Earth and Space Science

RESEARCH ARTICLE
10.1029/2020EA001265

Key Points:
- A machine-learning algorithm is proposed to increase the resolution of soil moisture products of remote sensing on the Tibetan Plateau.
- Topographical features are applied to estimate soil moisture in complex terrains, so that the accuracy of estimation could be improved.
- Random forest algorithm is applied to combine various features, and the results are expected to well fit the real situation.

Correspondence to:
Q. Chen,
cqq@cuit.edu.cn

Citation:
Chen, Q., Miao, F., Wang, H., Xu, Z.-X., Tang, Z., Yang, L., & Qi, S. (2020). Downscaling of satellite remote sensing soil moisture products over the Tibetan plateau based on the random forest algorithm: Preliminary Results. Earth and Space Science, 6, e2020EA001265. https://doi.org/10.1029/2020EA001265

Received 11 MAY 2020
Accepted 1 JUN 2020
Accepted article online 8 JUN 2020

Abstract

Soil moisture (SM) is an important index of soil drought, and it directly controls the energy balance and water cycle of the land surface. As an indicator and amplifier of global warming, the Tibetan Plateau (TP) is becoming warmer and wetter. Because of its particular geographical environment, large-scale measurements of SM on the TP can only be achieved by satellite remote sensing. The resolution of current SM product of the Soil Moisture Active Passive (SMAP) satellite is 36 km, which is insufficient for many practical applications. In this study, the Moderate Resolution Imaging Spectroradiometer (MODIS) and Digital Elevation Model (DEM) are applied to increase the resolution of SM down to 1 km using the Random Forest (RF) algorithm. The preliminary results of the proposed algorithm are evaluated by station observations and other reanalysis products. The downscaled results are more consistent with the in situ observations, the Land Data Assimilation System (CLDAS) from China Meteorological Administration (CMA), and the Global Land Data Assimilation System (GLDAS) from National Aeronautics and Space Administration (NASA) than the original SMAP product. The downsampling algorithm is most effective for grasslands. It is demonstrated that high-resolution SM products can be generated by fusing various features using machine-learning algorithms.

1. Introduction

Soil moisture (SM) plays a vital role in the energy balance and water cycle over the land surface. It affects the energy exchange between the atmosphere and the land surface by changing the surface albedo, soil heat capacity, surface evaporation, and the growth of vegetation, thus affecting climate change in local areas (Fischer et al., 2007). Both climate change and anthropogenic activities have an impact on plant dynamics; Tibetan Plateau (TP), which has been defined as the center of the “Third Pole,” is especially sensitive. The relationship between vegetation and climate is associated with surface fluxes, which is determined by the state of SM (Cai et al., 2014, 2015). Global warming has caused the TP to warm faster than the surrounding areas, at a rate of 1.5 times that of global warming, which is causing environmental changes such as the degradation of permafrost, glacial retreat, and melting snow (Yao et al., 2018; You et al., 2016, 2017). SM responds rapidly to changes in hydrology and water resources and interacts with the upper atmosphere through surface energy exchange and the water balance, leading to regional climatic feedback (Kuang & Jiao, 2016). Therefore, SM plays an important role in improving the prediction of climate change on a regional and global basis, crop yield estimation, environmental disaster monitoring, and natural and ecological issues regarding the environment. SM has therefore received a significant amount of attention in the forecasting of weather and floods, drought assessment, ecological monitoring, and the management of water resources (Zhu et al., 2010; Zuo et al., 2016). Currently, the accurate measurement of SM at a regional or global scale is relatively difficult. Only two methods are available for the measurement of SM: ground in situ measurements and the inversion of satellite remote sensing measurements. Techniques used in ground measurement include gravity, time domain reflectometry, capacitive sensors, neutron detectors, resistivity measurements, thermal pulse sensors, and fiber optic sensors, among others. These technologies are...
relatively mature, easy to install, and are capable of measuring SM at different depths with high temporal resolution (Peng et al., 2017). However, due to the complex interactions between various environmental variables such as the texture and structure of the soil, topographical features, land cover types (LCTs), and meteorological forcing, surface SM has great spatial variability at different scales (Vereecken et al., 2014), meaning that in situ observations may not be representative of adjacent areas (Collow et al., 2012). In areas such as the TP, with its high altitude, large area, and complex terrains, it is difficult to carry out the dynamic monitoring of SM over a wide area with a few ground stations. The economic, efficient, and synchronous feature of remote sensing technology provides the possibility for real-time dynamic monitoring of SM over large areas. Microwave remote sensing uses long wavelengths that can penetrate the atmosphere and clouds to achieve observations throughout the day in all weathers and is sensitive to SM information. Examples include the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) (Njoku et al., 2003), the Advanced Scatterometer (ASCAT) (Bartalis et al., 2007), the Soil Moisture and Ocean Salinity (SMOS) (Kerr et al., 2010), the Soil Moisture Active Passive (SMAP) (Entekhabi et al., 2010), and the European Space Agency’s Sentinel-1 satellite (Wagner et al., 2009). However, the above instruments only have low spatial resolutions of 25–50 km. Although this resolution may satisfy many global applications, many regional applications of agriculture and hydrology such as irrigation management and flood forecasting require spatial resolutions of a few kilometers or even tens of meters (Beyene et al., 2018; Ramírez-Hernández et al., 2017).

In order to overcome the limitation due to the coarse resolution, downscaling techniques are applied to improve the resolution of remote sensing SM products. Three groups of methods are widely used to achieve this goal: (1) The first group combines passive microwave and high-resolution active microwave data with visible light and thermal infrared data. Das et al. (2014) used radar backscatter data to downscale the bright temperature data; Montzka et al. (2016) proposed a simplified wavelet-based method in order to fuse the active and passive microwave data. SMAP solves the problem concerning the inconsistency in the observation times of the active and passive sensors. Many scholars (Leroux et al., 2016) have studied fusion methods to combine the active and passive data produced via SMAP. However, the active radar on SMAP stopped working after only 6 months of operation. Optical and thermal remote sensing can obtain land surface parameters with a higher spatial resolution than microwave remote sensing (Piles et al., 2016). Chen et al. (2019) and Ojha et al. (2019) applied the widely used surface temperature/vegetation index triangular feature space method. In addition, Srivastava et al. (2013) reduced the scale of SM monitoring with the use of artificial intelligence technologies such as artificial neural networks, support vector machine, and correlation vector machine. (2) The second group is based on geographic information and uses topography as the auxiliary information for downscaling. Cowley et al. (2017) applied the Equilibrium Moisture from Topography, Vegetation, and Soil (EMT + VS) model to develop a simple method using time-averaged precipitation and potential evapotranspiration (PET) and incorporated it into the EMT + VS model. Hoehn et al. (2017) generalized the EMT + VS model to accept multiple coarse grid cells. (3) The third group of methods is based on statistical models and surface physical models. Mascaro et al. (2011) downscaled the SM field by its fractal dimension characteristics. Shin and Mohanty (2013) obtained a fine-scale SM by optimizing the hydrological model. Kwon et al. (2017) proposed a multivariate stochastic SM estimation method based on the Gaussian-Mixture Nonstationary Hidden Markov Model (GM-NHMM). Verhoest et al. (2015) used statistics to ascertain the relationship between the fine-scale SM and the overlapping coarse-scale observations acquired by radiometers. Sahoo et al. (2013) used three-dimensional Ensemble Kalman Filter (3-D EnKF) and one-dimensional EnKF (1-D EnKF) methods to downscale AMSR-E SM from 25 to 1 km.

