Gap Filling of High-Resolution Soil Moisture for SMAP/Sentinel-1: A Two-Layer Machine Learning-Based Framework — Source link ⧉

Hanzi Mao, Dhruva Kathuria, Nick Duffield, Binayak P. Mohanty

Institutions: Texas A&M University

Related papers:

- Soil Moisture Retrieval Only Using Smap L-Band Radar Observations

- The SMAP Level 4 Surface and Root-zone Soil Moisture (L4_SM) Product

- Medium-Resolution Soil Moisture Retrieval Using the Bayesian Merging Method

- Smap Radiometer Soil Moisture Downscaling in Conus

- Investigation of SMAP Active–Passive Downscaling Algorithms Using Combined Sentinel-1 SAR and SMAP Radiometer Data

GAP FILLING OF HIGH-RESOLUTION SOIL MOISTURE FOR SMAP/SENTINEL-1: A TWO-LAYER MACHINE LEARNING-BASED FRAMEWORK

Hanzi Mao
Department of Computer Science and Engineering
Texas A&M University
College Station, TX 77843
hzmao@tamu.edu

Dhruva Kathuria
Biological and Agricultural Engineering
Texas A&M University
College Station, TX 77843
kathuria.dhruva@tamu.edu

Nick Duffield
Department of Electrical and Computer Engineering
Texas A&M University
College Station, TX 77843
duffieldng@tamu.edu

Binayak P. Mohanty
Biological and Agricultural Engineering
Texas A&M University
College Station, TX 77843
bmohanty@tamu.edu

ABSTRACT

As the most recent 3 km soil moisture product from the Soil Moisture Active Passive (SMAP) mission, the SMAP/Sentinel-1 L2_SM_SP product has a unique capability to provide global-scale 3 km soil moisture estimates through the fusion of radar and radiometer microwave observations. The spatial and temporal availability of this high-resolution soil moisture product depends on concurrent radar and radiometer observations which is significantly restricted by the narrow swath and low revisit schedule of the Sentinel-1 radars. To address this issue, this paper presents a novel two-layer machine learning-based framework which predicts the brightness temperature and subsequently the soil moisture at gap areas. The proposed method is able to gap-fill soil moisture satisfactorily at areas where the radiometer observations are available while the radar observations are missing. We find that incorporating historical radar backscatter measurements (30-day average) into the machine learning framework boosts its predictive performance. The effectiveness of the two-layer framework is validated against regional hold-out SMAP/Sentinel-1 3 km soil moisture estimates at four study areas with distinct climate regimes. Results indicate that our proposed method is able to reconstruct 3 km soil moisture at gap areas with higher Pearson correlation coefficient (47%/35%/20%/80% improvement of mean R, at Arizona/Oklahoma/Iowa/Arkansas) and lower unbiased Root Mean Square Error (20%/10%/7%/26% improvement of mean ubRMSE) when compared to the SMAP 33 km soil moisture product. Additional validation against airborne data and in-situ data from soil moisture networks is also satisfactory.

Keywords Soil Moisture · Machine Learning · Multi-Resolution Gap Filling · SMAP satellite · Sentinel-1 satellite · Spatial/Temporal Transfer Learning

1 Introduction

Soil moisture is an essential variable controlling the exchange of water and energy fluxes between the land surface and the atmosphere. Retrieving high-resolution soil moisture at local, regional and global scales is critical for best agricultural and irrigation practices [1], weather and climate predictions [2, 3], biogeochemical process characterizations [4] and drought and flood predictions [5, 6]. To complement the sparsely distributed ground-sampled soil moisture

measurements [7, 8, 9], various microwave remote sensing approaches have been developed over the last few decades to retrieve soil moisture estimates at a global scale such as the Advanced Scatterometer [10], the Advanced Microwave Scanning Radiometer-EOS [11], the Soil Moisture and Ocean Salinity (SMOS) [12], the Soil Moisture Active Passive (SMAP) mission [13], and the Sentinel-1 mission [14]. Global-scale microwave sensing soil moisture products have the advantage of capturing soil moisture heterogeneity over large spatial extents. This further benefits regional and global scale applications that are impacted by land-atmosphere interactions and/or soil moisture variability [15, 16, 17].

Existing microwave remote sensing techniques measure the near surface soil moisture ($\leq$ 5 cm) indirectly through its relationship with water content and soil dielectric constant. They can be classified into two categories according to the source of their signals, 1) the active radar measuring the backscattered signal after transmitting an active pulse and 2) the passive radiometer measuring the land surface emissions. Active radars typically retrieve soil moisture at a spatial resolution varying from several meters to ∼3 km. In addition to soil moisture, they are also highly sensitive to surface roughness and vegetation, which reduces the accuracy of the retrieved soil moisture estimates [18, 19]. On the contrary, the relatively mature passive-radiometers have the ability to provide accurate estimates of soil moisture due to their high sensitivity to soil moisture [20]. The high accuracy of the soil moisture estimates retrieved by the passive remote sensing is, however, compromised by its coarse resolution, ∼40 km for both the SMOS and SMAP passive soil moisture products [12, 21], which restricts the application of the resulting soil moisture products for regional hydro-climatological studies [22, 13, 23, 24].

Among the various methods proposed to downscale coarse-scale soil moisture [25, 26], the fusion of radar and radiometer microwave observations is one of the most promising and robust techniques to provide high resolution (1-10 km) soil moisture estimates at a global scale. The SMAP mission, specifically, adopts the active-passive retrieval algorithm proposed by [27]. In this two-layer retrieval algorithm, the radar backscatter measurements are first used to downscale the coarse-scale passive brightness temperature measurements, from which the high resolution soil moisture is then retrieved through the single channel algorithm (SCA) [28]. The original SMAP observatory was launched on January 31, 2015. Equipped with an L-band radiometer and an L-band radar as a single observation system, its mission was to combine 36 km brightness temperature and 3 km radar backscatter to provide 9 km soil moisture at a global scale [13, 29].

Unfortunately, the SMAP L-band radar unexpectedly ceased its operation on July 7, 2015. To recover the capability of providing high resolution soil moisture, one possible solution was to replace the SMAP L-band radar measurements with the C-band radar measurements from the European Space Agency (ESA)'s Sentinel-1 radars, Sentinel-1A and Sentinel-1B [30, 31]. Many studies have investigated the performance of integrating the C-band active and the L-band passive observations to retrieve high resolution soil moisture with promising results [30, 32, 33, 34].

As the most recent 3 km soil moisture product from the SMAP mission, the SMAP/Sentinel-1 product L2_SM_SP [35] combines the 33 km brightness temperature measurements from the SMAP enhanced passive soil moisture product L3_SM_P_E with the 3 km aggregated co-polarization (co-pol) and cross-polarization (cross-pol) backscatter measurements retrieved by the Sentinel-1 radars. [35] adopt the SMAP/Sentinel active-passive retrieval algorithm [36] similar to the original SMAP active-passive algorithm where high resolution brightness temperature and soil moisture are downscaled consecutively. One restriction of this product, however, is that the Sentinel-1 radar observations have much narrower swath width (∼250 km) compared with the relatively wide swath width (1000 km) of the SMAP radar and radiometer. As the SMAP/Sentinel active-passive retrieval algorithm requires concurrent radar and radiometer observations, the narrow swath of the Sentinel-1 radar observations significantly reduces the spatial coverage of the resulting daily soil moisture product compared with the SMAP enhanced passive 33 km soil moisture product (Figure 1). The temporal resolution also drops from a 3-day revisit schedule to a 6-day one when using the collocated Sentinel-1A and Sentinel-1B radar backscatter observations [13, 30, 37]. Additionally, considering that the Sentinel-1A and Sentinel-1B sensors provide backscatter measurements with different swath patterns and interleaved revisit schedules, the temporal resolution further drops to 12 days or more for some locations. The narrow swath and low revisit schedule of the SMAP/Sentinel-1 3 km soil moisture cannot achieve the required temporal resolution for many soil moisture related applications [26, 38].

