30). Both the SMAP L4

and STF_SSM datasets are exhibited at the 3-hour scale, while the Crop-CASMA SSM dataset is displayed at the daily scale. The basemap is from Esri, Earthstar Geographics, and the GIS User Community.

To illustrate the advantages of the STF_SSM dataset, we zoomed into a sub-region in the CONUS at a date with rainfall (2018-08-14) and showed SSM in the three different datasets in Figure 4. It is clear that both the 3-hour SMAP L4 SSM and STF_SSM datasets can capture increased SSM values from 1:30 to 7:30 in the southwest region of the sub-region. Moreover, The STF_SSM and Crop-CASMA SSM datasets provide more detailed spatial information than those in the SMAP L4 SSM product. The spatial texture of the STF_SSM dataset closely resembles that of the 1-km Crop-CASMA SSM dataset, which is smoother than that of the SMAP L4 SSM product.

Next, we selected two random pixels in Nebraska (shown in Figure 4) to exhibit intra-day SSM variation (Figure 5). Although the Crop-CASMA SSM scene provides spatial information at a 1-km resolution, it only provides SSM value at the daily scale. In contrast, both the SMAP L4 and STF_SSM datasets show the changes in SSM every 3 hours. Furthermore, the changing patterns of SSM over time are similar between the SMAP L4 and STF_SSM datasets (the CC values in Figure 5a and Figure 5b are 0.997 and 0.999), indicating the stability and consistency of the STF_SSM dataset.
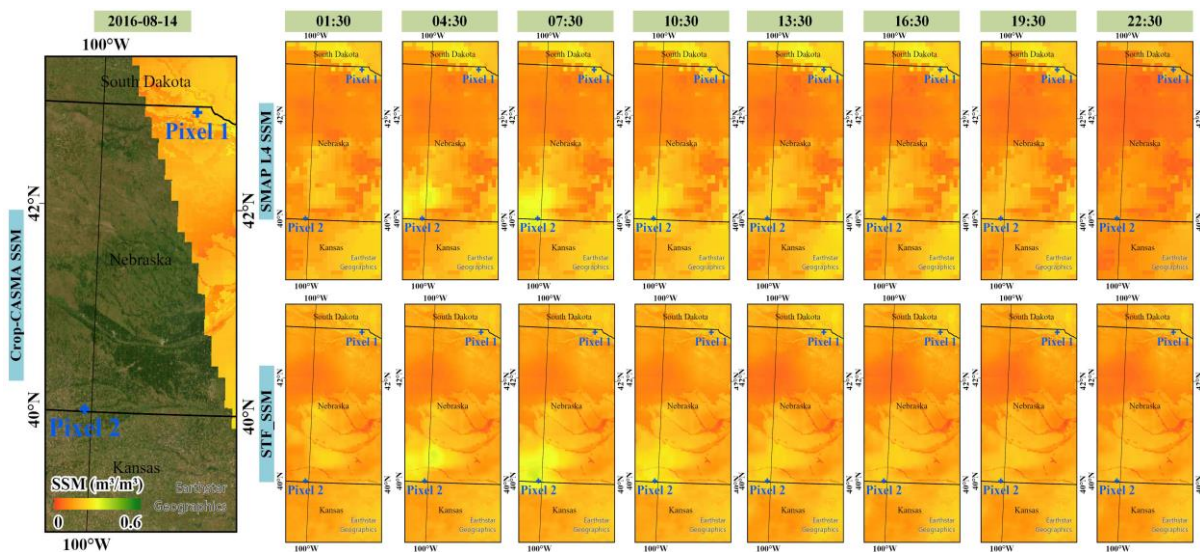

Figure 4. A sub-region in the western Continental United States (CONUS) to exhibit the spatial textures of the three Surface Soil Moisture (SSM) datasets on 2018-08-14. The basemap is from Esri, Earthstar Geographics, and the GIS User Community.
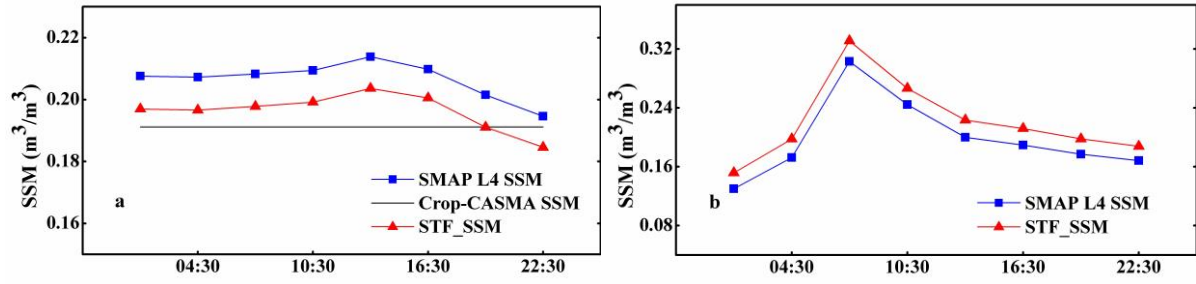
14

325

Figure 5. The intra-day Surface Soil Moisture (SSM) variation for two randomly selected pixels on 2018-08-14. The Crop-CASMA SSM dataset does not provide the intra-day SSM variation, shown as the flat black line. (a) and (b) refer to the pixels 1 and 2 in Figure 4, respectively. Since pixel 2 is located in the spatial gaps of the Crop-CASMA SSM scene, it

330 does not exhibit the flat black line in (b).

3.2 Validation based on daily soil moisture observations

At the daily scale, the three daily SSM datasets (the Crop-CASMA SSM, SMAP L4 SSM, and STF_SSM datasets) were compared against the daily *in-situ* data from 9 soil moisture observation networks. Figure 6 shows the SSM time-series acquired from four randomly

335 selected sites distributed across the CONUS: the Windy_Gap site in the CW3E network, CENT site in the MESONET network, DELVALLE site in the PBO_H2O network, and LCRA_4 site in the TxSON network. Although there are gaps in the *in-situ* data at these sites, the available *in-situ* data are sufficient for the validation of these three SSM datasets. Moreover, the SSM daily variations of the three SSM datasets are similar to those of the *in-situ* data. For example,

340 the CC values for the SMAP L4, Crop-CASMA, and STF_SSM datasets in Figure 6a (Windy_Gap site in the CW3E network) are 0.924, 0.879, and 0.886, respectively. In addition, there are some biases between different SSM datasets due to differences in spatial resolution and derived methods. Specifically, the SMAP records a lower minimum SSM value at the Windy_Gap site and a higher SSM maximum at the Delvalle site, compared with the Crop-
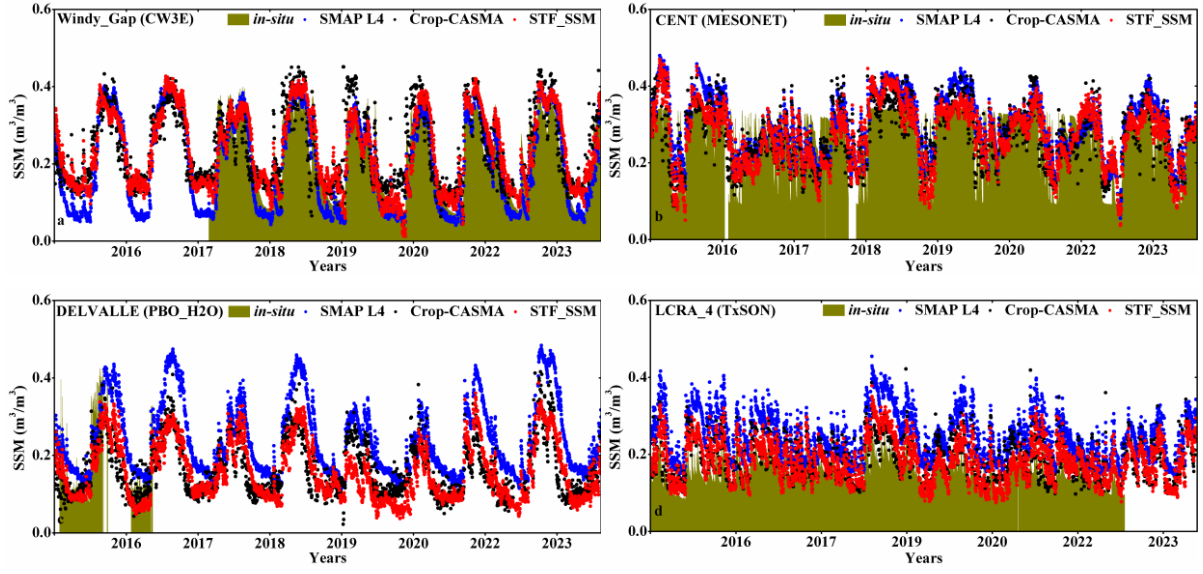
345 CASMA and STF_SSM datasets.

350

15

Figure 6. Temporal variations of daily surface soil moisture (SSM) from the SMAP L4 (blue), Crop-CASMA (black), STF_SSM (red), and *in-situ* observation data (dark yellow) at four different sites. (a) Windy_Gap site in the CW3E network. (b) CENT site in the MESONET network. (c) DELVALLE site in the PBO_H2O network. (d) LCRA_4 site in the TxSON network.