In summary, there are still problems with regards to the downscaling of remote sensing SM: (1) Downscaling is less effective in complex terrains, and further methods need to be explored. (2) There are many factors that influence SM, so methods for downscaling that use only a small number of factors may not be optimized. (3) The relationship between SM and the physical variables of the land surface is complex and nonlinear and cannot be accurately represented by linear models. In this study, various factors are combined in a machine-learning algorithm to downscale the SM passive microwave products of SMAP. The downscaled results over the TP are evaluated and compared with in situ observations, and SM products from the China Meteorological Administration Land Data Assimilation System (CLDAS) and the Global Land Data Assimilation System (GLDAS). The innovations in this paper are as follows: (1) to try and explore the
downscaling of remote sensing SM products in complex terrain of the TP. (2) A variety of land surface parameters is included, such as the Normalized Difference Vegetation Index (NDVI), the Enhanced Vegetation Index (EVI), the Land Surface Temperature (LST), and the LCT. Because of the characteristics of the complex terrain, geographic information such as longitude, latitude, elevation, slope, and the aspect is also integrated as inputs. (3) A machine-learning algorithm is applied in this study. This paper is organized as follows. Section 2 describes the study area and data used in this study, the methods are given in section 3, and the results and discussions are presented in section 4. Finally, the conclusions and discussions are given in section 5.

2. Study Area and Data

2.1. Study Area

The area studied in this investigation is TP (25°57′ to 39°55′N, 73°27′ to 104°38′E) in China. The TP is the highest and largest of the plateaus around the world, with an average elevation of more than 4,000 m and an area of approximately 2.5 × 10^6 km² (Zhang et al., 2002). The plateau is of higher elevation in the west and the north, as illustrated by the low-altitude area in the southeast and the high-altitude area in the northwest seen in (a) the Digital Elevation Model (DEM) and (b) the LCT of Figure 1. The slope of terrain is large, and the spatial heterogeneity is high. The typical types of plateau geomorphology (mountains, hills, alluvial fans, and intermountain basins) and land surface types (cold grasslands, alpine meadows, alpine barrens, lakes, and glaciers) are all well represented on the TP (Cao et al., 2017). Compared with other regions at the same latitude, the spatial distribution of vegetation varies greatly. From the southeast to the northwest, horizontal and vertical changes in the natural conditions lead to different vertical bands with distinct three-dimensional variations. The vegetation ranges from forest to shrub meadows, alpine meadows, alpine grasslands, and eventually to plateau barrens (Zheng, 1996). The TP has the widest and thickest distribution of frozen soil observed among the middle and low latitudes. The SM gradually increases from the northwest to the southeast (Chen et al., 2017; Ullah et al., 2018).
2.2. Data

2.2.1. MODIS Data
The Moderate Resolution Imaging Spectroradiometer (MODIS) is an important sensor on the Earth Observation System (EOS) Terra and Aqua satellites, which is widely used to monitor a variety of environments, including the land, sea, and atmosphere at the surface. The MODIS data used in this study are from the National Aeronautics and Space Administration (NASA) Land Processes Distributed Active Archive Center (LP DAAC), including the LST products MYD11A1 (https://e4ftl01.cr.usgs.gov/MOLA/MYD11A1.006/), the Vegetation Index (VI) products MOD13A2 (https://e4ftl01.cr.usgs.gov/MOLT/MOD13A2.006/), and the LCT products MCD12Q1 (https://e4ftl01.cr.usgs.gov/MOTA/MCD12Q1.006/). The tiles of the data products used in the study area are h23v05, h24v05, h25v05, h25v06, h26v05, and h26v06. (1) The LST product MYD11A1 is a daily L3 product with a 1 km resolution. (2) The VI product from the MOD13A2 data is a 16‐day synthetic 1 km resolution L3 product. This study used the NDVI and EVI products from these data. (3) The LCT annual product MCD12Q1 contains five land cover classification programs with a spatial resolution of 500 m. The International Geosphere‐Biosphere Programme (IGBP) classification was used in this study because it is suitable for use with global vegetation. This program consists of 17 types, including 11 natural vegetation types, 3 land development and inlaid land types, and 3 non‐grass land types (Justice et al., 1998).

2.2.2. DEM Data
DEM mainly describes information concerning the digital elevation of the ground. The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) GDEM V2 30 m products from the Geospatial Data Cloud website of the Computer Network Information Center, Chinese Academy of Sciences (http://www.gscloud.cn), were used in this study, which are in the TIFF format with spatial resolutions of up to 20 m vertically and 30 m horizontally. The ASTER GDEM product is currently the only high‐resolution elevation image data that cover the global. It was jointly developed by the Ministry of Economy, Trade, and Industry (METI) of Japan and NASA of USA and was released to the public for free. ASTER GDEM V2 applies an advanced algorithm to improve the spatial resolution accuracy and elevation accuracy of the data from ASTER GDEM V1 (Rexer & Hirt, 2014). It was officially released on 6 January 2015.

2.2.3. SMAP SM Data
The SMAP satellite designed by NASA was launched on 31 January 2015. It is the world’s first surface (approximately 5 cm underground) SM monitoring satellite that combines active and passive microwaves (L band). The main purpose of SMAP is to study the interactions between the water, energy, and carbon cycles that occur locally within a particular region; to estimate the global surface water heat flux; and to enhance the capability of drought monitoring, flood prediction, and meteorological forecasting. However, its active radar sensor was damaged on 7 July 2015, and only the passive SM data are available at present. The passive product is of 36 km spatial resolution and includes daily data at 6:00 am (satellite moving from north to south) and 6:00 pm (satellite moving from south to north), covering the global every 50 hr (Entekhabi et al., 2010). The SMAP data used in this study are from the National Snow and Ice Data Center (https://nsidc.org/data/SPL3SMP/versions/5) in an HDF5 format. Each file contains the AM and PM SM data within a single day. Five characteristic variables including longitude, latitude, AM/PM, and current/previous day can be extracted from the SMAP data file. A detailed explanation of these variables is described in section 3.2.