This paper thus focuses on downscaling/gap-filling soil moisture at areas where the SMAP/Sentinel active-passive retrieval algorithm cannot be applied due to missing radar backscatter measurements. [25] and [26] outline the various techniques to downscale passive remote sensing soil moisture. The techniques that do not rely on concurrent radar observations can be applied to gap-fill the SMAP/Sentinel-1 3 km soil moisture product, such as fusion of optical/thermal and microwave methods [39, 40, 41], data assimilation methods [42, 43], and machine learning methods [44, 45, 46, 47, 48]. Among these approaches, machine learning techniques have shown great potential to downscale soil moisture with its ability to handle noisy data and learning complex non-linear processes [49, 50]. Existing machine learning methods capture the relationship between coarse-scale and fine-scale soil moisture directly with the aid of auxiliary environmental variables, such as soil properties, topography, land cover, and meteorological forcings. The
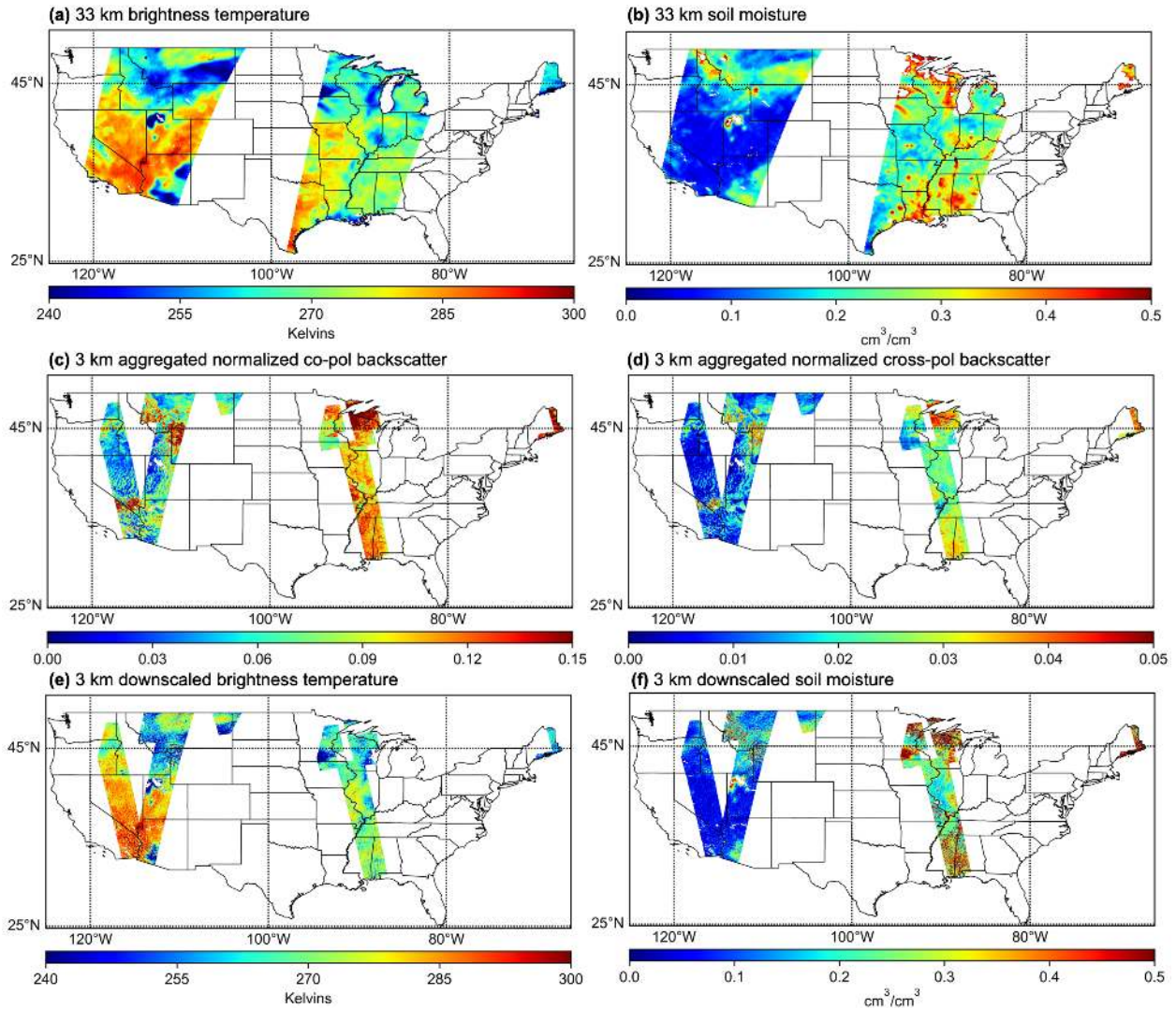
Figure 1: Spatial availability of the SMAP/Sentinel-1 active-passive 3 km soil moisture product (L2_SM_SP) and the SMAP enhanced passive 33 km soil moisture product (L3_SM_P_E) on June 16, 2018 for the Continental United States. (a) 33 km brightness temperature retrieved by the SMAP radiometer. (b) 33 km soil moisture converted from (a) using the single channel algorithm. (c) 3 km aggregated normalized co-pol backscatter retrieved by the Sentinel-1A/Sentinel-1B radars. (d) 3 km aggregated normalized cross-pol backscatter retrieved by the Sentinel-1A/Sentinel-1B radars. (e) 3 km downscaled brightness temperature from the SMAP-Sentinel active-passive algorithm. (f) 3 km downscaled soil moisture converted from (e) using the single channel algorithm.

fine-scale soil moisture estimates used to train/calibrate the machine learning models are restricted to ground data obtained from either catchment-scale campaigns [45, 48] or sparsely distributed soil moisture networks [44].

In this paper, estimates from the SMAP/Sentinel-1 active-passive 3 km soil moisture product serve as target variables to train the machine learning models. This greatly expands the training data size and has the potential to learn complex relationships that exists between coarse-scale and fine-scale soil moisture [51]. Directly using remote sensing soil moisture estimates for training a machine learning model also has the advantage of maintaining high consistency, or "high fidelity" [52] between the downscaled predictions and the SMAP/Sentinel-1 3 km soil moisture product. We propose a two-layer machine learning-based framework motivated by the SMAP/Sentinel active-passive retrieval algorithm to improve the consistency between the downscaled soil moisture and the retrieved 3 km soil moisture from the SMAP/Sentinel-1 soil moisture product. The downscaled brightness temperature and soil moisture are predicted consecutively using machine-learning models in the two layers of the framework. Instead of directly downscaling the 33 km soil moisture from the SMAP enhanced passive soil moisture product to 3 km, the 33 km brightness temperature (from the same product) is first downscaled to 3 km using a machine learning model. A separate machine learning model is then applied to convert the 3 km downscaled brightness temperature to 3 km soil moisture. This machine learning approach, different from the SMAP/Sentinel active-passive retrieval algorithm, removes the dependence on concurrent radar and radiometer observations, and thus can be applied to gap fill soil moisture at areas where the SMAP/Sentinel active-passive retrieval algorithm cannot be used. Compared with the conventional one-layer machine learning framework that directly downscales soil moisture from 33 km to 3 km, the two-layer framework avoids the dependence on the 33 km soil moisture that has additional uncertainties incurred by the indirect conversion from the 33 km brightness temperature. Moreover, it allows us to naturally introduce the average of historical (past 30 days) radar backscatter measurements as a key input to downscale the 33 km brightness temperature. This is critical as the vegetation and surface heterogeneity captured by the average of historical radar backscatter measurements can help produce more accurate downscaled brightness temperature, from which high resolution soil moisture with higher accuracy can be subsequently derived.

Using the proposed framework, downscaled soil moisture that is highly consistent with the SMAP/Sentinel-1 3 km soil moisture product can be reconstructed at gap areas where the SMAP/Sentinel active-passive retrieval algorithm fails. The framework is a general one though, and can embed any machine learning model to downscale brightness temperature/soil moisture. We implement both a general one-layer framework that directly downscales 33 km soil moisture to 3 km and our proposed two-layer framework with the Random Forest [53] serving as the predictive machine-learning model (for both the layers). The Random Forest model is specifically chosen here for its ability to handle complex non-linear relationships with reduced overfitting. It has also shown promising results in many soil moisture downscaling studies [44, 46, 54, 55].