Table 2. Accuracy of the daily surface soil moisture (SSM) from the SMAP L4, the Crop-CASMA and the STF_SSM datasets. Values in bold indicate the dataset with the best performance for each statistic in each row. Underlined values indicate second best.

	SMAP L4 SSM (9 km)					Crop-CASMA SSM (1 km)					STF_SSM (1 km)				
	CC	RMSE (m^3/m^3)	Bias (m^3/m^3)	ubRMSE (m^3/m^3)	KGE	CC	RMSE (m^3/m^3)	Bias (m^3/m^3)	ubRMSE (m^3/m^3)	KGE	CC	RMSE (m^3/m^3)	Bias (m^3/m^3)	ubRMSE (m^3/m^3)	KGE
ARM	**0.752**	0.089	**-0.001**	**0.058**	0.538	<u>0.712</u>	<u>0.086</u>	0.030	0.069	<u>0.589</u>	0.710	**0.078**	<u>0.023</u>	<u>0.062</u>	**0.595**
CW3E	**0.859**	**0.070**	**-0.012**	**0.042**	**0.408**	0.531	0.111	<u>-0.046</u>	0.078	-0.058	<u>0.818</u>	<u>0.093</u>	-0.053	<u>0.047</u>	<u>0.181</u>
FLUXNET-AMERIFLUX	**0.925**	**0.055**	**-0.012**	**0.051**	**0.758**	0.875	<u>0.069</u>	<u>-0.028</u>	0.062	<u>0.740</u>	<u>0.892</u>	0.070	-0.034	<u>0.061</u>	0.645
MESONET	**0.721**	<u>0.096</u>	<u>0.042</u>	<u>0.064</u>	0.509	0.633	0.102	0.046	0.077	0.507	<u>0.710</u>	**0.089**	**0.035**	**0.064**	**0.525**
PBO_H2O	**0.724**	<u>0.063</u>	**0.002**	<u>0.050</u>	0.415	0.614	0.066	0.020	0.055	0.404	<u>0.704</u>	**0.061**	<u>0.018</u>	**0.050**	**0.445**
SCAN	**0.660**	0.087	**-0.012**	0.056	**0.191**	0.470	0.102	<u>-0.012</u>	0.078	0.007	<u>0.621</u>	<u>0.089</u>	-0.019	<u>0.057</u>	<u>0.155</u>
SNOTEL	**0.622**	**0.102**	**0.027**	**0.074**	**0.254**	0.335	0.120	0.036	0.093	0.091	<u>0.561</u>	<u>0.107</u>	<u>0.031</u>	<u>0.078</u>	<u>0.202</u>
TxSON	**0.771**	0.085	<u>-0.019</u>	0.045	<u>0.361</u>	0.598	<u>0.079</u>	0.025	0.059	0.340	<u>0.751</u>	**0.065**	**0.018**	**0.043**	**0.521**
USCRN	**0.709**	0.084	**-0.017**	0.052	0.357	0.488	0.109	<u>-0.025</u>	0.081	0.166	<u>0.675</u>	<u>0.097</u>	-0.038	<u>0.054</u>	<u>0.282</u>
Mean	**0.749**	**0.081**	**-0.001**	**0.055**	**0.421**	0.584	0.094	0.005	0.072	0.310	<u>0.716</u>	<u>0.083</u>	<u>-0.002</u>	<u>0.057</u>	<u>0.395</u>

At the network level, the results of the accuracy assessment (Table 2) show that the 9-km SMAP L4 SSM product has the greatest accuracy among the three datasets. At a spatial resolution of 1 km, the generated STF_SSM dataset outperforms the Crop-CASMA SSM dataset. Specifically, the mean CC for the SMAP L4 SSM product is 0.749, which is 0.033 and 0.165 higher than the STF_SSM and Crop-CASMA datasets, respectively. The RMSE and ubRMSE for the SMAP L4 SSM dataset are 0.081 m^3/m^3 and 0.055 m^3/m^3, which are the smallest among the three datasets. The STF_SSM dataset has RMSE and ubRMSE values of 0.083 m^3/m^3 and 0.057 m^3/m^3, which is slightly higher than those for the SMAP L4 SSM product but smaller than those for the Crop-CASMA dataset. The Bias for the SMAP L4 SSM product (with a Bias of -0.001 m^3/m^3) and the STF_SSM dataset (with a Bias of -0.002 m^3/m^3) are closer to SSM observation data than that of the Crop-CASMA SSM dataset. The STF_SSM dataset has a mean KGE of 0.395, which is 0.026 smaller than that of the SMAP L4 but 0.185 larger than that of the Crop-CASMA dataset. At the three SSM observation networks of MESONET, PBO_H2O, and TxSON, the STF_SSM dataset demonstrates better performance compared to the other datasets. However, for the other networks, the SMAP L4 SSM product outperforms the other datasets in terms of accuracy.

3.3 Validation based on 3-hour soil moisture observations

The 9-km SMAP L4 product and the 1-km STF_SSM dataset were compared with the 3-hour *in-situ* data from 7 networks. We randomly selected 4 sites (Omega site in the ARM network, Tonzi_Ranch site in the FLUXNET-AMERIFLUX network, Bragg_Farm site in the SCAN network, and Medora_7_E site in the USCRN network) to exhibit the SSM comparison (Figure 7). The temporal variations of SSM time-series from the *in-situ* data, SMAP L4 product, and STF_SSM dataset are similar to each other, e.g., the CC values of the SMAP L4 and STF_SSM datasets in Figure 7a are 0.749 and 0.745 by referring to the *in-situ* data, revealing that both the SSM datasets can well capture the dynamics of SSM at the 3-hour scale. Moreover, the difference between the 9-km SMAP L4 product and 1-km STF_SSM dataset is small, indicating that the downscaling of our STF_SSM dataset does not introduce significant errors in SMAP L4 SSM. Therefore, our STF_SSM dataset can be regarded as a reliable high-resolution version of the SMAP L4 product.
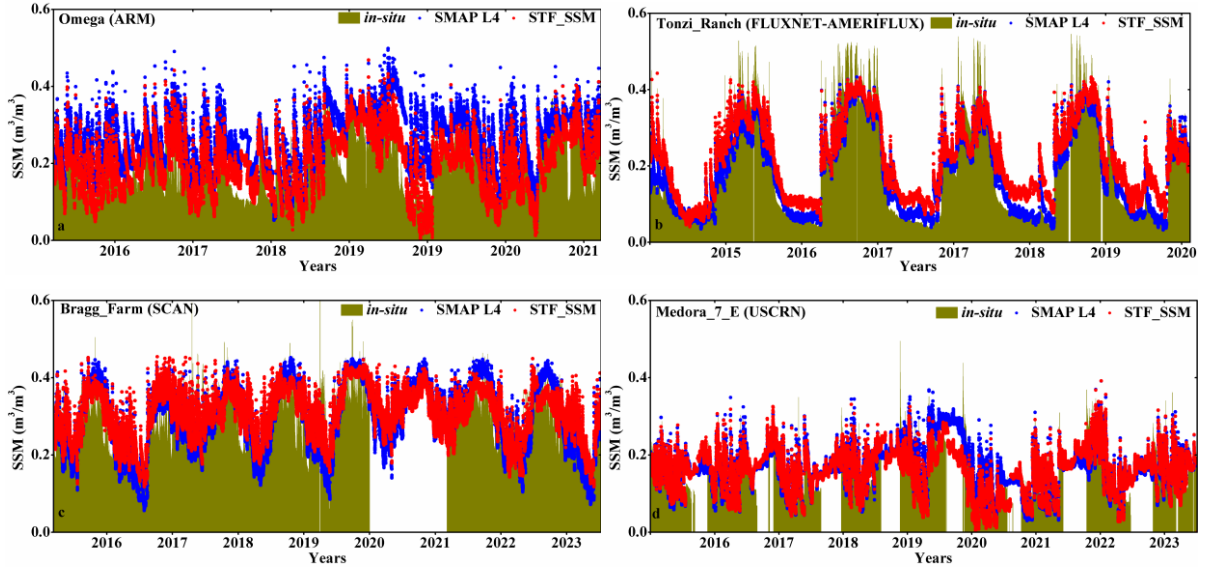
Figure 7. Temporal variations of 3-hour surface soil moisture (SSM) from SMAP L4 (blue), STF_SSM (red), and *in-situ* observation data (dark yellow) at four different sites. (a) Omega site in the ARM network. (b) Tonzi_Ranch site in the FLUXNET-AMERIFLUX network. (c) Bragg_Farm site in the SCAN network. (d) Medora_7_E site in the USCRN network.

The quantitative statistical metrics based on the 3-hour *in-situ* data are listed in Table 3. The results indicate that the 9-km SMAP L4 SSM product has better accuracy than the 1-km STF_SSM dataset. Specifically, the mean CC and KGE for the SMAP L4 SSM product are 0.728 and 0.414, which are 0.039 and 0.028 higher than those for the STF_SSM dataset. Furthermore, the SMAP L4 SSM product has a mean RMSE and ubRMSE of 0.086 m^3/m^3 and 0.059 m^3/m^3. Additionally, the Bias of the SMAP L4 SSM product is -0.005 m^3/m^3, which is closer to the SSM observation data than that of the STF_SSM dataset. For specific networks, we found that the CC and KGE of the SMAP L4 SSM dataset at the FLUXNET-AMERIFLUX network are 0.920 and 0.889, which are the highest among the 7 networks. On the other hand, the STF_SSM dataset provides an RMSE of 0.081 m^3/m^3 and 0.067 m^3/m^3 for the ARM and TxSON networks, which is 0.009 m^3/m^3 and 0.019 m^3/m^3 less than the SMAP L4 SSM dataset, respectively.