2.2.4. In Situ SM Data
The in situ data used in this study are soil volumetric water content at a depth of 0–10 cm measured by the automatic SM observation stations of China Meteorological Administration (CMA). These stations upload hourly data on a daily basis, mainly monitoring four SM elements at eight levels: 0–10, 10–20, 20–30, 30–40, 40–50, 50–60, 70–80, and 90–100 cm. The four elements measured include soil volumetric water content, soil relative humidity, soil mass content, and soil water availability (Cheng & Zhang, 2017). Zhang et al. (2014) compared SM data from the automatic stations and manual measurements and concluded that the data observed by the automatic stations can replace those from manual measurements.

2.2.5. CLDAS SM Data
CLDAS V2.0 is an equivalent latitude and longitude grid fusion analysis product that covers the Asian region (0–65°N, 60–160°E) with a spatial resolution of 0.0625° × 0.0625° and a temporal resolution of 1 hr, including atmospheric drive field products, surface temperature analysis products, SM products, soil temperature analysis products, and products related to the analysis of the relative humidity of the soil. This data set
utilizes observations from various ground and satellite sources and is developed using techniques such as the Space-Time Multiscale Analysis System (STMAS), Optimal Interpolation (OI), Cumulative Distribution Function (CDF) matching, physical inversion, and terrain correction. Its quality is better over China than other similar products with a higher spatial and temporal resolution (Cui et al., 2018; Qin et al., 2017). This paper uses the 0–5 cm (below the surface) SM product from China National Meteorological Information Center (http://data.cma.cn/data/detail/dataCode/NAFP_CLDAS2.0_RT.html).

### 2.2.6. GLDAS SM Data

GLDAS is a high-resolution global land surface simulation system that was jointly developed by NASA’s Goddard Space Flight Center (GSFC) and National Center for Environmental Prediction (NCEP). It provides global land surface data from 1979 to the present with a spatial resolution of 0.25° × 0.25° and 1° × 1°, on both a 3 hourly and monthly basis. GLDAS uses meteorological variables, radiation, and other driving data based on multisource observations, reanalysis data, and atmospheric assimilation products (Bao et al., 2017). This paper uses the GLDAS Noah Land Surface Model L4 3 hourly 0.25° × 0.25° V2.1 (GLDAS_NOAH025_3H.2.1) 0–10 cm layer data downloaded from the Goddard Earth Sciences Data and Information Services Center (GES DISC) (https://hydro1.gesdisc.eosdis.nasa.gov/data/GLDAS/GLDAS_NOAH025_3H.2.1), in an nc4 format. The SM products of SMAP and CLDAS provide water content in a volumetric form, while the unit of the GLDAS SM product is kg/m². The different units are converted using the following equation (Cui et al., 2018). More detailed information about the data used in this study is given in Table 1.

\[
\frac{\text{kg}}{m^3} \times \frac{1,000 \text{ mm}}{1 m} \times \frac{1}{\text{soil layer thickness (mm)}} = \text{volumetric water content}
\]  

(1)

### 3. Methods

#### 3.1. The Random Forest Algorithm

Random forest (RF), which can be used for both classification and regression, is an integrated learning algorithm based on the use of a decision tree. It was proposed by Breiman (2001) and improved by Cutler et al. (2011). It is adopted in this study because of its superiority in building the SM relationship model with surface variables (Im et al., 2016). As the predicted data are continuous, the regression model is applied. A regression RF model can be considered as the dependent variable that is affected by multiple independent variables, as shown in Equation 2:

\[
\{h(x, \bar{\theta}_t), t = 1, 2, 3...T\},
\]  

(2)

where \(\bar{\theta}_t\) is the generated independent and identically distributed regression tree, \(x\) is the independent variable, \(T\) is the number of regression trees generated, and \(h(x, \bar{\theta}_t)\) is the predicted value of this regression tree. The regression tree that grows each time is different due to the randomness of the independent variables. Using this method, the RF constructs a large number of regression trees, of which each tree has a predicted value. The final prediction of the RF will finally average all these values as shown in Equation 3:
\[ h(x) = \frac{1}{T} \sum_{t=1}^{T} \{ h(x, \theta_t) \}. \]  

(3)

To overcome the problem of overfitting and improve the accuracy of the model, the RF algorithm uses the ideas of bagging and random subspace. In the process of constructing a regression tree, the model uses the bootstrap method for resampling. Bootstrap is a repeat sampling method, where the original sample is returned statistically. By repeatedly sampling the original sample multiple times, the number sampled is the same as the original number of samples. The original data are returned after each sampling, and a new batch of samples can be obtained. Each new sample is then extracted to construct the regression tree. Using this method, a narrower interval of confidence can be obtained, and the returned sampling guarantees the independence between each tree. If \( A \) is the original sample and \( N \) is the number of samples in \( A \), then for each sample in \( A \), the probability of not being sampled is

\[ 1 - \frac{1}{N} \]

(4)

When \( N \to \infty \),

\[ \left(1 - \frac{1}{N}\right)^N \approx \frac{1}{e} \approx 0.368. \]

(5)

Therefore, approximately 36.8% of the data are not sampled every time. These data are known as the out-of-bag estimates, which can be used to assess the accuracy of the model. When there are \( M \) features in the samples, each child node of the tree randomly takes \( m \) features from the \( M \) features in the process of regression tree growth, forming a split feature set. The value of \( m \) remains unchanged throughout the forest growth process, and the growth of the node is performed by the best splitting method among the \( m \) features. This method guarantees the diversity and independence of each tree. After completing the above process, each tree is allowed to grow freely, and the RF is terminated by setting the number of trees permitted. After completing the growth of the forest, the model training process ends. The feature value \( X = x \) is then input into the model, and the predicted value \( y \) is calculated via the weighted averaging of the predicted value from each tree.

3.2. The Research Program

The raw data are first downloaded and preprocessed. The features of NDVI, EVI, LST, LCT, DEM, slope, and aspect in MODIS are then matched with SMAP in time and space. The 1 km resolution SM data are obtained by downscaling the matched data using RF. Finally, the results are evaluated and compared with in situ observations, CLDAS and GLDAS. The overall flowchart of this study is shown in Figure 2.