In this paper, relationships between the SMAP/Sentinel-1 3 km soil moisture/brightness temperature and the input features (ancillary data) are learned through machine learning models at regions where the 3 km soil moisture/brightness temperature are available. The learned models are then applied to predict 3 km soil moisture/brightness temperature at spatially/temporally neighboring gap regions where the Sentinel radar observations and thus the 3km soil moisture and brightness temperature values are missing. We validate the performance of the models in disparate regions across Contiguous US (CONUS) from April 1st, 2015 to September 30, 2018 by predicting hold-out SMAP-Sentinel-1 3 km soil moisture data. The efficacy of the proposed framework is also demonstrated by comparing it with in-situ measurements from soil moisture networks as well as with airborne soil moisture data retrieved during the SMAPVEx 2016 campaign in Iowa.

The contribution of this paper is threefold. First, it is the first study using machine learning to address the poor spatial and temporal coverage for downscaled soil moisture estimates retrieved by the radar-radiometer microwave combination approach. Second, it designs a novel two-layer machine learning framework motivated by the original SMAP/Sentinel active-passive retrieval algorithm to improve 3 km soil moisture predictions. Third, the 30-day average of historical radar backscatter measurements is introduced for the first time as a key input to aid in capturing subgrid soil moisture heterogeneity. The rest of the paper is organized as follows. Section 2 describes the data sets and study area. Section 3 gives a brief introduction of the SMAP/Sentinel active-passive retrieval algorithm. The specifics of the methodology are presented in Section 4. The results and validations are discussed in Section 5. The conclusions and discussion for future work are then provided in Section 6.

2   Data Sets and Study Area

2.1   Soil moisture and brightness temperature

The most recent SMAP/Sentinel-1 active-passive soil moisture product (L2_SM_SP) provides 3 km soil moisture and 3 km brightness temperature that can serve as the learning targets of our machine learning models. As a level

2 product, for any specific day, there are hundreds of data files each containing a sub-patch of the globally covered data. We generate daily composites from this level 2 product with the averaged value applied to overlapping estimates. Besides the 3 km brightness temperature and soil moisture, 3 km radar backscatter cross-section measurements and other physically-based parameters are available in L2_SM_SP from which averages of historical (past 30 days) values can be generated for both covered and gap regions. We take the 33 km soil moisture and 33 km brightness temperature (values are posted on 9 km resolution grids) from the SMAP enhanced passive soil moisture product (L3_SM_P_E). For both the aforementioned SMAP soil moisture products, only the descending overpass, or local 6 am, data is used in this paper considering the thermal equilibrium effect [56]. These products are gridded using the Equal-Area Scalable Earth-2 (EASE2) projection [57] where the 3 km subgrids are nested consistently inside the 9 km grid [58]. In-situ soil moisture observations from two soil moisture networks, Soil Climate Analysis Network (SCAN) [59] and U.S. Climate Reference Network (USCRN) [60] were used for additional validation at selected study areas from April, 2015 to September, 2018. Airborne soil moisture data for South Fork, Iowa (August 2016) was retrieved using a Passive Active L-band System (PALS) equipped aircraft and is available on the National Snow and Ice Data Center (NSIDC) website.

2.2 Ancillary data for proposed two-layer framework

Ancillary data used in the SMAP/Sentinel active-passive retrieval algorithm are also used in the proposed two-layer framework to capture the fine-scale heterogeneities in soil moisture and brightness temperature. Soil properties represented by the percent sand, percent clay, and bulk density were obtained from Soil Geographic (STATSGO) Data Base [61] with 1 km resolution. Land cover class was extracted from the MCD12Q1 [62] with 500 m resolution. Both the soil properties and land cover class are rescaled to 3 km through averaging. Effective soil temperature, scattering albedo, vegetation water content, vegetation opacity, and roughness coefficient were obtained from the 33 km SMAP enhanced passive soil moisture product and rescaled to 3 km using bilinear interpolation.

2.3 Ancillary data for traditional one-layer framework

Different from our proposed two-layer framework, existing machine learning methods normally predict soil moisture by capturing its relationship with other environmental variables, such as soil properties, topography, land cover, and meteorological forcings. We also implement this one-layer framework to compare with the proposed two-layer approach. Again, we take the soil properties from the STATSGO and landcover class from MCD12Q1. The elevation data were obtained from the NASA Shuttle Radar Topography Mission Global 3 arc second product (SRTMGL3) [63]. Additional land surface controls represented by leaf area index and land surface temperature were extracted from the MODIS products, MCD15A3H (Myneni et al., 2015) with 500 m resolution, and MOD11A1 (Wan et al., 2015) with 1 km resolution respectively. Precipitation forcing data were obtained from the Global Precipitation Measurement [64] 3IMERGDF product with ~11 km resolution. All these inputs were resampled to the 3-km EASE2-Grid projection through linear averaging (for inputs with resolutions higher than 3 km) or using bilinear interpolation (for inputs with resolutions lower than 3 km) to match the grid size of the learning targets. Slope was then computed from the resampled elevation data sets. Land surface temperature and precipitation data have a daily temporal resolution while the leaf area index has a 4-day temporal resolution. We assume that the leaf area index stays the same for each 4-day period.

Additional post-processing was applied to land surface temperature and precipitation data. As retrieval for land surface temperature depends on the satellite thermal infrared channels, invalid surface temperature values are generated for surfaces that are obscured by clouds [65]. For these grids with invalid values, we search for historical values within a two-week finite window and select the most recent valid historical values as substitutes. For precipitation, to incorporate its lagged effect on vegetation, surface characteristics and soil moisture, instead of using the current-day precipitation values, we keep historical precipitation values from the most recent three days separately as inputs to the machine learning models.

2.4 Study Area

We apply the proposed framework separately at four US states with distinct climate regimes and land surface hetero-geneities: Arizona, Oklahoma, Iowa, and Arkansas (Figure 2). Arizona has primarily an "arid" climate (BSk and BWh classification), Oklahoma is "warm-temperate" (Cfa classification), Iowa is primarily "snow" with fully-humid hot summer (Dfa classification) while Arkansas is "warm temperate" (Csc classification) according to the Köppen-Gieger classification [66]. While Arizona, Oklahoma, and Iowa are predominantly composed of open shrublands (70.8%), grasslands (75.5%), and croplands (85.1%) respectively, Arkansas is covered with forest (28.7%), croplands (25.5%) and woody savannah (19%) [62]. Arizona has a high mean elevation (above mean sea level) of 1283.1 m while Oklahoma, Iowa and Arkansas have mean elevations of 398.3 m, 323.5 m, and 166.2 m respectively [63]. In Figure 2, we show daily examples of the 33 km and 3 km soil moisture. Arizona with the "arid-steppe-hot" climate generally