19

Table 3. Accuracy of the 3-hour surface soil moisture (SSM) from the SMAP L4 SSM product and the STF_SSM datasets. Values in bold indicate the dataset with better performance for each statistic in each row.

	SMAP L4 SSM (9 km)					STF_SSM (1 km)				
	CC	RMSE (m^3/m^3)	Bias (m^3/m^3)	ubRMSE (m^3/m^3)	KGE	CC	RMSE (m^3/m^3)	Bias (m^3/m^3)	ubRMSE (m^3/m^3)	KGE
ARM	**0.744**	0.090	**0.001**	**0.060**	0.544	0.696	**0.081**	0.023	0.065	**0.593**
CW3E	**0.818**	0.074	**-0.008**	**0.046**	**0.398**	0.777	0.097	-0.047	0.051	0.216
FLUXNET-AMERIFLUX	**0.920**	0.060	**-0.009**	**0.055**	**0.746**	0.889	0.073	-0.030	0.066	0.645
SCAN	**0.628**	0.097	**-0.011**	**0.065**	**0.305**	0.582	0.097	-0.016	0.066	0.304
SNOTEL	**0.512**	**0.111**	**0.021**	**0.084**	**0.189**	0.457	0.114	0.033	0.086	0.136
TxSON	**0.769**	0.086	**-0.018**	0.047	0.358	0.750	**0.067**	0.020	**0.045**	**0.514**
USCRN	**0.708**	**0.088**	**-0.012**	**0.055**	**0.357**	0.674	0.097	-0.032	0.057	0.295
Mean	**0.728**	**0.086**	**-0.005**	**0.059**	**0.414**	0.689	0.090	-0.007	0.062	0.386

3.4 SSM data accuracy across land cover types

Generally, the accuracy of satellite-derived SSM datasets varies between land cover types, because the penetration capacity of remote sensing signals can be affected by land cover types. Therefore, it is necessary to assess the performance of the SSM datasets under different land cover types. According to the NLCD land cover product from 2015 to 2023, we separated the CONUS into nine types, including the developed, barren, forest, shrub, grassland, pasture, crops, wetlands, and changed. The former eight types are the existing categories (e.g., shrub and grassland) or composite categories (e.g., forest is composed of deciduous, evergreen, and mixed forest) in the NLCD product. The "changed" category refers to the areas where land use has changed between 2015 and 2023.

At the 3-hour scale, it is seen from Figures 8a and 8b that the SMAP L4 product has slightly better performance than the generated STF_SSM dataset for most land cover types. As shown in Table 4, the mean CC and RMSE values of the SMAP L4 product are 0.611 and 0.107, which are 0.042 and 0.002 better than those of the STF_SSM dataset, respectively. However, both SSM datasets exhibit lower accuracy in wetlands compared to other land cover types. Meanwhile, the SSM accuracy has a larger variation across forest pixels, shrub pixels, and developed area pixels than the other land cover types. This is primarily because the topsoil in these land cover types is covered by woody plants and man-made features, which influences

SSM observations and leads to a loss of accuracy. The highest accuracy is observed in barren

435 (with an RMSE of 0.083 m^3/m^3 for the STF_SSM dataset) and grassland areas (with an RMSE

of 0.083 m^3/m^3 for the SMAP L4 product), as these types tend to have less cover. The land

cover changes observed in the "changed" category also provide satisfactory accuracy. This is

because approximately 50% of the land cover change samples consist of barren or grassland

areas, which contribute to higher SSM accuracy. These patterns are also reflected in the daily

440 SSM datasets (Figures 8c and 8d). Moreover, the Crop-CASMA dataset (with mean CC and

RMSE values of 0.440 and 0.111, respectively) has a lower performance than the STF_SSM

dataset across all land use covers, highlighting the reliability of the generated STF_SSM dataset

at a 1-km spatial resolution.
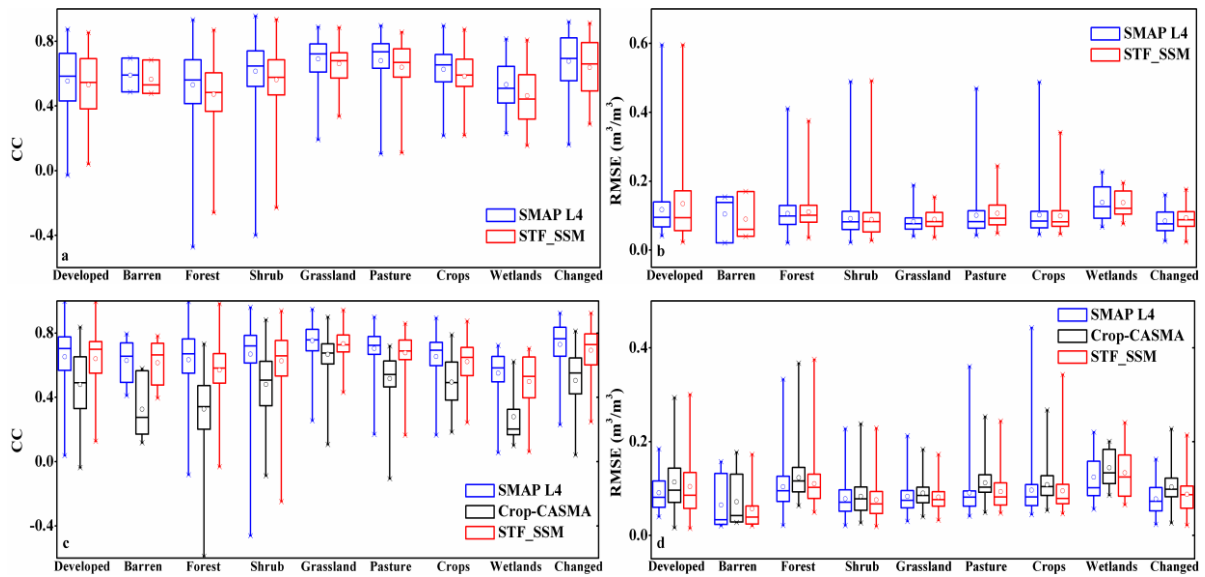
Figure 8. Accuracy of the surface soil moisture (SSM) datasets under different land cover types. The "changed" type refers to areas where land cover type changed between 2015 and 2023. (a) and (b) are the correlation coefficient (CC) and the root mean square error (RMSE) of SSM datasets at the 3-hour scale. (c) and (d) are the CC and RMSE of SSM datasets at the daily

450 scale. For each land cover type, the number of validation sites is 45 (developed), 7 (barren), 190 (forest), 211 (shrub), 148 (grassland), 97 (pasture), 69 (crops), 10 (wetlands), and 49 (changed), respectively.

455

460

Table 4. Mean correlation coefficient (CC) and root mean square error (RMSE) of surface soil moisture (SSM) datasets under different land cover types.

Land cover type (number of validation sites)	3-hour				Daily					
	SMAP L4		STF_SSM		SMAP L4		Crop-CASMA		STF_SSM	
	CC	RMSE (m^3/m^3)	CC	RMSE (m^3/m^3)	CC	RMSE (m^3/m^3)	CC	RMSE (m^3/m^3)	CC	RMSE (m^3/m^3)
Developed (45)	0.554	0.117	0.530	0.135	0.653	0.091	0.480	0.114	0.640	0.104
Barren (7)	0.591	0.105	0.564	0.083	0.601	0.103	0.208	0.117	0.576	0.092
Forest (190)	0.530	0.151	0.472	0.156	0.634	0.104	0.327	0.123	0.570	0.110
Shrub (211)	0.615	0.091	0.562	0.089	0.669	0.078	0.481	0.083	0.627	0.076
Grassland (148)	0.691	0.083	0.662	0.089	0.753	0.083	0.669	0.090	0.734	0.082
Pasture (97)	0.680	0.101	0.639	0.107	0.704	0.091	0.517	0.112	0.678	0.093
Crops (69)	0.626	0.102	0.585	0.099	0.654	0.097	0.494	0.108	0.623	0.095
Wetlands (10)	0.532	0.138	0.463	0.138	0.551	0.124	0.279	0.145	0.498	0.134
Changed (49)	0.677	0.084	0.639	0.094	0.730	0.078	0.505	0.104	0.693	0.088
Mean	0.611	0.108	0.569	0.110	0.661	0.094	0.440	0.111	0.627	0.097