1. Data downloading and preprocessing: The ArcGIS software is used to splice files of the study area stored in multiple tiles into the same raster file. The high-resolution DEM and LCT are resampled to 1 km using the nearest neighbor method. In order to ensure spatial alignment, the projection coordinate system of all the data is unified with WGS84 EASE-Grid 2.0 of the SMAP data. The slope and aspect are generated from the DEM data and added to the features of the RF algorithm. All data are clipped using the TP vector boundary to obtain data from inside the study area.

2. Time and space matching: The observation time of SMAP is used as the reference, and other data are selected with the smallest time difference to SMAP. A similar process is applied to space matching where the nearest points of other data are selected according to the latitude and longitude of SMAP.

3. Model building and downscaling: Since the resolution of the SMAP data is 36 km and that of the DEM and MODIS data is 1 km, 36*36 features are matched for the same tag, and it is necessary to select those that are optimal. To solve this problem, the shortest distance method is applied, and the distance stored as a new feature. Each SMAP file contains both data of morning and afternoon, so the AM/PM feature is added. Since the SMAP data are not able to cover the entire study area in a single day, the data from the previous day are also used. A new categorical feature “Current/previous day” is added to distinguish the SMAP data between the current day and the previous day. Continuous data (NDVI, EVI, LST, DEM, Distance, Slope, and Aspect) and categorical data (LCT, AM/PM, and Current/previous day) are thereby
obtained. Categorical data, which are incalculable, need to be independently encoded in order to avoid the computer mistakenly thinking of it as continuous data. Ice, snow, frozen soil, water bodies, and so forth are considered as incapable of inverting SM by satellites. Grid points included in these areas are therefore assigned invalid values and are eliminated from the training examples. Such areas occupy only a very small proportion of the total, so this does not affect the results. After the above processing, the training samples are all ready for RF modeling. In order to ensure the randomness of the sample used for training, the samples are randomly sampled within the data set, with a ratio of the number of training set samples to the number of test set samples of 7:3. After the data partition, the training data are input to the RF regression module for machine learning. An optimal model is obtained when the parameters are adjusted such that the test score and the out-of-bag test score reach a high level at the same time and the difference between the two is small. In this study, the model scored 0.834 in the test set and the out-of-bag data test scored 0.824. Applying this model, all of the data are preprocessed and input into the model for the 1 km scale data prediction. The 1 km downscaled SM is hereafter referred to as D1SM.

4. Evaluation and comparison: The SM data from CLDAS and GLDAS over the same study area are matched with the D1SM in time and space. They are evaluated with in situ observations. The results of evaluation and comparison are detailed in section 4.

Downscaling of SM is challenging over complex terrain as TP since SM has large variability both in time and space. This preliminary study mainly focuses on the spatial variation of SM without considering its temporal evolution. First, it is because the accurate downsampling of SM for a spatial full coverage is the basis for a comprehensive evaluation of its temporal

Table 2

| Variables     | Time                  |
|---------------|-----------------------|
| SMAP SM       | 6:00 and 18:00, 10/18–19/2017 |
| NDVI EVI      | 10/16/2017            |
| LST           | 10/19/2017            |
| LCT           | 2017                  |
| DEM           | —                     |
| CLDAS SM      | 18:00, 10/19/2017     |
| GLDAS SM      | —                     |
| In situ SM    | —                     |
evolution (Jia et al., 2017; Zhang & Huang, 2013). Second, the spatial characteristics of SM are more important than its temporal evolution in the studied area. Due to the use of the MODIS LST daily data, the current algorithm is not applicable over areas with cloud coverage. The SMAP data cannot cover the entire study area in a single day; for full spatial coverage, the data from 2 days (18 and 29 October 2017) with minimal cloud cover were used to develop the downscaling model. Since the main focus of the current study is the feasibility and performance of the algorithm, the use of data in cloudless days is acceptable. Better solutions to circumvent the problem of cloud contamination will be the topics of subsequent studies. The time for other data is chosen using the nearest neighbor method due to the different time resolutions. After downscaling, the data at 18:00, 19 October 2017 are used for evaluation and comparison. However, SMAP uses data of 6:00 and 18:00, 18–19 October 2017 in order to cover the study area. In summary, the time of data used in this study is listed in Table 2.

4. Results and Discussions

4.1. Correlation Analysis of Model Variables

The correlation coefficients (CCs) between the independent and dependent variables in the built model are given in Table 3, which passed the 95% confidence level except for the aspect feature. It can be seen from Table 3 that the CCs range from −0.03 and 0.6 with an absolute average value of 0.3, indicating that SM is not determined by any individual factors but by a combination of multiple factors. Therefore, if only a small number of independent variables are used for downscaling, other influencing factors will be ignored. It is therefore reasonable to use more characteristic variables and to automatically calculate the impact weights via machine learning. It can be seen from Table 3 that the highest CCs are NDVI, EVI, and Longitude with values of 0.6, 0.58, and 0.56, respectively. This indicates that vegetation has a significant influence on SM. The larger the VI, the more moist the soil, because the vegetation retains the water in the soil. The positive of Longitude is attributed to the drier air toward the west inland and the vegetation changes from forest and meadow to grassland and desert.

| Features          | CC   |
|-------------------|------|
| NDVI              | 0.60 |
| EVI               | 0.58 |
| Longitude         | 0.56 |
| LCT               | −0.43|
| DEM               | −0.38|
| Latitude          | −0.26|
| LST               | −0.23|
| Current/previous day | −0.22|
| Slope             | 0.16 |
| Distance          | 0.12 |
| AM/PM             | −0.06|
| Aspect            | −0.03|

Figure 3. Locations and DEMs of the stations in Sichuan Province over the TP.
4.2. Analysis of In Situ Evaluation

After preprocessing, the numbers of samples of the D1SM, SMAP, CLDAS, and GLDAS are 2,094,860, 3,438, 55,002, and 3,454, respectively. The D1SM, SMAP, CLDAS, and GLDAS data are evaluated using the SM in situ observations from the meteorological stations located in Sichuan Province. These stations mainly distribute in the southeast of TP (the red rectangle in Figure 1a) and are labeled in Figure 3. The stations are relatively dense over this area of the TP. The DEM values of the stations are listed in Table 4 in descending order.