have dry conditions compared with the other three states (see Figure 2(a) and 2(b)). Even with low soil moisture values, high levels of heterogeneity can still be seen at the 3 km resolution, especially for areas with latitude lower than 35°N. Oklahoma, Iowa and Arkansas have soil moisture values ranging from 0.0 cm^3/cm^3 to 0.5 cm^3/cm^3. While the high resolution soil moisture has more variability at Arkansas (Figure 2(g) and 2(h)), Oklahoma and Iowa are spatially homogeneous at 3 km (Figure 2(c), 2(d) and Figure 2(e), 2(f)).
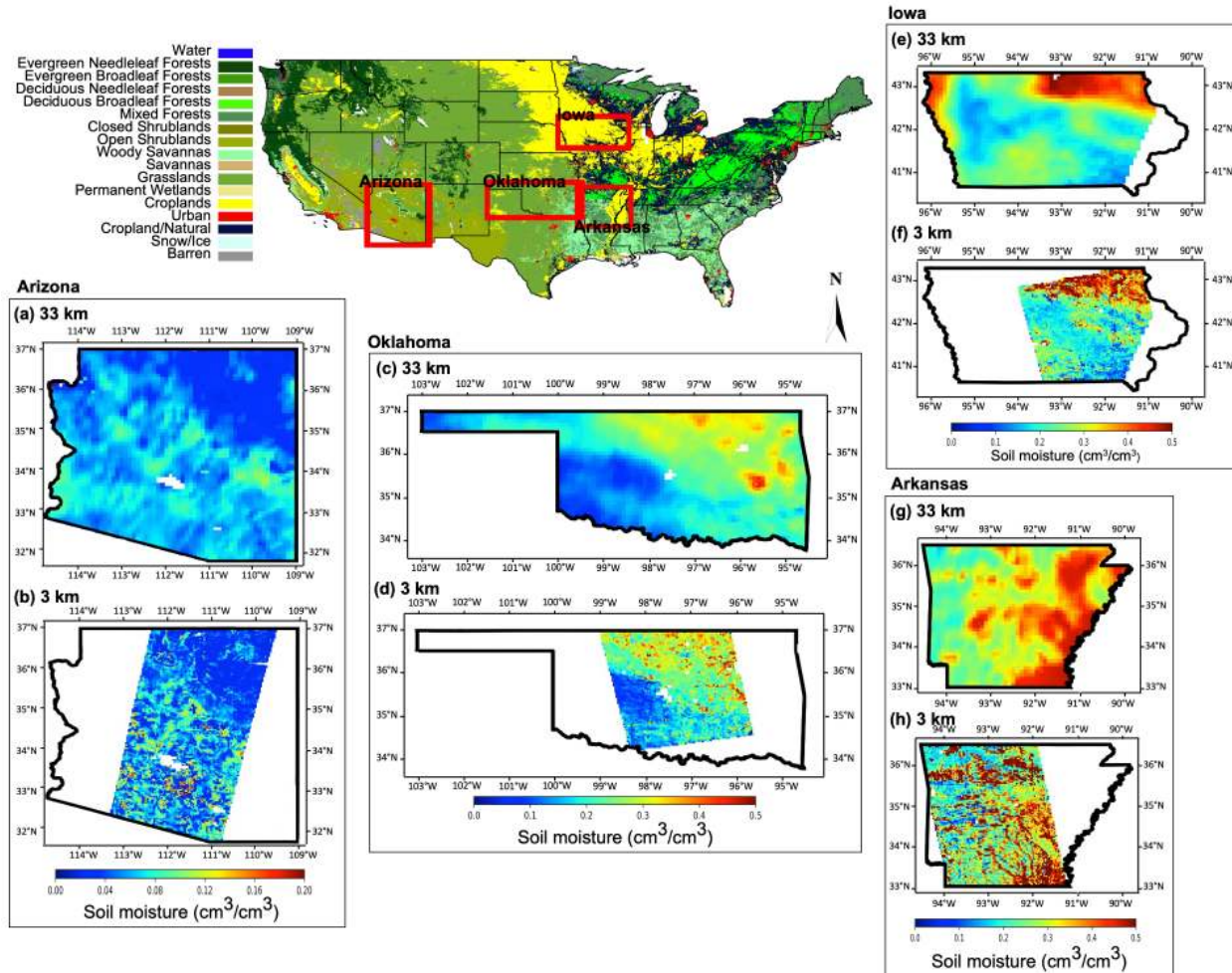

Figure 2: Locations of the four study areas. The background shows the MODIS IGBP land cover class. (a) 33 km soil moisture and (b) 3km soil moisture at Arizona on June 13th, 2018. (c) 33 km soil moisture and (d) 3 km soil moisture at Oklahoma on June 1st, 2018. (e) 33 km soil moisture and (f) 3 km soil moisture at Iowa on June 9th, 2018. (g) 33 km soil moisture and (h) 3 km soil moisture at Arkansas on June 3rd, 2018. 33 km soil moisture is converted from 33 km brightness temperature through the SCA algorithm while 3 km soil moisture is the downscaled product from the SMAP-Sentinel active-passive algorithm.

3   The SMAP Active-Passive Retrieval Algorithm

We provide here an overview of the retrieval algorithm adopted by the SMAP/Sentinel 3 km soil moisture product [27, 67, 36] that motivates the design of our proposed two-layer machine learning-based framework. This retrieval algorithm adopts a two-layer strategy where coarse resolution SMAP radiometer brightness temperature is first downscaled using overlapping high resolution Sentinel radar backscatter data followed by soil moisture retrieval through the SCA (Figure 3(a)).

The algorithm first disaggregates the coarse-scale brightness temperature measurements (T_B, at polarization v or h) with the aid of fine-scale radar backscatter measurements. Both co-pol (σ_{vv} or σ_{hh}) and cross-pol (σ_{vh}) radar backscatter cross-section (in normalized/linear scale) are used in the algorithm. Cross-pol measurements are included in the
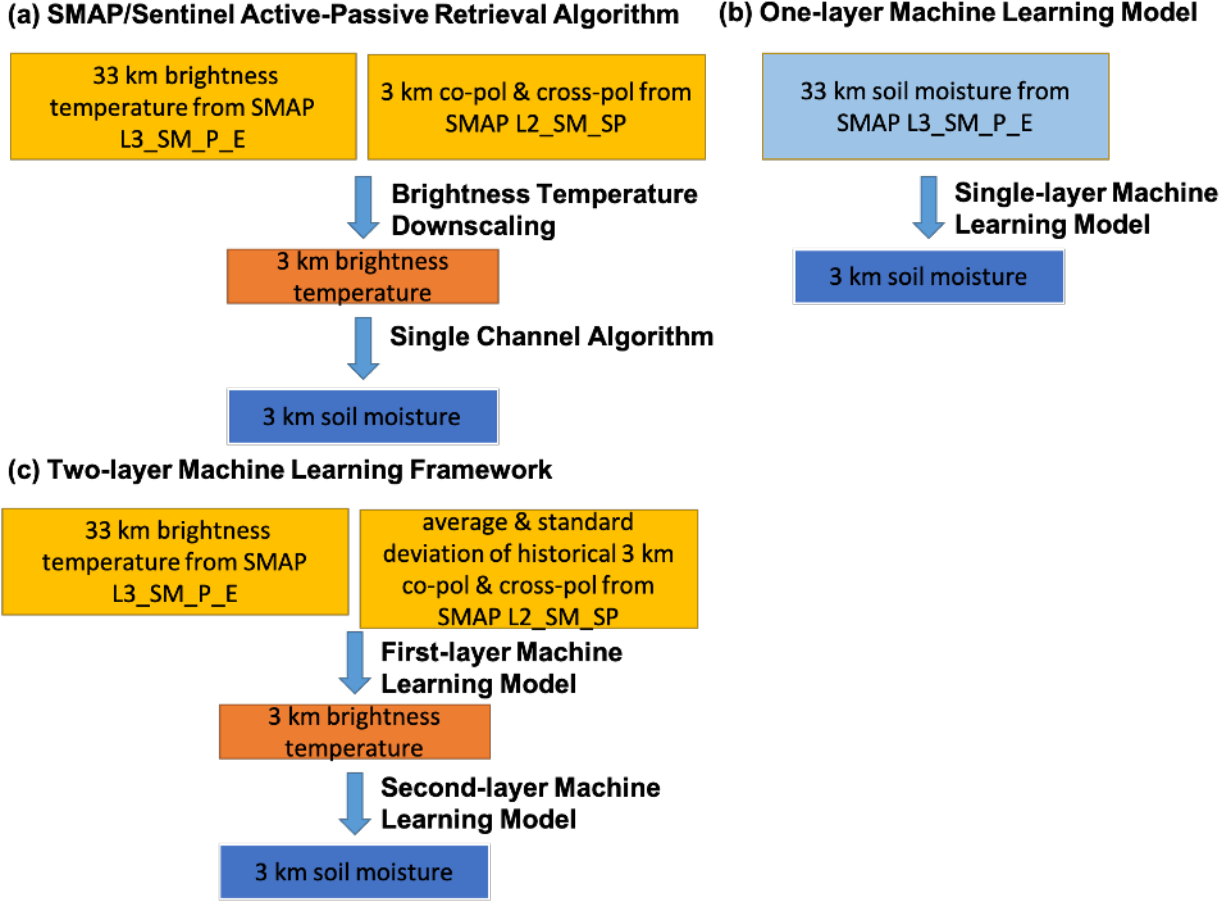
Figure 3: Architecture comparison of (a) the SMAP active-passive retrieval algorithm, (b) the one-layer machine learning model, and (c) the two-layer machine learning framework. Only key inputs are listed here.