3.5 SSM data accuracy across topographic conditions

As a significant soil-forming factor, terrain is one of the determinants of SSM variations. Particularly, SSM could have strong spatial variability in areas with complex topographic conditions. We analyzed the accuracy (including CC and RMSE) of the 1-km Crop-CASMA and STF_SSM datasets under different topographic conditions. As shown in Figures 9a and 9b, both the Crop-CASMA and STF_SSM datasets show a decrease in the CC value with increasing elevation. However, the STF_SSM dataset shows a slower decline in accuracy (with a slope of -0.055) compared to the Crop-CASMA dataset (with a slope of -0.100). Under complex terrain conditions (i.e., larger slope), the accuracy of both SSM datasets is reduced. It can be seen from Figures 9c and 9d that the CC of the Crop-CASMA dataset decreases more sharply as the slope increases (with a slope of -0.023) while the CC in the generated STF_SSM dataset declines more gradually (with a slope of -0.011). Likewise, Figures 9e and 9f shows that the RMSE values of both SSM datasets increase with elevation. According to the intercept, the STF_SSM dataset has a slightly greater RMSE than the Crop_CASMA dataset, especially at high altitudes. Meanwhile, with an increase in slope, the STF_SSM dataset has a slower rise

in RMSE values than the Crop-CASMA dataset (Figure 9g and 9h). This suggests that the STF_SSM dataset is more reliable than the Crop-CASMA dataset in complex terrain conditions.
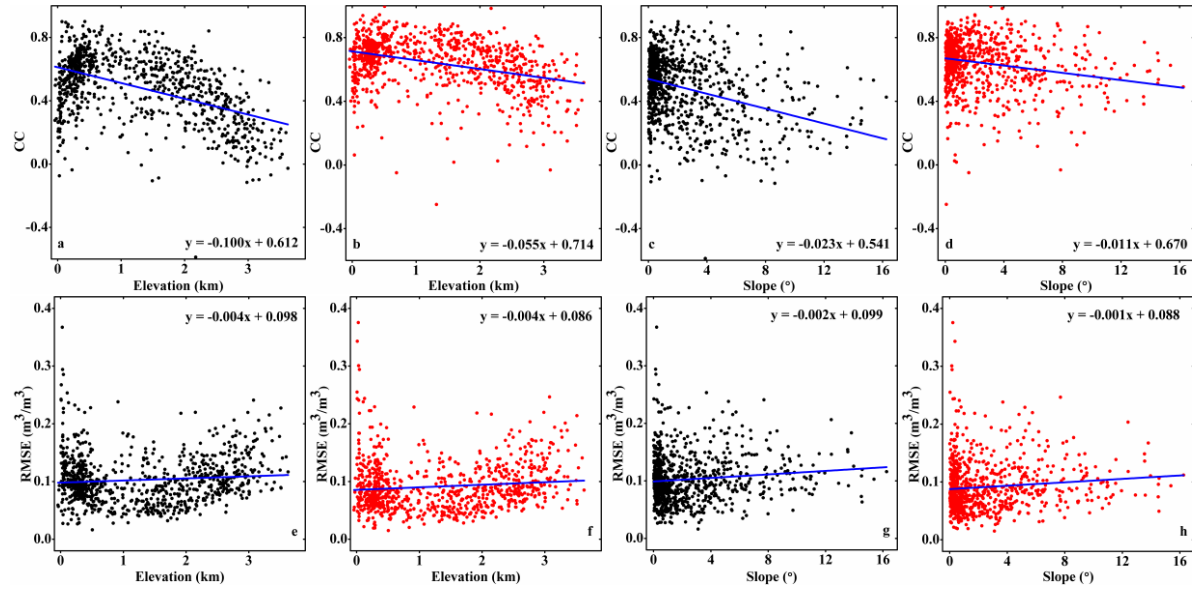

Figure 9. Accuracy of the estimated surface soil moisture (SSM) from the 1-km Crop-CASMA and generated STF_SSM datasets along changing topographic conditions, denoted by elevation and slope. (a), (c), (e) and (g) refer to the 1-km Crop-CASMA SSM dataset. (b), (d), (f) and (h) are the generated STF_SSM dataset.

4. Discussion

4.1 Implications

Ma et al. (2021) adopted the daily 25-km, 10-day composited CCI SSM time-series to define agricultural drought events. However, this 10-day composition approach may not capture the rapidly developed drought phenomena (e.g., flash drought). In contrast, the generated 3-hour, 1-km STF_SSM dataset has advantages over the CCI dataset, because the STF_SSM dataset provides more detailed and continuous SSM information in both the temporal and spatial dimensions. This implies that the STF_SSM dataset can detect both long-term and flash drought events at a finer scale.

Figure 10 exhibits the SSM variation under four drought events in Oklahoma (January to February, 2018), Alabama (May to December, 2016), California (January to December, 2020), and Nevada (January to December, 2020) **(U.S. Drought Monitor, 2024)**. The SSM in our dataset shows a clear response to the drought event. As shown in Figure 10a, during the Oklahoma's drought in early 2018 **(Shephard et al., 2021)**, SSM is about 0.08 m^3/m^3 lower than the multi-year average value from 2015 to 2023. The drought was gradually alleviated in February. Alabama's drought in 2016 began around May and continued into December (Figure 10b) **(Noel et al., 2020)**, as the SSM value began to deviate from the average in May and remained in a lower range. In 2020, California and Nevada suffered extreme and long-term droughts **(Williams et al., 2022),** Figures 10c and 10d show that SSM values in the two states were generally less than the average and continued for the entire year.
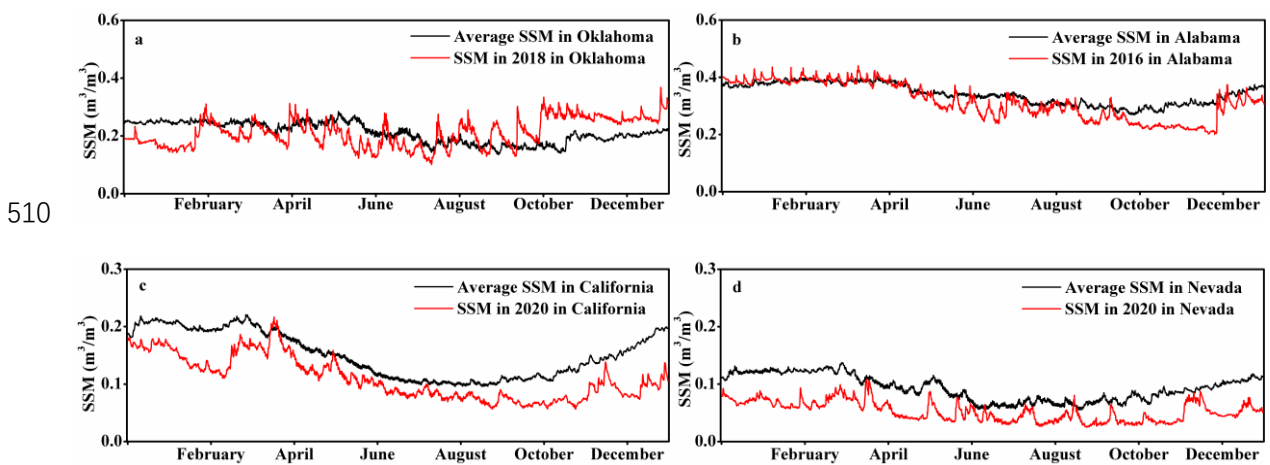

Figure 10. Surface soil moisture (SSM) variations under drought events. Black lines refer to the average SSM values calculated from 2015 to 2023. Red lines represent the SSM values for

the corresponding year. (a) the drought in Oklahoma in 2018. (b) the drought in Alabama in 2016. (c) the drought in California in 2020. (b) the drought in Nevada in 2020.

In addition to droughts, SSM is also sensitive to flooding. When a flooding event begins, the SSM value is usually rapidly increased over a short period. To highlight the advantages of the developed 3-hour SSM dataset, we portrayed the SSM variation under two flooding events. Figure 11a shows the flooding in 2015 in Williamsburg, South Carolina, because of extreme precipitation from 2015-10-01 to 2015-10-05. It can be seen that the SSM value in this region began to increase dramatically from the evening of 2015-10-01 to 2015-10-02, and remained at a high level until 2015-10-05. Figure 11b presents the flooding in 2017 in Jefferson, Texas, due to Hurricane Harvey. We found that the SSM value had started to rise on the evening of 2017-08-24 before Hurricane Harvey reached landfall fully (2017-08-25), and peaked on 2017-08-27.
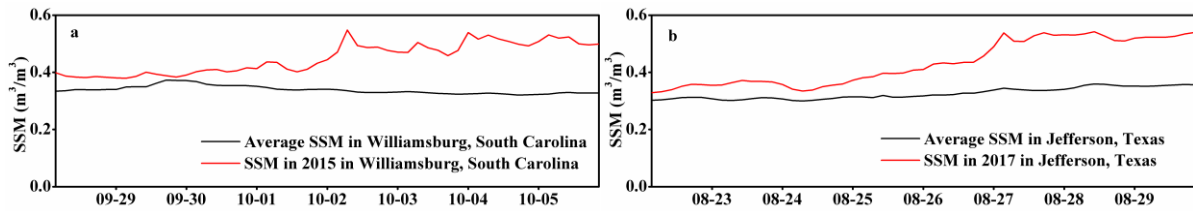

Figure 11. Surface soil moisture (SSM) variations under flood events. Black lines represent the average SSM values calculated from 2015 to 2023. Red lines are the SSM values for the corresponding year. (a) the flood in Williamsburg, South Carolina in 2015. (b) the flood in Jefferson, Texas in 2017.

It is clear from the mentioned analysis that SSM information is closely linked to drought and flooding. This suggests that SSM can be applied to identify these events and quantify their severity. Thus, the developed STF_SSM dataset has great potential for application, especially in agriculture. For instance, near real-time crop conditions could be observed directly by dynamically monitoring SSM. It will provide a rational basis for refining irrigation management. In addition, severe drought and flooding affect crop yields, implying that SSM information with fine spatio-temporal resolution also has the potential to play an important role in crop yield estimation.