The SM values of D1SM, SMAP, CLDAS, and GLDAS are compared to the ground observations in Figure 4a with their differences relative to the ground observations shown in Figure 4b. Despite the limited samples, it is evident that the D1SM best matches the ground observations in most

| Table 4 | DEM Values of the Stations |
|---------|-----------------------------|
| Station | DEM (m)                     |
| Shigu   | 4,200.0                     |
| Daocheng| 3,727.7                     |
| Ganzi   | 3,393.5                     |
| Rangtang| 3,292.9                     |
| Jiulong | 2,925.0                     |
| Maerkang| 2,664.4                     |
| Heishui | 2,400.1                     |
| Muli    | 2,379.0                     |
| Jinchuan| 2,168.9                     |
| Maoxian | 1,590.1                     |

Figure 4. (a) The SM values of the ground stations, D1SM, SMAP, CLDAS, and GLDAS in the same locations. (b) The SM differences between the D1SM, SMAP, CLDAS, GLDAS, and the ground observations of some stations over the TP.
4.3. Overall Comparison

4.3.1. Statistical Characteristics

To verify the effect of the algorithm, the changes before and after the downscaling are analyzed through comparing with CLDAS and GLDAS. Figure 5 shows the boxplots for the four SM data sets. The mean value of SMAP is smaller (drier) than those of CLDAS and GLDAS. The dry tendency of SMAP is partly corrected by the D1SM. The GLDAS data are generally higher because its measured depth is 0–10 cm while the others are 0–5 cm. The closer the soil layer to the ground surface, the stronger the water dispersion and the downward infiltration. Therefore, the SM values near the ground are lower than those far from the ground (Minet et al., 2011).

To further describe the statistical characteristics of the data, some indicators for the above four data sets were calculated. Figure 6 shows the probability density function curves with detailed statistics (mean, standard deviation, kurtosis, and skewness) listed in Table 5. As Figure 6 and Table 5 show, the modes of the D1SM, SMAP, and CLDAS are near 0.05–0.1 m³/m³. The larger mode (0.16 m³/m³) of GLDAS are likely resulted from its different measurement depth. Among these data sets, the probability density function curve of D1SM has three peaks similar to CLDAS and GLDAS while the original SMAP has only two peaks. The standard deviations (0.07, 0.08, 0.07, and 0.08 m³/m³) of the four data sets are small and similar, indicating that the data distributions are concentrated. The difference between the means of the D1SM and the CLDAS (0.12 and 0.16 m³/m³, respectively) is small, indicating that the two are generally consistent. The kurtoses are all greater than 3, indicating that their distributions are steeper than the normal distribution. The kurtosis of the SMAP is significantly different from others, and the skewness is all positive, and the D1SM data have a smaller difference than the SMAP data. In summary, the D1SM is more consistent with the CLDAS and GLDAS than SMAP.

4.3.2. Spatial Distribution

Figure 7 shows the SM spatial distribution of the D1SM, SMAP, CLDAS, and GLDAS. The four data sets show the similar regional characters. The SM values in the eastern and southeastern parts of TP are higher than those of the northwest inland areas, which is likely related to the precipitation distribution and the land surface types. Because there is less precipitation in the northwest and more in the southeast, the vegetation shows a transition from the subtropical evergreen broad-leaved forest in the southeast to the high-cold desert in the northwest, and the SM undergoes a corresponding change (Yang et al., 2017). There is less rainfall in the northwestern part of TP, where most of the land surface is barren with sparse vegetation. The capacity to retain water is relatively poor, resulting in the low SM values and lower SM change seen in this area. In the central regions, the vegetation is relatively good, and the SM values lie in the medium range. In the eastern and southern regions, and especially the southeast, the dense vegetation has good water-fixing capacity, and with the abundance of rivers, the sufficient water and heat conditions, and the relatively low elevation, the SM is maintained at a high level. The southern mountainous and evergreen broad-leaved forest belt is a low-altitude area downstream of the Yarlung Zangbo River, where the SM is easily maintained. The Hengduan Mountains lie in the southeast, through which the Nujiang River, the Lancang River, and the Jinsha River flow, and to the east is the Yellow River Basin. The SM is therefore well maintained in these areas. The soil around lakes and at the confluence of rivers is obvious high value, and the high value of Qinghai Lake in the northeast is particularly notable in D1SM.
The general consistency of the D1SM with the spatial distribution of the other three data sets demonstrates the accuracy of the D1SM and the feasibility of this method. Compared with the original SMAP data, the D1SM is more refined, and the local differences are more obvious, so the fine-scale SM values can be obtained. The D1SM at the southwestern edge of TP are in agreement with those from SMAP but are slightly different from those from CLDAS and GLDAS. GLDAS has higher values due to its deeper measurement depth.

### 4.3.3. Land Cover Type Analysis

To analyze the overall data distributions of SM in different land cover types (LCTs) and examine the effect of the downscaling method in detail, the four most common types of land cover are selected over the TP according to their frequencies in the histogram. These include Class 10 Grasslands (52.53%), Class 16 barren lands (38.17%), Class 8 Woody Savannas (2.12%), and Class 1 Evergreen Needleleaf Forests (1.47%), which can be seen in Figure 1b. We analyzed the SM boxplots of these four land types in the four data sets. As shown in Figure 8a, the D1SM boxplots are closer to that of SMAP than those of CLDAS and GLDAS. Satellite-based products including the D1SM and the SMAP show dry Barren land and wet Woody Savannas and Evergreen Needleleaf Forests. The D1SM and the CLDAS are closest in terms of Grasslands, indicating that the downscaling method works best for Grasslands and improves the accuracy of the original SMAP because of more samples and the better effect of VI in Grasslands than that in Barren land. In addition, the SM boxplots of all LCTs of the D1SM are shown in Figure 8b. The highest SM is in Cropland/Natural Vegetation Mosaics (C14), followed by Evergreen Broadleaf Forests (C2). The lowest SM is in Open Shrublands (C7), followed by Barren land (C16). Statistical errors were high for Permanent Snow and Ice (C15) and Water Bodies (C17) because they were removed when building the model. There are no data on Deciduous Needleleaf Forests (C3) because their cover was much too small. The means of SM of other land types are all close to 0.2 m³/m³.

### Table 5

|                  | Mean (m³/m³) | Standard deviation (m³/m³) | Kurtosis | Skewness |
|------------------|--------------|----------------------------|----------|----------|
| D1SM             | 0.12         | 0.07                       | 4.86     | 1.21     |
| SMAP             | 0.12         | 0.08                       | 9.09     | 2.07     |
| CLDAS            | 0.16         | 0.07                       | 4.62     | 0.81     |
| GLDAS            | 0.21         | 0.08                       | 3.68     | 0.94     |

The general consistency of the D1SM with the spatial distribution of the other three data sets demonstrates the accuracy of the D1SM and the feasibility of this method. Compared with the original SMAP data, the D1SM is more refined, and the local differences are more obvious, so the fine-scale SM values can be obtained. The D1SM at the southwestern edge of TP are in agreement with those from SMAP but are slightly different from those from CLDAS and GLDAS. GLDAS has higher values due to its deeper measurement depth.