algorithm to capture the subgrid variations of brightness temperature within a coarse-scale footprint due to the variations in vegetation characteristics and surface roughness. The vegetation characteristics and surface roughness are assumed to be invariant within a short period of time. Suppose that we have brightness temperature measurements at a coarse support "C" and radar backscatter measurements at a fine support "M". Within a single pixel of the grid at support C, the disaggregated brightness temperature of a sub-pixel at support M can be obtained using

$$T_B(M) = [\frac{T_B(C)}{T_S} + \beta(C) \cdot \{[\sigma_{vv}(M) - \sigma_{vv}(C)] + \Gamma \cdot [\sigma_{vh}(C) - \sigma_{vh}(M)]\}] \cdot T_S. \tag{1}$$

T_S is the effective surface temperature of the near-surface soil (top ∼5 cm). Parameter β is derived from retrieved brightness temperature, radar backscatter, and other physically-based parameters, such as scattering albedo, and vegetation opacity, to estimate the covariation of the emission and backscatter observations. The sensitivity parameter, $\Gamma \equiv [\delta\sigma_{vv}(M)/\delta\sigma_{vh}(M)]$, is used to convert the subgrid heterogeneity captured by the cross-pol deviations $[\sigma_{vh}(C) - \sigma_{vh}(M)]$ to co-pol deviations. Once the downscaled brightness temperature is obtained, the SCA [28] is applied to retrieve the downscaled soil moisture.

Table 1 lists the inputs required for downscaling brightness temperature and SCA separately. For some of the derived inputs like $\sigma_{vv}(M)$, we also list their corresponding data sources. Physically-based parameters like scattering albedo and vegetation opacity depend on vegetation structures, polarization, microwave frequency, etc. They are currently obtained through a land cover-driven lookup table that contains parameter values under different land cover classes [68]. It can be seen from the table that the SMAP/Sentinel active-passive retrieval algorithm, especially the brightness temperature downscaling, requires concurrent radar and radiometer observations on the same day. The narrow imaging swath and low revisit schedule of the Sentinel radars thus severely restricts the spatial and temporal coverage of the resulting downscaled soil moisture estimates. In the next section, we propose a machine learning framework to account for this shortcoming.

7

Table 1: Inputs of the SMAP/Sentinel Active-Passive Retrieval Algorithm.

Brightness Temperature Downscaling	$T_B(C), T_S,$ $\sigma_{vv}(M), \sigma_{vv}(C) \leftarrow \sigma_{vv}(F),$ $\sigma_{vh}(M), \sigma_{vh}(C) \leftarrow \sigma_{vh}(F),$ $\beta(C) \leftarrow T_B(C), T_S, \sigma_{vv}(M), \sigma_{vh}(M),$ scattering albedo, vegetation opacity, incidence angle $\Gamma(C) \leftarrow \sigma_{vv}(M), \sigma_{vh}(M)$
Single Channel Algorithm	$T_B, T_S,$ vegetation water content, vegetation opacity sand & clay percentages, bulk density scattering albedo, roughness coefficient land cover class

4 Methodology

4.1 Two-layer machine learning Framework

To downscale soil moisture that has a high consistency with the SMAP/Sentinel-1 3 km soil moisture estimates at gap areas, we hypothesize that it would be beneficial to learn a two-layered machine learning framework whose structure and inputs design are guided by the SMAP/Sentinel active-passive retrieval algorithm.

First, we formulate the relationship learned by the first-layer machine learning model to disaggregate the 33 km brightness temperature to 3 km. Following the SMAP/Sentinel active-passive algorithm, we rewrite equation (1) as

$$\frac{T_B(3km) - T_B(33km)}{T_S} = \beta \cdot \{[\sigma_{vv}(3km) - \sigma_{vv}(33km)] + \Gamma \cdot [\sigma_{vh}(33km) - \sigma_{vh}(3km)]\}. \tag{2}$$

The left side of equation (2) can serve as the learning target of our first-layer machine learning model. It is available at areas where the radar and radiometer observations are available and missing at the gap areas where the radar observations are absent.

The parameters on the right side of equation (2) are restricted by the availability of radar observations and therefore incur gap areas for the SMAP/Sentinel-1 3 km soil moisture product. To remove this dependence, we define the function f_1 that will be learned by our first-layer machine learning model as

$$y = \frac{T_B(3km) - T_B(33km)}{T_S} = f_1(rb^{hist}, \beta^{hist}, \Gamma^{hist}, \omega, \tau, lat, lon). \tag{3}$$

ω and τ refer to the scattering albedo and vegetation opacity respectively and can be obtained from the SMAP enhanced passive soil moisture product. They are thus available in both covered and gap areas. Latitude and longitude were added to this framework, and all following frameworks, to aid the machine learning models to learn more localized patterns. Instead of directly using the radar observations $\sigma_{vv}(3km)$ and $\sigma_{vh}(3km)$ that are not available at gap areas, we use the average of historical radar backscatter measurements rb^{hist} within a 30-day window as substitutes to capture the subgrid heterogeneity in vegetation and surface physical characteristics. This substitution is based on the knowledge that radar backscatter measurements can capture the subgrid land-surface heterogeneity and can be assumed to have less variability, compared with the dynamics of soil moisture, within a finite temporal window [27, 67]. To the best of our knowledge, this is the first study utilizing the average of historical radar backscatter measurements for downscaling/gap-filling soil moisture. In addition to the average, we also include the standard deviation of the historical 3 km radar backscatter measurements as an input to reflect the extent to which they change during the past 30 days. 30-day historical averages of β (β^{hist}) and Γ (Γ^{hist}) are also introduced to improve the efficacy of the models.

Using these ancillary data, the first-layer machine learning model will be trained at areas where the learning target and inputs are available. The learned model can then be used to obtain the predicted $\hat{y}$ at areas where $T_B(3km)$ is missing. We use the "hat" symbol over a variable to denote that it is a predicted value. The 3 km brightness temperature at these gap areas can, therefore, be reconstructed through

$$\hat{T}_B(3km) = T_B(33km) + \hat{y} \cdot T_S. \tag{4}$$

After downscaling the 33 km brightness temperature, the downscaled 3 km soil moisture is obtained through the second-layer machine learning model f_2 given by

$$\widehat{SM}(3km) = f_2(\hat{T}_B(3km), T_S, VWC, S, \omega, \tau, h, L, lat, lon). \tag{5}$$

8

VWC, S, h, and L refer to vegetation water content, soil properties including sand & clay percentages and bulk density, roughness coefficient and land cover class respectively. These features, along with the predicted 3 km brightness temperature, effective surface temperature, scattering albedo, and vegetation opacity, also form inputs to the original SCA algorithm. We call this machine learning framework as "two-layer".

4.2 Traditional one-layer machine learning framework

In addition to the two-layer machine learning-based framework, we also implement a one-layer framework which directly downscales soil moisture from 33 km to 3 km using ancillary data in Section 2.3. It also has the ability to gap-fill soil moisture at missing areas as long as the ancillary data is available. The function f_3 learned by this framework can be written as:

$$\widehat{SM}(3km) = f_3(SM(33km), \mathcal{H}, lat, lon), \tag{6}$$

where $\mathcal{H}$ denotes the feature set including leaf area index, land surface temperature, soil properties, elevation, slope, land cover, and precipitation and have been shown to be the major drivers of soil-moisture variability in past studies [69, 70, 71, 72, 73]. In this paper, we call this framework as "one-layer".

Figure 3(b) and 3(c) shows the structures of the one-layer framework and the proposed two-layer framework respectively. Comparing Figure 3(a) and Figure 3(c), it can be seen that the two-layer machine learning-based framework follows a similar structure as the original SMAP/Sentinel active-passive retrieval algorithm while using a machine learning paradigm. Its performance over the one-layer framework in terms of downscaling soil moisture at gap areas will be demonstrated through extensive machine learning-based experiments discussed in Section 5.