4.2 Accuracy and latency time of updated STF_SSM data

Drought monitoring needs real-time or near real-time soil moisture data, which requires

25

high data accuracy and short latency time. Since the strategies of characteristics extraction for training the STF_SSM dataset from 2015 to 2023 depend on the complete SSM time-series data over the entire year, it is difficult to directly update the near real-time data using this strategy. Therefore, we propose three alternative strategies for characteristics extraction to examine on a randomly selected date in 2023 (2023-01-10). This first alternative is using all data in 2022; the second alternative is using available data in 2023 (before 2023-01-10); the third is using all data in 2022 plus available data in 2023. By referring to the corresponding STF_SSM scenes (predict using all data in 2023), we found that using available data in 2023 has the greatest performance among the three strategies of characteristics extraction, with a mean CC and RMSE of 0.999 and 0.006 m^3/m^3 (Table 5). Meanwhile, the CC and KGE of using all data in 2022 plus available data in 2023 are 0.964 and 0.947, which are 0.004 and 0.009 higher than those using data only in 2022. This means that using characteristics extracted from temporally adjacent data in 2023 improves fusion accuracy.

Due to a data-driven approach to production, real-time updates of the STF_SSM data have unavoidable latency time. This is because the latency time of the STF_SSM dataset depends on that of other auxiliary data. According to the investigation from the official website, the near real-time SMAP L4 SSM product and Crop-CASMA SSM dataset usually have three and two days of latency, respectively. Thus, if only available data within the year are adopted to update the STF_SSM data, the latency time of the near real-time STF_SSM scene is at least three days.

Table 5. Accuracy of near real-time STF_SSM data (on 2023-01-10) production using different strategies of characteristics extraction.

	CC	RMSE (m^3/m^3)	Bias (m^3/m^3)	ubRMSE (m^3/m^3)	KGE
Accuracy using all data in 2022	0.960	0.035	-0.010	0.033	0.938
Accuracy using all data in 2022 plus available data in 2023	0.964	0.032	-0.008	0.031	0.947
Accuracy using available data in 2023	0.999	0.006	0.000	0.006	0.992

4.3 Analysis of different fine SSM datasets

Currently, some high-resolution SSM datasets have been published and used. We listed six 1-km SSM datasets at a large scale and exhibited the details of these datasets in Table 6, such as the spatial resolution, temporal resolution, and accuracy. Given the differences in validation

methods, spatial and temporal coverages, and statistical metrics, etc, it is difficult to harmonize these datasets to the same standard to quantify accuracy. Therefore, before using the data, it is necessary to further select the more suitable SSM dataset according to the requirements.

Table 6. Comparative analysis for six high-resolution SSM datasets. Statistical metrics for accuracy include root mean square error (RMSE), unbiased root mean square error (ubRMSE), and unbiased root mean square deviation(ubRMSD).

Reference	Area	Spatial resolution	Temporal resolution	Accuracy (m^3/m^3)
This study	CONUS	1-km	3-hour	ubRMSE = 0.057
(Vergopolan et al., 2021)	CONUS	30-m	6-hour	RMSE = 0.07
(Fang et al., 2022)	Global	1-km	daily	ubRMSE = 0.063
(Zheng et al., 2023)	Global	1-km	daily	ubRMSE = 0.045
(Han et al., 2023)	Global	1-km	daily	ubRMSE = 0.050
(Song et al., 2022)	China	1-km	daily	ubRMSD = 0.074

4.4 Uncertainties and future works

Although the STF_SSM dataset has good performance in representing the fast changes in soil moisture, two uncertainties in the data generation need to be noted. First, the spatio-temporal fusion model used in this study is a data-driven method, which depends on the stable and accurate SMAP L4 SSM product and Crop-CASMA SSM dataset. If either of these datasets stops updating or contains significant errors, the generation and accuracy of the STF_SSM dataset will be impacted. Second, many environmental and ecological variables affect the SSM, such as precipitation, vegetation, temperature, evaporation, and terrain. However, these variables are not fully considered in the STF_SSM production, decreasing the interpretability of the STF_SSM dataset.

Currently, geostationary satellites have a large potential to provide hourly and spatially fine auxiliary variables to produce SSM datasets. Fusing the auxiliary variables from the geostationary satellites for the generation of hourly SSM data is an ongoing work. However, the extensive data acquisition and necessary preprocessing steps can significantly increase the time cost of data production without leading to a substantial improvement in accuracy. Thus, balancing data accuracy and generation efficiency is necessary for the downscaling of the SSM dataset in the future. Compared with existing SSM datasets, the generated SSM dataset in this

590   study has advantages in terms of spatio-temporal resolution. It is worthwhile to further explore its potential applications, such as monitoring drought severity and occurrences, quantifying wildfire danger levels, evaluating responses of agriculture and natural ecosystems to soil moisture dynamics, and understanding local and regional hydrological processes.

5. Conclusions

In this study, we developed a spatio-temporal fusion model to generate the first spatially seamless 3-hour, 1-km STF_SSM dataset in the CONUS from 2015 to 2023. This dataset integrated the 3-hour, 9-km SMAP L4 SSM product from NASA and the daily, 1-km Crop-CASMA SSM dataset from USDA-NASS. The former provided fine temporal resolution and the latter provided detailed spatial details. To deal with the mismatch between the two datasets in terms of temporal resolution and spatial coverage, the HCTSA-based time-series mining method was used to extract spatially seamless characteristics from both SSM time-series data for each year. Four characteristics extracted at 1-km and 9-km spatial resolutions (minimum, maximum, mean, and median of SSM time-series) were employed as inputs of the fusion model. By coupling with each 3-hour, 9-km SMAP L4 scene, the downscaled 3-hour, 1-km STF_SSM scene was simulated, in turn. Data validation at the daily scale showed that the generated 1-km STF_SSM dataset (with a mean CC and ubRMSE of 0.716 and 0.057 m^3/m^3) outperforms the 1-km Crop-CASMA SSM dataset (with a mean CC and ubRMSE of 0.584 and 0.072 m^3/m^3) when compared to *in-situ* measurements. At the 3-hour scale, the accuracy of the 9-km SMAP L4 SSM product (with a mean CC and ubRMSE of 0.728 and 0.059 m^3/m^3) is slightly higher than that of the 1-km STF_SSM dataset (with a mean CC and ubRMSE of 0.689 and 0.062 m^3/m^3). Additionally, the STF_SSM dataset has a better performance than the Crop-CASMA dataset under complex terrain conditions. Overall, the generated 3-hour, 1-km STF_SSM dataset is reliable and has great potential for applications at various spatio-temporal scales. The proposed STF_SSM dataset can be freely acquired from https://doi.org/10.6084/m9.figshare.28188011.

Data availability

The STF_SSM dataset is available at https://doi.org/10.6084/m9.figshare.28188011 (Yang
et al., 2025). The code for generating the STF_SSM dataset can be downloaded at
https://github.com/hhhhhaoxuanyang/STF_SSM-dataset.git.

Author contributions

Conceptualization: HY. Data curation: HY, JY, TEO. Formal analysis: HY, JY, MX.
Funding acquisition: JY, TEO, MX. Methodology: HY. Software: HY, JY. Resources: HY, JY,
TEO, ESK, MX, CBZ. Supervision: JY, TEO, CBZ. Validation: HY, TEO, ESK. Visualization:
HY. Writing – original draft preparation: HY. Writing – review & editing: HY, JY, TEO, ESK,
MX, CBZ. All authors have reviewed and approved the manuscript.

Competing interests

At least one of the (co-)authors is a member of the editorial board of Earth System Science
Data.

References

650 Abbaszadeh, P., Moradkhani, H., and Zhan, X.: Downscaling SMAP Radiometer Soil Moisture Over the CONUS Using an Ensemble Learning Method, Water Resources Research, 55, 324–344, https://doi.org/10.1029/2018WR023354, 2019.

Abbaszadeh, P., Moradkhani, H., Gavahi, K., Kumar, S., Hain, C., Zhan, X., Duan, Q., Peters-Lidard, C., and Karimiziarani, S.: High-Resolution SMAP Satellite Soil Moisture Product: Exploring the

655 Opportunities, Bulletin of the American Meteorological Society, 102, 309–315, https://doi.org/10.1175/BAMS-D-21-0016.1, 2021.

Abowarda, A. S., Bai, L., Zhang, C., Long, D., Li, X., Huang, Q., and Sun, Z.: Generating surface soil moisture at 30 m spatial resolution using both data fusion and machine learning toward better water resources management at the field scale, Remote Sensing of Environment, 255, 112301,

660 https://doi.org/10.1016/j.rse.2021.112301, 2021.

Alaminie, A. A., Annys, S., Nyssen, J., Jury, M. R., Amarnath, G., Mekonnen, M. A., and Tilahun, S. A.: A comprehensive evaluation of satellite-based and reanalysis soil moisture products over the upper Blue Nile Basin, Ethiopia, Science of Remote Sensing, 10, 100173, https://doi.org/10.1016/j.srs.2024.100173, 2024.