### 4.3.3. Land Cover Type Analysis

To analyze the overall data distributions of SM in different land cover types (LCTs) and examine the effect of the downscaling method in detail, the four most common types of land cover are selected over the TP according to their frequencies in the histogram. These include Class 10 Grasslands (52.53%), Class 16 barren lands (38.17%), Class 8 Woody Savannas (2.12%), and Class 1 Evergreen Needleleaf Forests (1.47%), which can be seen in Figure 1b. We analyzed the SM boxplots of these four land types in the four data sets. As shown in Figure 8a, the D1SM boxplots are closer to that of SMAP than those of CLDAS and GLDAS. Satellite-based products including the D1SM and the SMAP show dry Barren land and wet Woody Savannas and Evergreen Needleleaf Forests. The D1SM and the CLDAS are closest in terms of Grasslands, indicating that the downscaling method works best for Grasslands and improves the accuracy of the original SMAP because of more samples and the better effect of VI in Grasslands than that in Barren land. In addition, the SM boxplots of all LCTs of the D1SM are shown in Figure 8b. The highest SM is in Cropland/Natural Vegetation Mosaics (C14), followed by Evergreen Broadleaf Forests (C2). The lowest SM is in Open Shrublands (C7), followed by Barren land (C16). Statistical errors were high for Permanent Snow and Ice (C15) and Water Bodies (C17) because they were removed when building the model. There are no data on Deciduous Needleleaf Forests (C3) because their cover was much too small. The means of SM of other land types are all close to 0.2 m³/m³.

### Table 5

|                  | Mean (m³/m³) | Standard deviation (m³/m³) | Kurtosis | Skewness |
|------------------|--------------|----------------------------|----------|----------|
| D1SM             | 0.12         | 0.07                       | 4.86     | 1.21     |
| SMAP             | 0.12         | 0.08                       | 9.09     | 2.07     |
| CLDAS            | 0.16         | 0.07                       | 4.62     | 0.81     |
| GLDAS            | 0.21         | 0.08                       | 3.68     | 0.94     |

The general consistency of the D1SM with the spatial distribution of the other three data sets demonstrates the accuracy of the D1SM and the feasibility of this method. Compared with the original SMAP data, the D1SM is more refined, and the local differences are more obvious, so the fine-scale SM values can be obtained. The D1SM at the southwestern edge of TP are in agreement with those from SMAP but are slightly different from those from CLDAS and GLDAS. GLDAS has higher values due to its deeper measurement depth.

### 4.3.3. Land Cover Type Analysis

To analyze the overall data distributions of SM in different land cover types (LCTs) and examine the effect of the downscaling method in detail, the four most common types of land cover are selected over the TP according to their frequencies in the histogram. These include Class 10 Grasslands (52.53%), Class 16 barren lands (38.17%), Class 8 Woody Savannas (2.12%), and Class 1 Evergreen Needleleaf Forests (1.47%), which can be seen in Figure 1b. We analyzed the SM boxplots of these four land types in the four data sets. As shown in Figure 8a, the D1SM boxplots are closer to that of SMAP than those of CLDAS and GLDAS. Satellite-based products including the D1SM and the SMAP show dry Barren land and wet Woody Savannas and Evergreen Needleleaf Forests. The D1SM and the CLDAS are closest in terms of Grasslands, indicating that the downscaling method works best for Grasslands and improves the accuracy of the original SMAP because of more samples and the better effect of VI in Grasslands than that in Barren land. In addition, the SM boxplots of all LCTs of the D1SM are shown in Figure 8b. The highest SM is in Cropland/Natural Vegetation Mosaics (C14), followed by Evergreen Broadleaf Forests (C2). The lowest SM is in Open Shrublands (C7), followed by Barren land (C16). Statistical errors were high for Permanent Snow and Ice (C15) and Water Bodies (C17) because they were removed when building the model. There are no data on Deciduous Needleleaf Forests (C3) because their cover was much too small. The means of SM of other land types are all close to 0.2 m³/m³.
4.4. Difference Analysis

The differences between the D1SM and CLDAS (GLDAS) are assessed in the boxplots (Figure 9a) and the probability density function curves (Figure 9b). It is clear from Figure 9a that the means of the differences are close to 0 m$^3$/m$^3$, which indicates that the differences between the D1SM and the CLDAS (GLDAS) data are small. The small differences suggest that the downscaling algorithm is effective. The values are mainly negative, which indicates that the D1SM is a little lower overall. The probability density function curves
of the differences in Figure 9b demonstrate the differences between the D1SM and CLDAS (GLDAS). The distributions are close to the normal distribution.

To further study the consistency of the D1SM and the CLDAS (GLDAS) data, we analyzed the spatial distributions of the differences in SM data between the D1SM and the CLDAS (GLDAS) over the TP. It can be seen from Figure 10 that the spatial differences are small overall, which is consistent with the conclusions from Figure 9. Most of the differences from CLDAS are between −0.1 and 0.1 m³/m³ and typically negative. However, the D1SM is higher than CLDAS in a few small parts located in the south low-altitude evergreen broadleaf forest area, the Qinghai Lake area, and the Yellow River basin in the east of TP. This is probably due to the overestimation over the SM rich regions relative to CLDAS. At the same time, the D1SM is lower than GLDAS in most areas, especially in the eastern hinterland and the northwest. This is reasonable since the data depth of GLDAS is deeper. Nevertheless, in a few areas where the altitude is less than 3,000 m (e.g., the Qaidam Basin, Qinghai Lake Area, the turn of the Yellow River, and evergreen broadleaf forest in the south), the differences are still positive. This means that this algorithm tends to overestimate at low altitudes compared to GLDAS.

4.5. Correlation Analysis

Scatter plots and the linear fittings (red lines) of the SM data between the D1SM and CLDAS and GLDAS are shown in Figure 11. Table 6 lists the correlation analysis and the error analysis data of the D1SM with respect to CLDAS and GLDAS. Because the resolutions of CLDAS and GLDAS are 6 and 25 km respectively, whereas the resolution of the D1SM is 1 km, we use the nearest neighbor method for pairwise matching, applying the coarse scale as the standard. Therefore, there is a certain error in the spatial alignment. Despite this, the correlation and consistency are still good. As seen in Table 6, these CCs passed a 95% confidence test, with higher CCs (0.67 and 0.70 m³/m³) and lower Mean Absolute Errors (MAEs) (0.05 and 0.09 m³/m³) and Root Mean Squared Errors (RMSEs) (0.03 and 0.08 m³/m³). In summary, the D1SM is consistent with CLDAS and GLDAS.