4.3 Additional machine learning frameworks for comparison

For a comprehensive evaluation, we also compare the proposed "two-layer" with the following machine-learning frameworks:

- Two-layer+ : "Two-layer" framework but with ancillary data from both Sections 2.2 and 2.3 as inputs.
- One-layer + (SM): "One-layer" framework but with ancillary data from both Sections 2.2 and 2.3 as inputs.
- One-layer + (BT) : "One-layer" framework which converts 33 km brightness temperature (instead of 33 km soil moisture) directly into 3 km soil moisture and with ancillary data from both Sections 2.2 and 2.3 as inputs.

The "two-layer+" and "one-layer+ (SM)" allows us to evaluate the effect of adding "extra" covariate information in the machine-learning frameworks. The "one-layer + (BT)" is also evaluated to demonstrate the potential advantage of using a two-layer framework as opposed to directly downscaling brightness temperature to 3 km soil moisture.

4.4 Random Forest

In this paper, for all the machine-learning frameworks, the Random Forest [53] will be used as the "machine-learning model" for downscaling soil moisture/brightness temperature (i.e. the model used in f_1, f_2 and f_3 in equations 3, 5 and 6 respectively). Proposed in 2001, the Random Forest model quickly gained its reputation for being able to learn complex nonlinear relationships between the target variable and input features while resisting to the overfitting problem. Its superior performance is achieved through an ensemble of individually learned decision trees combined with the bagging method and random selections of features. During the training process of the Random Forest model, multiple hierarchical decision trees are learned independently with the goal of maximizing the variance reduction for each split. The individually grown decision trees can, however, easily suffer from the overfitting problem where significantly disparate performances can be observed on the training and test sets. To tackle this issue, the Random Forest model averages the predictions generated by all trees in the forest with additional strategies to reduce the correlation between them. The correlation reduction is done through bagging and the random selections of features. During the training process, each decision tree is exposed to different data subsets that are randomly sampled with replacement from the same training set. Additionally, randomly selected subset of the features are considered for each split. The Random Forest model generalizes well on test set when strong decision trees are constructed with controlled correlations between them.

4.5 Spatio-Temporal Regional Learning

For a machine-learning framework, we test the following three hypotheses in this paper: high resolution soil moisture/brightness temperature at gap areas can be reconstructed from the models learned at (1) spatial neighboring regions, (2) temporal neighboring regions, and (3) both spatial and temporal neighboring regions. We define these three settings

as the spatial regional learning, the temporal regional learning, and the spatio-temporal regional learning respectively. For the temporal regional learning and the spatio-temporal regional learning, a 30-day window is adopted to search for the most recent temporal neighboring regions for the gap regions. As shown in Figure 4, machine learning models are built using data available at covered spatial and/or temporal neighboring regions (inside the orange frames) and are subsequently used to predict high-resolution soil moisture at the gap region (inside the red frame) where the original SMAP/Sentinel-1 active-passive soil moisture product is missing. In this paper, separate machine learning models are trained in this paper for one specific region on different days to dynamically capture the relationship between downscaled soil moisture/brightness temperature and designed inputs. The success of these machine learning models depends on how effectively information learned at spatial/temporal neighboring regions can be extrapolated to the gap region.
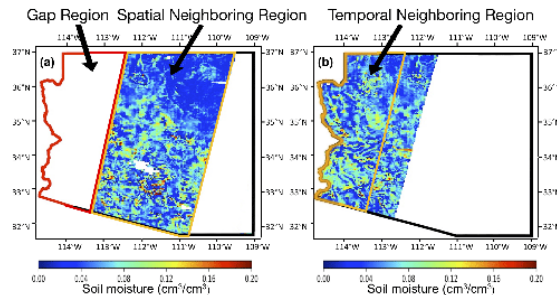

Figure 4: 3 km soil moisture swath at Arizona on (a) June 13, 2018 and (b) June 5th, 2018 from SMAP L2_SM_SP.

The three regional learning settings enable expanded training data from spatio-temporally neighboring data regions. The expanded training data enabled by the regional learning settings allows the models to learn complex relationships between coarse-scale and fine-scale soil moisture and achieve higher predictive capabilities. Note that we focus on gap-filling the "neighboring" spatio-temporal regions in this paper which fall into the same, or, at least, similar hydroclimates as the regions where the models are learned. Building models that can be applied to "almost everywhere" or discussing whether a model learned under one hydroclimate is applicable to another is beyond the scope of this paper and will be explored in the future.

5 Results and Discussions

We validate the downscaled soil moisture at four selected study areas, Arizona, Oklahoma, Iowa, and Arkansas, for the three and a half year period between April 1st, 2015 and September 30th, 2018. Besides the one-layer framework (equation 6), we also compare the performance of our proposed two-layer framework with the 33 km soil moisture from the SMAP enhanced passive soil moisture product. As the SMAP enhanced passive soil moisture product is posted on 9 km EASE2 grids, we resample the 33 km soil moisture to 3 km by simply assuming that all the 3 km EASE2 subgrids lying inside one 9 km EASE2 grid have the same value. We denote this product as the "resampled 33 km soil moisture". The resampled 33 km soil moisture estimates serve as a reference for comparison with the one-layer and the proposed two-layer frameworks. In terms of metrics, the correlation coefficient (R) and unbiased root mean square error (ubRMSE) are used to quantify and compare the overall agreement between the predicted 3 km soil moisture and the retrieved ones. Average metrics from five machine learning runs are reported for all the experiments.

5.1 33 km versus one-layer versus two-layer

To demonstrate the performance of our proposed two-layer framework, we split the covered regions (where the 3 km soil moisture estimates are available) into training and test regions. Machine learning models learned at the training region are used to predict the downscaled soil moisture/brightness temperature at the test region. At each of the selected study areas, Arizona, Oklahoma, Iowa and Arkansas, there are $2 \sim 3$ different major swath patterns with $\sim$12 day interleaved revisit schedules. As shown in Figure 5, we select two swath patterns for each of the study areas, the main swath A and the shadow swath B. The overlap of the two swaths will be used as a filter to obtain the hold-out test region. For any specific day which has the main swath A available, the filter will be applied to obtain the test region (as shown in the red frames of Figure 5 (a), (c), (e) and (g)), while the rest will be used as spatial neighboring regions for the training process (as shown in the orange frames in Figure 5 (a), (c), (e) and (g)). For the temporal regional learning task, a 30-day sliding window is adopted to search for the most recent shadow swath B. Once the most recent shadow swath B is obtained, the filter of the test region is applied to obtain the temporal neighboring region (as shown in the orange frames of Figure 5 (b), (d), (f) and (h)).
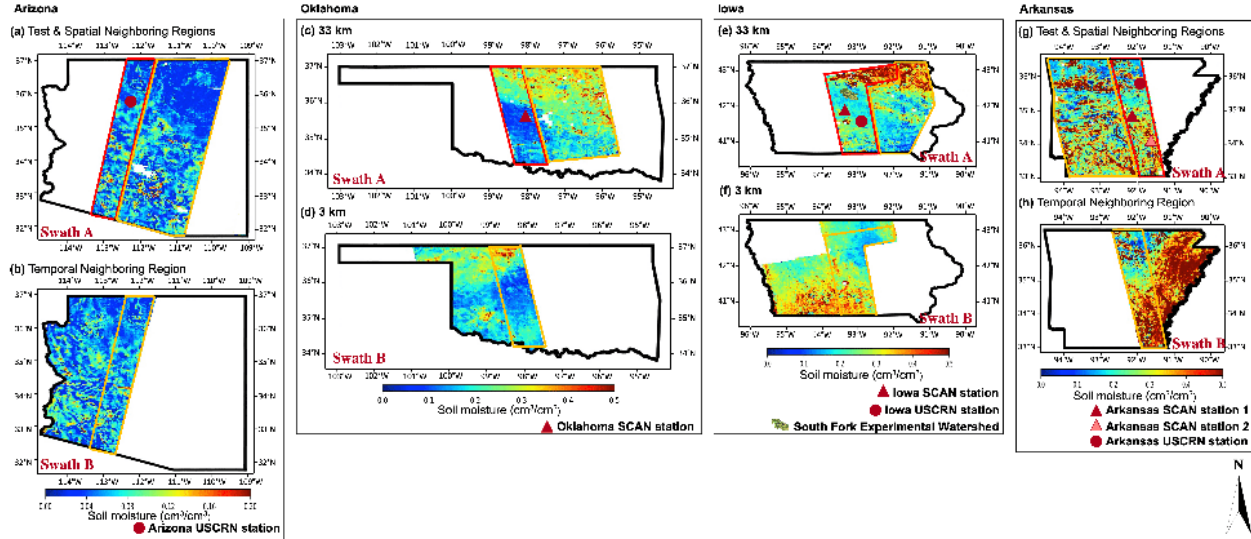
Figure 5: Train and test splits at Arizona, Oklahoma, Iowa and Arkansas. (a) selected swath A at Arizona on June 13th, 2018. (b) selected swath B at Arizona on June 5th, 2018. (c) selected swath A at Oklahoma on June 1st, 2018. (d) selected swath B at Oklahoma on May 24th, 2018. (e) selected swath A at Iowa on June 9th, 2018. (f) selected swath B at Iowa on June 3rd, 2018. (g) selected swath A at Arkansas on July 8th, 2018. (h) selected swath B at Arkansas on June 21th, 2018.