665 Brocca, L., Gaona, J., Bavera, D., Fioravanti, G., Puca, S., Ciabatta, L., Filippucci, P., Mosaffa, H., Esposito, G., Roberto, N., Dari, J., Vreugdenhil, M., and Wagner, W.: Exploring the actual spatial resolution of 1 km satellite soil moisture products, Science of The Total Environment, 945, 174087, https://doi.org/10.1016/j.scitotenv.2024.174087, 2024.

Carlson, T.: An Overview of the "Triangle Method" for Estimating Surface Evapotranspiration and Soil

670 Moisture from Satellite Imagery, Sensors, 7, 1612–1629, https://doi.org/10.3390/s7081612, 2007.

Carlson, T. N., Perry, E. M., and Schmugge, T. J.: Remote estimation of soil moisture availability and fractional vegetation cover for agricultural fields, Agricultural and Forest Meteorology, 52, 45–69, https://doi.org/10.1016/0168-1923(90)90100-K, 1990.

Carlson, T. N., Gillies, R. R., and Schmugge, T. J.: An interpretation of methodologies for indirect

675 measurement of soil water content, Agricultural and Forest Meteorology, 77, 191–205, https://doi.org/10.1016/0168-1923(95)02261-U, 1995.

Chan, S. K., Bindlish, R., O'Neill, P. E., Njoku, E., Jackson, T., Colliander, A., Chen, F., Burgin, M., Dunbar,

32

S., Piepmeier, J., Yueh, S., Entekhabi, D., Cosh, M. H., Caldwell, T., Walker, J., Wu, X., Berg, A., Rowlandson, T., Pacheco, A., McNairn, H., Thibeault, M., Martinez-Fernandez, J., Gonzalez-Zamora, A., Seyfried, M., Bosch, D., Starks, P., Goodrich, D., Prueger, J., Palecki, M., Small, E. E., Zreda, M., Calvet, J.-C., Crow, W. T., and Kerr, Y.: Assessment of the SMAP Passive Soil Moisture Product, IEEE Trans. Geosci. Remote Sensing, 54, 4994–5007, https://doi.org/10.1109/TGRS.2016.2561938, 2016.

Chen, F., Crow, W. T., Bindlish, R., Colliander, A., Burgin, M. S., Asanuma, J., and Aida, K.: Global-scale evaluation of SMAP, SMOS and ASCAT soil moisture products using triple collocation, Remote Sensing of Environment, 214, 1–13, https://doi.org/10.1016/j.rse.2018.05.008, 2018.

Colliander, A., Yang, Z., Mueller, R., Sandborn, A., Reichle, R., Crow, W., Entekhabi, D., and Yueh, S.: Consistency Between NASS Surveyed Soil Moisture Conditions and SMAP Soil Moisture Observations, Water Resources Research, 55, 7682–7693, https://doi.org/10.1029/2018WR024475, 2019.

Cui, H., Jiang, L., Du, J., Zhao, S., Wang, G., Lu, Z., and Wang, J.: Evaluation and analysis of AMSR-2, SMOS, and SMAP soil moisture products in the Genhe area of China, JGR Atmospheres, 122, 8650–8666, https://doi.org/10.1002/2017JD026800, 2017.

Das, N. N., Entekhabi, D., Dunbar, R. S., Chaubell, M. J., Colliander, A., Yueh, S., Jagdhuber, T., Chen, F., Crow, W., O'Neill, P. E., Walker, J. P., Berg, A., Bosch, D. D., Caldwell, T., Cosh, M. H., Collins, C. H., Lopez-Baeza, E., and Thibeault, M.: The SMAP and Copernicus Sentinel 1A/B microwave active-passive high resolution surface soil moisture product, Remote Sensing of Environment, 233, 111380, https://doi.org/10.1016/j.rse.2019.111380, 2019.

Djamai, N., Magagi, R., Goïta, K., Merlin, O., Kerr, Y., and Roy, A.: A combination of DISPATCH downscaling algorithm with CLASS land surface scheme for soil moisture estimation at fine scale during cloudy days, Remote Sensing of Environment, 184, 1–14, https://doi.org/10.1016/j.rse.2016.06.010, 2016.

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

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., Kimball, J., Piepmeier, J. R., Koster, R. D., Martin, N.,

McDonald, K. C., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., Spencer, M. W., Thurman, S. W., Tsang, L., and Van Zyl, J.: The Soil Moisture Active Passive (SMAP) Mission, Proc. IEEE, 98, 704–716, https://doi.org/10.1109/JPROC.2010.2043918, 2010.

Fan, D., Zhao, T., Jiang, X., García-García, A., Schmidt, T., Samaniego, L., Attinger, S., Wu, H., Jiang, Y., Shi, J., Fan, L., Tang, B.-H., Wagner, W., Dorigo, W., Gruber, A., Mattia, F., Balenzano, A., Brocca, L., Jagdhuber, T., Wigneron, J.-P., Montzka, C., and Peng, J.: A Sentinel-1 SAR-based global 1-km resolution soil moisture data product: Algorithm and preliminary assessment, Remote Sensing of Environment, 318, 114579, https://doi.org/10.1016/j.rse.2024.114579, 2025.

Fang, B., Lakshmi, V., Cosh, M., Liu, P., Bindlish, R., and Jackson, T. J.: A global 1-km downscaled SMAP soil moisture product based on thermal inertia theory, Vadose Zone Journal, 21, e20182, https://doi.org/10.1002/vzj2.20182, 2022.

Fulcher, B. D. and Jones, N. S.: hctsa : A Computational Framework for Automated Time-Series Phenotyping Using Massive Feature Extraction, Cell Systems, 5, 527-531.e3, https://doi.org/10.1016/j.cels.2017.10.001, 2017.

Fulcher, B. D., Little, M. A., and Jones, N. S.: Highly comparative time-series analysis: the empirical structure of time series and their methods, J. R. Soc. Interface., 10, 20130048, https://doi.org/10.1098/rsif.2013.0048, 2013.

Green, J. K., Seneviratne, S. I., Berg, A. M., Findell, K. L., Hagemann, S., Lawrence, D. M., and Gentine, P.: Large influence of soil moisture on long-term terrestrial carbon uptake, Nature, 565, 476–479, https://doi.org/10.1038/s41586-018-0848-x, 2019.

Guevara, M., Taufer, M., and Vargas, R.: Gap-free global annual soil moisture: 15 km grids for 1991–2018, Earth Syst. Sci. Data, 13, 1711–1735, https://doi.org/10.5194/essd-13-1711-2021, 2021.

Guillod, B. P., Orlowsky, B., Miralles, D. G., Teuling, A. J., and Seneviratne, S. I.: Reconciling spatial and temporal soil moisture effects on afternoon rainfall, Nat Commun, 6, 6443, https://doi.org/10.1038/ncomms7443, 2015.

Han, Q., Zeng, Y., Zhang, L., Wang, C., Prikaziuk, E., Niu, Z., and Su, B.: Global long term daily 1 km surface soil moisture dataset with physics informed machine learning, Sci Data, 10, 101, https://doi.org/10.1038/s41597-023-02011-7, 2023.

Homer, C., Dewitz, J., Jin, S., Xian, G., Costello, C., Danielson, P., Gass, L., Funk, M., Wickham, J., Stehman, S., Auch, R., and Riitters, K.: Conterminous United States land cover change patterns 2001–2016 from

the 2016 National Land Cover Database, ISPRS Journal of Photogrammetry and Remote Sensing, 162, 184–199, https://doi.org/10.1016/j.isprsjprs.2020.02.019, 2020.

Hu, F., Wei, Z., Zhang, W., Dorjee, D., and Meng, L.: A spatial downscaling method for SMAP soil moisture through visible and shortwave-infrared remote sensing data, Journal of Hydrology, 590, 125360, https://doi.org/10.1016/j.jhydrol.2020.125360, 2020.

Jagdhuber, T., Baur, M., Akbar, R., Das, N. N., Link, M., He, L., and Entekhabi, D.: Estimation of active-passive microwave covariation using SMAP and Sentinel-1 data, Remote Sensing of Environment, 225, 458–468, https://doi.org/10.1016/j.rse.2019.03.021, 2019.

Jiang, H., Shen, H., Li, X., Zeng, C., Liu, H., and Lei, F.: Extending the SMAP 9-km soil moisture product using a spatio-temporal fusion model, Remote Sensing of Environment, 231, 111224, https://doi.org/10.1016/j.rse.2019.111224, 2019.

Jiang, M., Shen, H., Li, J., and Zhang, L.: Generalized spatio-temporal-spectral integrated fusion for soil moisture downscaling, ISPRS Journal of Photogrammetry and Remote Sensing, 218, 70–86, https://doi.org/10.1016/j.isprsjprs.2024.10.012, 2024.

Jin, S., Dewitz, J., Danielson, P., Granneman, B., Costello, C., Smith, K., and Zhu, Z.: National Land Cover Database 2019: A New Strategy for Creating Clean Leaf-On and Leaf-Off Landsat Composite Images, J Remote Sens, 3, 0022, https://doi.org/10.34133/remotesensing.0022, 2023.

Jin, Y., Ge, Y., Wang, J., Chen, Y., Heuvelink, G. B. M., and Atkinson, P. M.: Downscaling AMSR-2 Soil Moisture Data With Geographically Weighted Area-to-Area Regression Kriging, IEEE Trans. Geosci. Remote Sensing, 56, 2362–2376, https://doi.org/10.1109/TGRS.2017.2778420, 2018.