5. Conclusions and Discussions

With the rapid changes in global warming, the SM in most of the TP has been humidified over the past three decades (Yin et al., 2016). Changes in the SM of the TP affect its sensible heat, which in turn affects the formation of the Asian monsoon, leading to significant climate changes in Asia. Due to the harsh natural environment and the cold and complex geographical features, there are few SM observation stations on the TP, and these tend to be concentrated on a few small areas, leading to a scarcity in ground SM data for this area. In recent years, with the development of satellite remote sensing technology, the large-scale detection of SM has been realized, but the resolution of remote sensing SM products is low and cannot meet the needs of fine-scale applications. In order to downscale the spatial scale of SMAP 36 km SM products to 1 km over the TP, this study uses the RF algorithm to build a model for predicting the SM using VI, LST, and LCT products with MODIS and DEM terrain data, and experimental downscaling to estimate the SM data at a 1 km scale. The overall conclusions are as follows:

1. The relationship between SM and the environmental variables of the land surface in complex terrains is complex and nonlinear; therefore, it is difficult to create effective simulations using traditional statistical models, while machine-learning algorithms have advantages for the simulation of multivariate, nonlinear complex relationships. In this paper, we use the RF algorithm of machine learning, combined with multisource remote sensing data and terrain elements, to design and implement the downscaling method for use with the SM data, which can solve the limitations of traditional algorithms for application on complex terrains.

2. Using the prediction principle of the RF algorithm, NDVI, EVI, LST, LCT, DEM, longitude, latitude, and other environmental factors at the surface that are closely related to SM are selected as auxiliary variables to establish a data sample set for training. The downscaling prediction of SM is based on the constructed training model. The preliminary results show that the RF algorithm can fit the relationship between SM and the relative auxiliary variables in order to predict fine-scale SM.

3. The preliminary results are given by analyzing the correlations between SM and variables, in situ evaluation, statistical characteristics, spatial distribution, LCTs, differences, and correlations. The results show that (a) the SM data are affected by various factors and the most important factors are NDVI, EVI, and longitude in this model. (b) The consistency between the D1SM and the in situ observations is better than that of the SMAP, CLDAS, and GLDAS. (c) The D1SM is generally consistent with the comparative data in spatial distribution but provides more details. (d) The satellite-based data are drier in Barren and wetter in Woody Savannas and Evergreen Needleleaf Forests. The algorithm works best in Grasslands. The D1SM is wettest in Cropland/Natural Vegetation Mosaics, whereas the driest LCT is Open Shrublands. (e) Differences of the data sets in spatial distributions are little, with concentrated statistics distributions. The D1SM is lower than CLDAS (GLDAS) in most areas, but higher than CLDAS and GLDAS in wet SM areas and low-altitude areas, respectively.

The above results indicate that it is feasible to downscale the SM products over complex terrain using machine-learning methods. The key contributions of this study are listed as follows. First, this study has laid a practical foundation for the research of SM downscaling over complex terrain. Based on this study, different characteristic variables, more advanced machine-learning algorithms, and long-term input data can be used for further studies. Second, it shows the potential to downscale long-term SM products over TP. Third, this study provides a substitute for the malfunctioned 3 km SMAP product to obtain fine-scale SM again. Lastly, this method can be potentially applied to other variables requiring downscaling (e.g., radiation, wind speed, humidity, and precipitation). Despite the above contributions, there are still some limitations in the study that requires further study. (1) This study only considered spatial domain information. Due to the temporal evolution of SM, the time factor (e.g., season) should be added as an independent variable in the model. (2) Since the MODIS land products are only available in clear-sky conditions, the proposed algorithm may not be applicable in the existence of cloud cover. This limitation may be addressed through the application of cloud restoration algorithms in future studies (Shen et al., 2014). (3) A few important factors

---

**Table 6**

|                      | D1SM(Y) and CLDAS(X) | D1SM(Y) and GLDAS(X) |
|----------------------|----------------------|----------------------|
| Equation             | Y = 0.64 * X + 0.02  | Y = 0.62 * X − 0.00  |
| R²                   | 0.45                 | 0.49                 |
| Average X (m³/m³)    | 0.16                 | 0.21                 |
| Average Y (m³/m³)    | 0.13                 | 0.12                 |
| CC                   | 0.67                 | 0.70                 |
| MAE (m³/m³)          | 0.05                 | 0.09                 |
| RMSE (m³/m³)         | 0.03                 | 0.08                 |
(e.g., precipitation, albedo, and evapotranspiration) are not used as inputs to the machine-learning model.

Radar-based high-resolution precipitation analysis will be considered in subsequent studies to improve the accuracy of the model.

References

Bao, H. Y., Yang, K., & Wang, C. H. (2017). Characteristics of GLDAS soil-moisture data on the Tibetan Plateau. *Sciences in Cold and Arid Regions*, 9(2), 127–141.

Bartalis, Z., Wagner, W., Naemini, V., Hasenauer, S., Scipal, K., Bonekamp, H., et al. (2007). Initial soil moisture retrievals from the METOP-A advanced scatterometer (ASCAT). *Geophysical Research Letters*, 34, L20401. https://doi.org/10.1029/2007GL031088

Beyene, A., Cornelis, W., Verhoest, N. E. C., Tilahun, S., Alamirew, T., Adgo, E., et al. (2018). Estimating the actual evapotranspiration and deep percolation in irrigated soils of a tropical floodplain, northwest Ethiopia. *Agricultural Water Management*, 202, 42–56. https://doi.org/10.1016/j.agwat.2018.01.022

Breiman, L. (2001). *RANDOM FORESTS*. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cai, D., Fraelich, K., Sielmann, F., Guan, Y., Guo, S., Zhang, L., & Zhu, X. (2014). Climate and vegetation: An ERA-Interim and GMIMS NDI analysis. *Journal of Climate*, 27(13), 5111–5118. https://doi.org/10.1175/JCLI-D-13-00674.1

Cao, W., Fraelich, K., Sielmann, F., Zhang, L., Zhu, X., Guo, S., & Guan, Y. (2015). Vegetation dynamics on the Tibetan Plateau (1982–2006): An attribution by ecohydrological diagnostics. *Journal of Climate*, 28(11), 4575–4584. https://doi.org/10.1175/JCLI-D-14-00692.1