This training and test splitting strategy mimics how the regional learning is applied to fill/downscale soil moisture at real gap areas where the 3 km soil moisture estimates are indeed missing. It also allows us to compare the performance of the three regional learning settings, the spatial regional learning, temporal regional learning, and the spatial-temporal regional learning. Figure 6 shows, as an example, the test region at Arizona on June 13 and its corresponding spatial neighboring region and temporal neighboring region for the training. Table 2 lists the number of days covered for each quarter of a year at the selected study areas where the main swath A is available for test and spatial neighboring regions and the shadow swath B is available for a temporal neighboring regions search. It can be seen from the table that Arizona, Iowa, and Arkansas have no coverage in 2015 and the first three quarters of 2016 which is due to the limited/no coverage of the SMAP/Sentinel-1 3 km soil moisture product during that period. The average sizes (number of pixels available) of spatial training and test regions and their corresponding ratio for each area are also listed in the table. The sizes of temporal training regions are the same as the sizes of the test sets.

Table 2: Number of Days Covered for Each Quarter of a Year (April 1st, 2015 - September 30th, 2018) & Average Sizes of Spatial Training Regions and Test Regions.

Year	2015			2016				2017				2018			Training size	Test size	Ratio
Quarter	2	3	4	1	2	3	4	1	2	3	4	1	2	3			
Arizona	-	-	-	-	-	-	-	2	4	6	7	6	7	3	11463	3676	3.1
Oklahoma	-	2	-	-	4	4	5	1	5	2	3	4	6	8	4306	2330	1.8
Iowa	-	-	-	-	-	6	-	4	5	7	6	2	7	7	7773	2512	3.1
Arkansas	-	-	-	-	-	-	3	7	5	7	5	7	7	7	4782	2215	2.1

Figure 7 and Figure 8 compare the performance (in terms of R and ubRMSE respectively) of the resampled 33 km soil moisture, one-layer framework and our proposed two-layer framework under different regional learning settings at available test regions. At each selected study area, performance metrics of different DOYs are represented by boxplots to show a general pattern of how different frameworks perform. It can be seen that the proposed two-layer framework generally outperforms both the resampled 33 km soil moisture product and the one-layer framework in terms of R and ubRMSE under all three regional learning settings. The improvement in R is significant, especially at Arizona and Arkansas. Taking the spatio-temporal regional learning as an example, the mean R at Arizona is increased to 0.728 for the two-layer framework compared to 0.572 (27% improvement) for the one-layer framework and 0.493 (47% improvement) for the 33 km soil moisture, while the mean R at Arkansas is increased to 0.744 for the two-layer
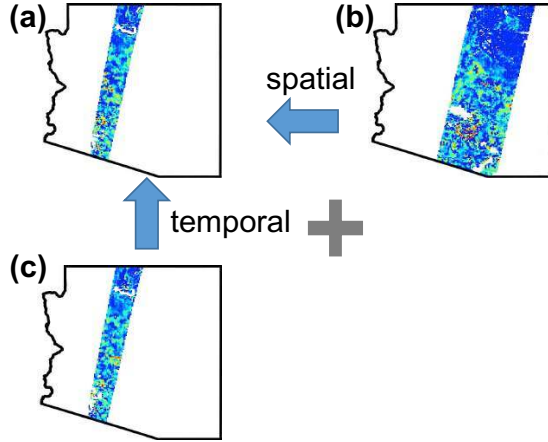
Figure 6: Spatial/temporal regional learning for (a) test area at Arizona on June 13th, 2018. (b) spatial neighboring region on the same day. (c) search for temporal neighboring regions within 30-day sliding window is applied to obtain the most recent temporal neighboring region on June 5th, 2018.

framework compared to 0.651 (14% improvement) for the one-layer framework and 0.413 (80% improvement) for the 33 km soil moisture. Exceptions of improvement in performance are noticed at Oklahoma and Iowa for the spatial regional learning and temporal regional learning where relatively insignificant changes in the chosen metrics are found when comparing the resampled 33 km soil moisture product, the one-layer framework, and the two-layer framework. This is, however, expected as the retrieved 3 km soil moisture has much lesser spatial heterogeneity compared with the 33 km soil moisture product at Oklahoma and Iowa as discussed in Section 2.4.

To visually compare the performance of these frameworks, Figure 9, as an example, focuses on a daily test region on June 13, 2018 at Arizona to show the disaggregated soil moisture produced by the resampled 33 km soil moisture product, the one-layer framework and the proposed two-layer framework. It can be seen that although the one-layer framework under the spatial-temporal regional learning has the ability to reconstruct fine-scale soil moisture to some extent (Figure 9(c)), its downscaling quality is still inferior to the proposed two-layer framework, even to the one under the spatial regional learning only (Figure 9(d)). With the proposed two-layer framework, more fine-scale soil moisture heterogeneities can be predicted/downscaled with improvement in both R and ubRMSE, especially under the spatial-temporal regional learning setting (Figure 9(f)). Figure 10 shows the pixel-by-pixel comparison of the correlation of the resampled 33 km, one-layer framework and the proposed two-layer framework under the spatial-temporal regional learning respectively with the SMAP/Sentinel-1 observations. The correlation values are calculated for all available values inside each 33 km × 33 km pixel of the test region. As the SMAP 33 km enhanced passive soil moisture product is actually posted on a 9 km resolution grid, it has non-zero correlation values inside the 33 km × 33 km pixels. Figure 10, again, shows that our two-layer framework increases the correlation between the gap-filled predicted soil moisture and the SMAP/Sentinel-1 3 km observations compared with both the resampled 33 km soil moisture and the one-layer predictions, though as mentioned earlier, the improvement is moderate in Iowa and Oklahoma.

Since the above discussion summarizes the performance of the proposed framework for the entire duration of the data, it is also worthwhile to look at how the framework performs seasonally. To achieve this, we group the results from all years into four quarters (January–March, April–June, July–September, and October–September) and compare R and ubRMSE across these time periods. (Figure 11). It can be seen that the two-layer framework consistently outperforms the 33-km soil moisture and the one-layer framework across seasons. It can also be seen that for the majority of the seasons, the scatter of the boxplot for the two-layer is also less, making the predictions consistently accurate across all days in a season.