Jin, Y., Ge, Y., Liu, Y., Chen, Y., Zhang, H., and Heuvelink, G. B. M.: A Machine Learning-Based Geostatistical Downscaling Method for Coarse-Resolution Soil Moisture Products, IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing, 14, 1025–1037, https://doi.org/10.1109/JSTARS.2020.3035386, 2021.

Karthikeyan, L. and Mishra, A. K.: Multi-layer high-resolution soil moisture estimation using machine learning over the United States, Remote Sensing of Environment, 266, 112706, https://doi.org/10.1016/j.rse.2021.112706, 2021.

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

https://doi.org/10.1016/j.rse.2017.10.026, 2018.

Kim, H., Lee, S., Cosh, M. H., Lakshmi, V., Kwon, Y., and McCarty, G. W.: Assessment and Combination of SMAP and Sentinel-1A/B-Derived Soil Moisture Estimates With Land Surface Model Outputs in the Mid-Atlantic Coastal Plain, USA, IEEE Trans. Geosci. Remote Sensing, 59, 991–1011, https://doi.org/10.1109/TGRS.2020.2991665, 2021.

Kim, J. and Hogue, T. S.: Improving Spatial Soil Moisture Representation Through Integration of AMSR-E and MODIS Products, IEEE Trans. Geosci. Remote Sensing, 50, 446–460, https://doi.org/10.1109/TGRS.2011.2161318, 2012.

Krueger, E. S., Ochsner, T. E., and Brorsen, B. W.: Soil Moisture Information Improves Drought Risk Protection Provided by the USDA Livestock Forage Disaster Program, Bulletin of the American Meteorological Society, 105, E1153–E1169, https://doi.org/10.1175/BAMS-D-23-0087.1, 2024.

Kumar, S. V., Dirmeyer, P. A., Peters-Lidard, C. D., Bindlish, R., and Bolten, J.: Information theoretic evaluation of satellite soil moisture retrievals, Remote Sensing of Environment, 204, 392–400, https://doi.org/10.1016/j.rse.2017.10.016, 2018.

Lakshmi, V. and Fang, B.: SMAP-Derived 1-km Downscaled Surface Soil Moisture Product, Version 1, https://doi.org/10.5067/U8QZ2AXE5V7B, 2023.

Li, Q., Shi, G., Shangguan, W., Nourani, V., Li, J., Li, L., Huang, F., Zhang, Y., Wang, C., Wang, D., Qiu, J., Lu, X., and Dai, Y.: A 1 km daily soil moisture dataset over China using in situ measurement and machine learning, Earth Syst. Sci. Data, 14, 5267–5286, https://doi.org/10.5194/essd-14-5267-2022, 2022.

Liu, L., Gudmundsson, L., Hauser, M., Qin, D., Li, S., and Seneviratne, S. I.: Soil moisture dominates dryness stress on ecosystem production globally, Nat Commun, 11, 4892, https://doi.org/10.1038/s41467-020-18631-1, 2020a.

Liu, P.-W., Bindlish, R., Fang, B., Lakshmi, V., O'Neill, P. E., Yang, Z., Cosh, M. H., Bongiovanni, T., Bosch, D. D., Collins, C. H., Starks, P. J., Prueger, J., Seyfried, M., and Livingston, S.: Assessing Disaggregated SMAP Soil Moisture Products in the United States, IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing, 14, 2577–2592, https://doi.org/10.1109/JSTARS.2021.3056001, 2021.

Liu, P.-W., Bindlish, R., O'Neill, P., Fang, B., Lakshmi, V., Yang, Z., Cosh, M. H., Bongiovanni, T., Collins, C. H., Starks, P. J., Prueger, J., Bosch, D. D., Seyfried, M., and Williams, M. R.: Thermal Hydraulic Disaggregation of SMAP Soil Moisture Over the Continental United States, IEEE J. Sel. Top. Appl.

Earth Observations Remote Sensing, 15, 4072–4092, https://doi.org/10.1109/JSTARS.2022.3165644, 2022.

800 Liu, Y., Jing, W., Wang, Q., and Xia, X.: Generating high-resolution daily soil moisture by using spatial downscaling techniques: a comparison of six machine learning algorithms, Advances in Water Resources, 141, 103601, https://doi.org/10.1016/j.advwatres.2020.103601, 2020b.

Long, D., Bai, L., Yan, L., Zhang, C., Yang, W., Lei, H., Quan, J., Meng, X., and Shi, C.: Generation of spatially complete and daily continuous surface soil moisture of high spatial resolution, Remote Sensing
805 of Environment, 233, 111364, https://doi.org/10.1016/j.rse.2019.111364, 2019.

Ma, H., Zeng, J., Chen, N., Zhang, X., Cosh, M. H., and Wang, W.: Satellite surface soil moisture from SMAP, SMOS, AMSR2 and ESA CCI: A comprehensive assessment using global ground-based observations, Remote Sensing of Environment, 231, 111215, https://doi.org/10.1016/j.rse.2019.111215, 2019.

810 Ma, S., Zhang, S., Wang, N., Huang, C., and Wang, X.: Prolonged duration and increased severity of agricultural droughts during 1978 to 2016 detected by ESA CCI SM in the humid Yunnan Province, Southwest China, CATENA, 198, 105036, https://doi.org/10.1016/j.catena.2020.105036, 2021.

Ma, X., Wang, Q., Tong, X., and Atkinson, P. M.: A deep learning model for incorporating temporal information in haze removal, Remote Sensing of Environment, 274, 113012,
815 https://doi.org/10.1016/j.rse.2022.113012, 2022a.

Ma, X., Wang, Q., and Tong, X.: A spectral grouping-based deep learning model for haze removal of hyperspectral images, ISPRS Journal of Photogrammetry and Remote Sensing, 188, 177–189, https://doi.org/10.1016/j.isprsjprs.2022.04.007, 2022b.

McPherson, R. A., Fiebrich, C. A., Crawford, K. C., Kilby, J. R., Grimsley, D. L., Martinez, J. E., Basara, J.
820 B., Illston, B. G., Morris, D. A., Kloesel, K. A., Melvin, A. D., Shrivastava, H., Wolfinbarger, J. M., Bostic, J. P., Demko, D. B., Elliott, R. L., Stadler, S. J., Carlson, J. D., and Sutherland, A. J.: Statewide Monitoring of the Mesoscale Environment: A Technical Update on the Oklahoma Mesonet, Journal of Atmospheric and Oceanic Technology, 24, 301–321, https://doi.org/10.1175/JTECH1976.1, 2007.

Meng, X., Zeng, J., Yang, Y., Zhao, W., Ma, H., Letu, H., Zhu, Q., Liu, Y., Wang, P., and Peng, J.: High-
825 resolution soil moisture mapping through passive microwave remote sensing downscaling, TIG, 100105, https://doi.org/10.59717/j.xinn-geo.2024.100105, 2024.

Merlin, O., Rudiger, C., Al Bitar, A., Richaume, P., Walker, J. P., and Kerr, Y. H.: Disaggregation of SMOS

Soil Moisture in Southeastern Australia, IEEE Trans. Geosci. Remote Sensing, 50, 1556–1571, https://doi.org/10.1109/TGRS.2011.2175000, 2012.

830   Min, X., Li, D., Shangguan, Y., Tian, S., and Shi, Z.: Characterizing the accuracy of satellite-based products to detect soil moisture at the global scale, Geoderma, 432, 116388, https://doi.org/10.1016/j.geoderma.2023.116388, 2023.

Molero, B., Merlin, O., Malbéteau, Y., Al Bitar, A., Cabot, F., Stefan, V., Kerr, Y., Bacon, S., Cosh, M. H., Bindlish, R., and Jackson, T. J.: SMOS disaggregated soil moisture product at 1 km resolution:

835   Processor overview and first validation results, Remote Sensing of Environment, 180, 361–376, https://doi.org/10.1016/j.rse.2016.02.045, 2016.

Montzka, C., Bogena, H., Zreda, M., Monerris, A., Morrison, R., Muddu, S., and Vereecken, H.: Validation of Spaceborne and Modelled Surface Soil Moisture Products with Cosmic-Ray Neutron Probes, Remote Sensing, 9, 103, https://doi.org/10.3390/rs9020103, 2017.

840   Moran, M. S., Clarke, T. R., Inoue, Y., and Vidal, A.: Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation index, Remote Sensing of Environment, 49, 246–263, https://doi.org/10.1016/0034-4257(94)90020-5, 1994.

Noel, M., Bathke, D., Fuchs, B., Gutzmer, D., Haigh, T., Hayes, M., Poděbradská, M., Shield, C., Smith, K., and Svoboda, M.: Linking Drought Impacts to Drought Severity at the State Level, Bulletin of the

845   American Meteorological Society, 101, E1312–E1321, https://doi.org/10.1175/BAMS-D-19-0067.1, 2020.

Peng, J., Loew, A., Merlin, O., and Verhoest, N. E. C.: A review of spatial downscaling of satellite remotely sensed soil moisture, Reviews of Geophysics, 55, 341–366, https://doi.org/10.1002/2016RG000543, 2017.

850   Rabus, B., Eineder, M., Roth, A., and Bamler, R.: The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar, ISPRS Journal of Photogrammetry and Remote Sensing, 57, 241–262, https://doi.org/10.1016/S0924-2716(02)00124-7, 2003.