Cao, W., Cheng, Y., Wu, J., Li, J., & Wang, S. (2017). Spatial variability of permafrost soil moisture on the slope of the Qinghai-Tibet plateau (in Chinese). *Advances in Water Science*, 28(1), 32–40. https://doi.org/10.1007/s12205-017-01004

Chen, Q., Miao, F., Xu, Z.-X., Wang, H., Yang, L., & Tang, Z. (2019). Downscaling of remote sensing soil moisture products based on TVDI in complex terrain areas. Paper presented at 2019 International Conference on Meteorology Observations, Chengdu, China. https://doi.org/10.1109/icmo49322.2019.9026012

Chen, Y., Yang, K., Qin, J., Cui, Q., Lu, H., La, Z., et al. (2017). Evaluation of SMAP, SMOS, and AMSR2 soil moisture retrievals against observations from two methods on the Tibetan Plateau. *Journal of Geophysical Research: Atmospheres*, 122, 5780–5792. https://doi.org/10.1002/2016JD026388

Cheng, Z., & Zhang, X. (2017). Evaluation method of data from soil moisture automatic observation station (in Chinese). *Chinese Agricultural Science Bulletin*, 33(35), 113–118.

Collow, T. W., Robock, A., Basara, J. B., & Illston, B. G. (2012). Evaluation of SMOS retrievals of soil moisture over the central United States with currently available in situ observations. *Journal of Geophysical Research*, 117, D09113. https://doi.org/10.1029/2011JD017095

Cowley, G. S., Niemann, J. D., Green, T. R., Seyfried, M. S., Jones, A. S., & Grazulis, P. J. (2017). Impacts of precipitation and potential evapotranspiration patterns on downsampling soil moisture in regions with large topographic relief. *Water Resources Research*, 53, 1553–1574. https://doi.org/10.1002/2016WR019907

Cui, Y., Qin, J., Jing, W., & Tan, G. (2018). Applicability evaluation of merged soil moisture in GLDAS and CLDAS products over Qinghai-Tibetan Plateau (in Chinese). *Plateau Meteorology*, 37(1), 123–136. https://doi.org/10.7522/j.issn.1000-0534.2017.00035

Cutler, A., Cutler, D. R., & Stevens, J. R. (2011). Random forests. In C. Zhang, & Y. Ma (Eds.), *Ensemble machine learning: Methods and applications* (Vol. 41, pp. 157–175). Berlin: Springer Science+Business Media. https://doi.org/10.1007/978-1-4419-9326-7_5

Das, N. N., Entekhabi, D., Njoku, E. G., Shi, J. J. C., Johnson, J. T., & Collander, A. (2014). Tests of the SMAP combined radar and radiometer algorithm using airborne field campaign observations and simulated data. *IEEE Transactions on Geoscience and Remote Sensing*, 52(4), 2018–2028. https://doi.org/10.1109/TGRS.2013.2257605

Entekhabi, D., Njoku, E. G., O’Neill, P. E., Kellogg, E. H., Crow, W. T., Edelstein, W. N., et al. (2010). The Soil Moisture Active Passive (SMAP) mission. *Proceedings of the IEEE*, 98(5), 704–716. https://doi.org/10.1109/JPROC.2010.2043918

Fischer, E. M., Seneviratne, S. I., Vidalé, P. L., Lüthi, D., & Schar, C. (2007). Soil moisture-atmosphere interactions during the 2003 European summer heat wave. *Journal of Climate*, 20(20), 5081–5099. https://doi.org/10.1175/JCLI4288.1

Hesham, D. C., Niemann, J. D., Green, T. R., Jones, A. S., & Grazulis, P. J. (2017). Downsampling soil moisture over regions that include multiple coarse-resolution grid cells. *Remote Sensing of Environment*, 199, 387–200. https://doi.org/10.1016/j.rse.2017.07.021

Im, J., Park, S., Rhee, J., Baik, J., & Choi, M. (2016). Downscaling of AMSR-E soil moisture with MODIS products using machine learning approaches. *Environmental Earth Sciences*, 75(15), 1120. https://doi.org/10.1007/s12665-016-5917-5

Jia, Y. C., Xie, M. W., & Jiang, H. T. (2017). Daily estimate of global 36 km grid soil moisture (in Chinese). *Journal of Geo-information Science*, 19(6), 854–860. https://doi.org/10.7646/SPF.JIF.2017.06.0054

Justice, C. O., Vermote, E., Townshend, J. R. G., Defries, R., Roy, D. P., Hall, D. K., et al. (1998). The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing*, 36(4), 1228–1249. https://doi.org/10.1109/36.701075

Kerr, Y. H., Waldteufel, P., Wigneron, J. P., Delwart, S., Cabot, F., Boutin, J., et al. (2010). The SMOS mission: New tool for monitoring key elements of the global water cycle. *Proceedings of the IEEE*, 98(5), 666–687. https://doi.org/10.1109/JPROC.2010.2040302

Kuang, X., & Xiao, J. J. (2016). Review on climate change on the Tibetan Plateau during the last half century. *Journal of Geophysical Research: Atmospheres*, 121, 3979–4007. https://doi.org/10.1002/2015JD024728

Kwon, M., Kwon, H. W., & Han, D. (2017). A spatial downsampling of soil moisture from rainfall, temperature, and AMSR2 using a Gaussian-Mixture Nonstationary Hidden Markov Model. *Journal of Hydrology*, 564, 1194–1207. https://doi.org/10.1016/j.jhydrol.2017.12.015

Leroux, D., Das, N. N., Entekhabi, D., Collander, A., Njoku, E., Jackson, T. J., & Yueh, S. (2016). Active-passive soil moisture retrievals during the SMAP validation experiment 2012. *IEEE Geoscience and Remote Sensing Letters*, 13(4), 475–479. https://doi.org/10.1109/lgrs.2015.2491163

Mascaro, G., Viven, E. R., & Deidda, R. (2011). Soil moisture downsampling across climate regions and its emergent properties. *Journal of Geophysical Research: Atmospheres*, 116, D22214. https://doi.org/10.1029/2011JD016231

Minet, J., Whyhudi, A., Bogaert, P., Vancooeter, M., & Sébastien, L. (2011). Mapping shallow soil moisture profiles at the field scale using full-waveform inversion of ground penetrating radar data. *Geoderma*, 161(3–4), 225–237. https://doi.org/10.1016/j.geoderma.2010.12.023

Montzka, C., Jagdhuber, T., Horn, R., Bogaen, H. R., Hajnsek, I., Reigber, A., & Vereecken, H. (2016). Investigation of SMAP fusion algorithms with airborne active and passive L-band microwave remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 54(7), 3878–3889. https://doi.org/10.1109/tgrs.2016.2529659