Sabaghy, S., Walker, J. P., Renzullo, L. J., Akbar, R., Chan, S., Chaubell, J., Das, N., Dunbar, R. S., Entekhabi, D., Gevaert, A., Jackson, T. J., Loew, A., Merlin, O., Moghaddam, M., Peng, J., Peng, J., Piepmeier, J.,

855   Rüdiger, C., Stefan, V., Wu, X., Ye, N., and Yueh, S.: Comprehensive analysis of alternative downscaled soil moisture products, Remote Sensing of Environment, 239, 111586, https://doi.org/10.1016/j.rse.2019.111586, 2020.

Shephard, N. T., Joshi, O., Meek, C. R., and Will, R. E.: Long-term growth effects of simulated-drought, mid-rotation fertilization, and thinning on a loblolly pine plantation in southeastern Oklahoma, USA, Forest Ecology and Management, 494, 119323, https://doi.org/10.1016/j.foreco.2021.119323, 2021.

860

Song, P., Huang, J., and Mansaray, L. R.: An improved surface soil moisture downscaling approach over cloudy areas based on geographically weighted regression, Agricultural and Forest Meteorology, 275, 146–158, https://doi.org/10.1016/j.agrformet.2019.05.022, 2019.

Song, P., Zhang, Y., Guo, J., Shi, J., Zhao, T., and Tong, B.: A 1 km daily surface soil moisture dataset of enhanced coverage under all-weather conditions over China in 2003–2019, Earth Syst. Sci. Data, 14, 2613–2637, https://doi.org/10.5194/essd-14-2613-2022, 2022.

865

Souza, A. G. S. S., Ribeiro Neto, A., and Souza, L. L. D.: Soil moisture-based index for agricultural drought assessment: SMADI application in Pernambuco State-Brazil, Remote Sensing of Environment, 252, 112124, https://doi.org/10.1016/j.rse.2020.112124, 2021.

870 Tavakol, A., Rahmani, V., Quiring, S. M., and Kumar, S. V.: Evaluation analysis of NASA SMAP L3 and L4 and SPoRT-LIS soil moisture data in the United States, Remote Sensing of Environment, 229, 234–246, https://doi.org/10.1016/j.rse.2019.05.006, 2019.

U.S. Drought Monitor: https://droughtmonitor.unl.edu/DmData/TimeSeries.aspx, last access: 31 December 2024.

875 Vergopolan, N., Chaney, N. W., Pan, M., Sheffield, J., Beck, H. E., Ferguson, C. R., Torres-Rojas, L., Sadri, S., and Wood, E. F.: SMAP-HydroBlocks, a 30-m satellite-based soil moisture dataset for the conterminous US, Sci Data, 8, 264, https://doi.org/10.1038/s41597-021-01050-2, 2021.

Wang, L., Fang, S., Pei, Z., Wu, D., Zhu, Y., and Zhuo, W.: Developing machine learning models with multisource inputs for improved land surface soil moisture in China, Computers and Electronics in
880 Agriculture, 192, 106623, https://doi.org/10.1016/j.compag.2021.106623, 2022.

Wang, Q. and Atkinson, P. M.: Spatio-temporal fusion for daily Sentinel-2 images, Remote Sensing of Environment, 204, 31–42, https://doi.org/10.1016/j.rse.2017.10.046, 2018.

Wang, Q., Tang, Y., Tong, X., and Atkinson, P. M.: Virtual image pair-based spatio-temporal fusion, Remote Sensing of Environment, 249, 112009, https://doi.org/10.1016/j.rse.2020.112009, 2020.

885 Wang, X., Lü, H., Crow, W. T., Zhu, Y., Wang, Q., Su, J., Zheng, J., and Gou, Q.: Assessment of SMOS and SMAP soil moisture products against new estimates combining physical model, a statistical model, and in-situ observations: A case study over the Huai River Basin, China, Journal of Hydrology, 598, 126468,

https://doi.org/10.1016/j.jhydrol.2021.126468, 2021.

Wang, Z., Zhao, T., Shi, J., Wang, H., Ji, D., Yao, P., Zheng, J., Zhao, X., and Xu, X.: 1-km soil moisture
retrieval using multi-temporal dual-channel SAR data from Sentinel-1 A/B satellites in a semi-arid
watershed, Remote Sensing of Environment, 284, 113334, https://doi.org/10.1016/j.rse.2022.113334,
2023.

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 Sensing of Environment,
225, 30–44, https://doi.org/10.1016/j.rse.2019.02.022, 2019.

Wen, F., Zhao, W., Wang, Q., and Sanchez, N.: A Value-Consistent Method for Downscaling SMAP Passive
Soil Moisture With MODIS Products Using Self-Adaptive Window, IEEE Trans. Geosci. Remote
Sensing, 58, 913–924, https://doi.org/10.1109/TGRS.2019.2941696, 2020.

Williams, A. P., Cook, B. I., and Smerdon, J. E.: Rapid intensification of the emerging southwestern North
American megadrought in 2020–2021, Nat. Clim. Chang., 12, 232–234,
https://doi.org/10.1038/s41558-022-01290-z, 2022.

Xu, M., Yao, N., Yang, H., Xu, J., Hu, A., Gustavo Goncalves De Goncalves, L., and Liu, G.: Downscaling
SMAP soil moisture using a wide & deep learning method over the Continental United States, Journal
of Hydrology, 609, 127784, https://doi.org/10.1016/j.jhydrol.2022.127784, 2022.

Xu, Z., Sun, H., Gao, J., Wang, Y., Wu, D., Zhang, T., and Xu, H.: PhySoilNet: A deep learning downscaling
model for microwave satellite soil moisture with physical rule constraint, International Journal of
Applied Earth Observation and Geoinformation, 135, 104290,
https://doi.org/10.1016/j.jag.2024.104290, 2024.

Yang, H., Yang, J., and Ochsner, T. E.: A 3-hour, 1-km surface soil moisture dataset in Continental United
States, Figshare [data set], https://doi.org/10.6084/m9.figshare.28188011, 2025.

Yang, H. and Wang, Q.: Reconstruction of a spatially seamless, daily SMAP (SSD_SMAP) surface soil
moisture dataset from 2015 to 2021, Journal of Hydrology, 621, 129579,
https://doi.org/10.1016/j.jhydrol.2023.129579, 2023.

Yang, H., Wang, Q., Zhao, W., Tong, X., and Atkinson, P. M.: Reconstruction of a Global 9 km, 8-Day SMAP
Surface Soil Moisture Dataset during 2015–2020 by Spatiotemporal Fusion, J Remote Sens, 2022,
2022/9871246, https://doi.org/10.34133/2022/9871246, 2022.

Yang, H., Wang, Q., Ma, X., Liu, W., and Liu, H.: Digital Soil Mapping Based on Fine Temporal Resolution

Landsat Data Produced by Spatiotemporal Fusion, IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing, 16, 3905–3914, https://doi.org/10.1109/JSTARS.2023.3267102, 2023.

920 Yang, H., Wang, Q., and Liu, W.: A stepwise method for downscaling SMAP soil moisture dataset in the CONUS during 2015–2019, International Journal of Applied Earth Observation and Geoinformation, 130, 103912, https://doi.org/10.1016/j.jag.2024.103912, 2024.

Zhang, C., Yang, Z., Zhao, H., Sun, Z., Di, L., Bindlish, R., Liu, P.-W., Colliander, A., Mueller, R., Crow, W., Reichle, R. H., Bolten, J., and Yueh, S. H.: Crop-CASMA: A web geoprocessing and map service

925 based architecture and implementation for serving soil moisture and crop vegetation condition data over U.S. Cropland, International Journal of Applied Earth Observation and Geoinformation, 112, 102902, https://doi.org/10.1016/j.jag.2022.102902, 2022.

Zhao, H., Li, J., Yuan, Q., Lin, L., Yue, L., and Xu, H.: Downscaling of soil moisture products using deep learning: Comparison and analysis on Tibetan Plateau, Journal of Hydrology, 607, 127570,

930 https://doi.org/10.1016/j.jhydrol.2022.127570, 2022.

Zhao, W., Sánchez, N., Lu, H., and Li, A.: A spatial downscaling approach for the SMAP passive surface soil moisture product using random forest regression, Journal of Hydrology, 563, 1009–1024, https://doi.org/10.1016/j.jhydrol.2018.06.081, 2018.

Zhao, W., Wen, F., Wang, Q., Sanchez, N., and Piles, M.: Seamless downscaling of the ESA CCI soil moisture

935 data at the daily scale with MODIS land products, Journal of Hydrology, 603, 126930, https://doi.org/10.1016/j.jhydrol.2021.126930, 2021.

Zheng, C., Jia, L., and Zhao, T.: A 21-year dataset (2000–2020) of gap-free global daily surface soil moisture at 1-km grid resolution, Sci Data, 10, 139, https://doi.org/10.1038/s41597-023-01991-w, 2023.

Zhu, L., Wang, H., Zhao, T., Li, W., Li, Y., Tong, C., Deng, X., Yue, H., and Wang, K.: Disaggregation of

940 remote sensing and model-based data for 1 km daily seamless soil moisture, International Journal of Applied Earth Observation and Geoinformation, 125, 103572, https://doi.org/10.1016/j.jag.2023.103572, 2023.