5.2   Comparison with additional frameworks

Since in the previous section, we found that the spatio-temporal regional learning outperforms the spatial-only and temporal-only approaches, we now compare the two-layer framework with machine-learning frameworks defined in Section 4.3 under the spatio-temporal regional learning using the same test-train split criterion as Section 5.1. Figure 12 compares the various frameworks in terms of the mean R between the predictions and observation in the test region for all four study areas. We find that for all the study-areas, there is insignificant improvement between the "two-layer" and "two-layer+" framework denoting that adding the extra features from Section 2.3 has no effect in improving soil
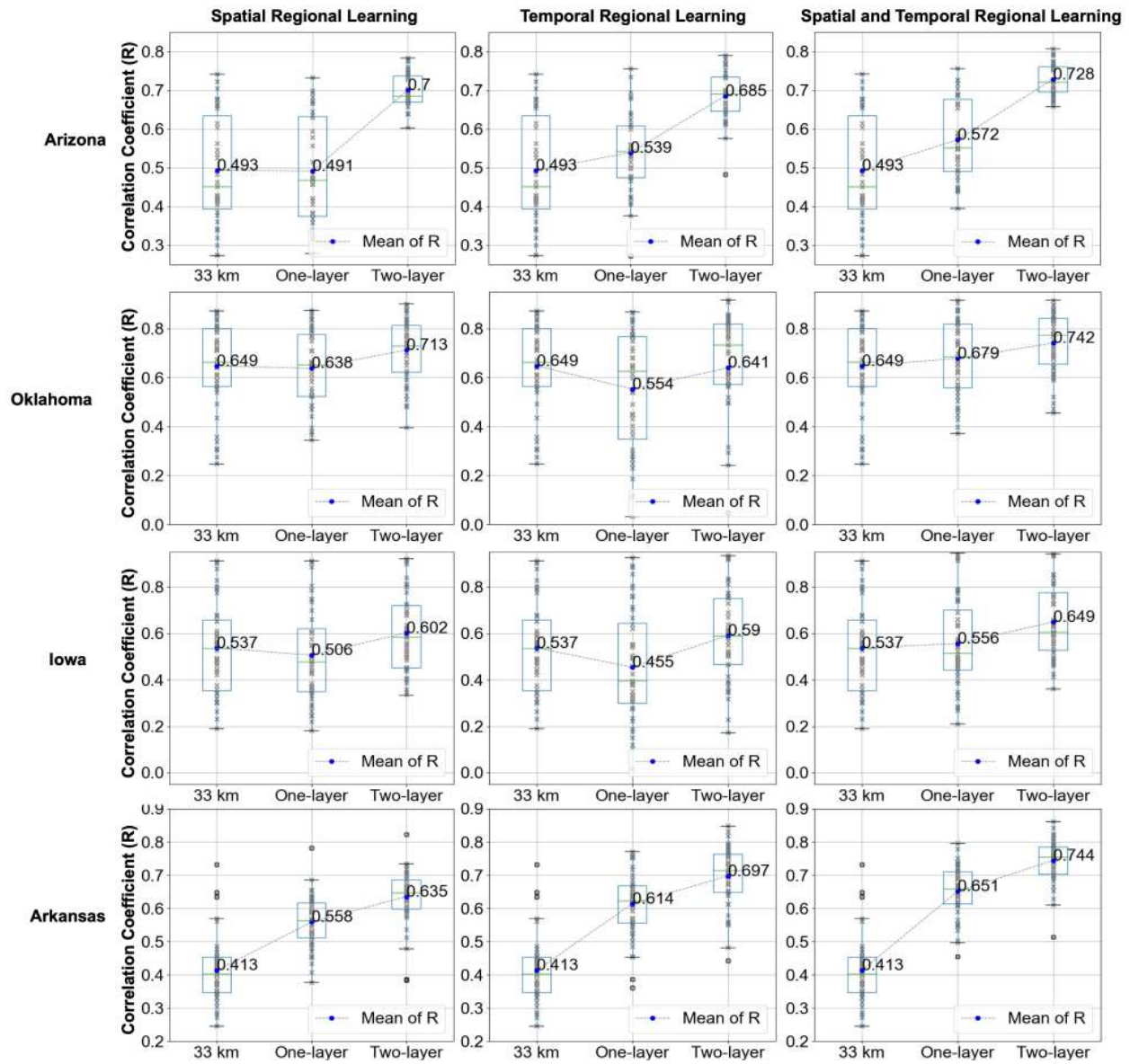
Figure 7: Correlation coefficient (R) comparisons of the resampled 33 km, one-layer model and our proposed two-layer framework for the selected study areas under different regional learning settings.
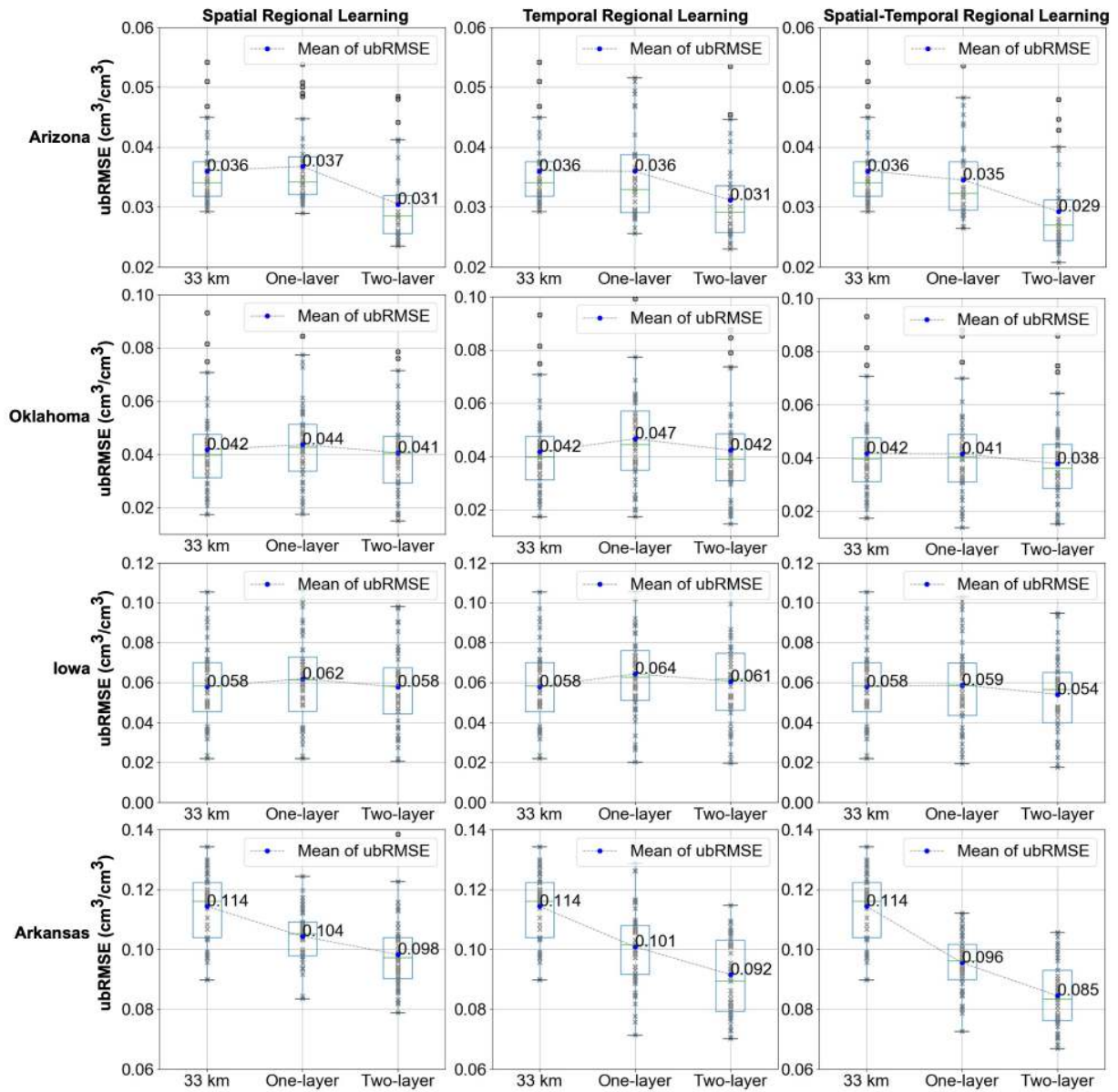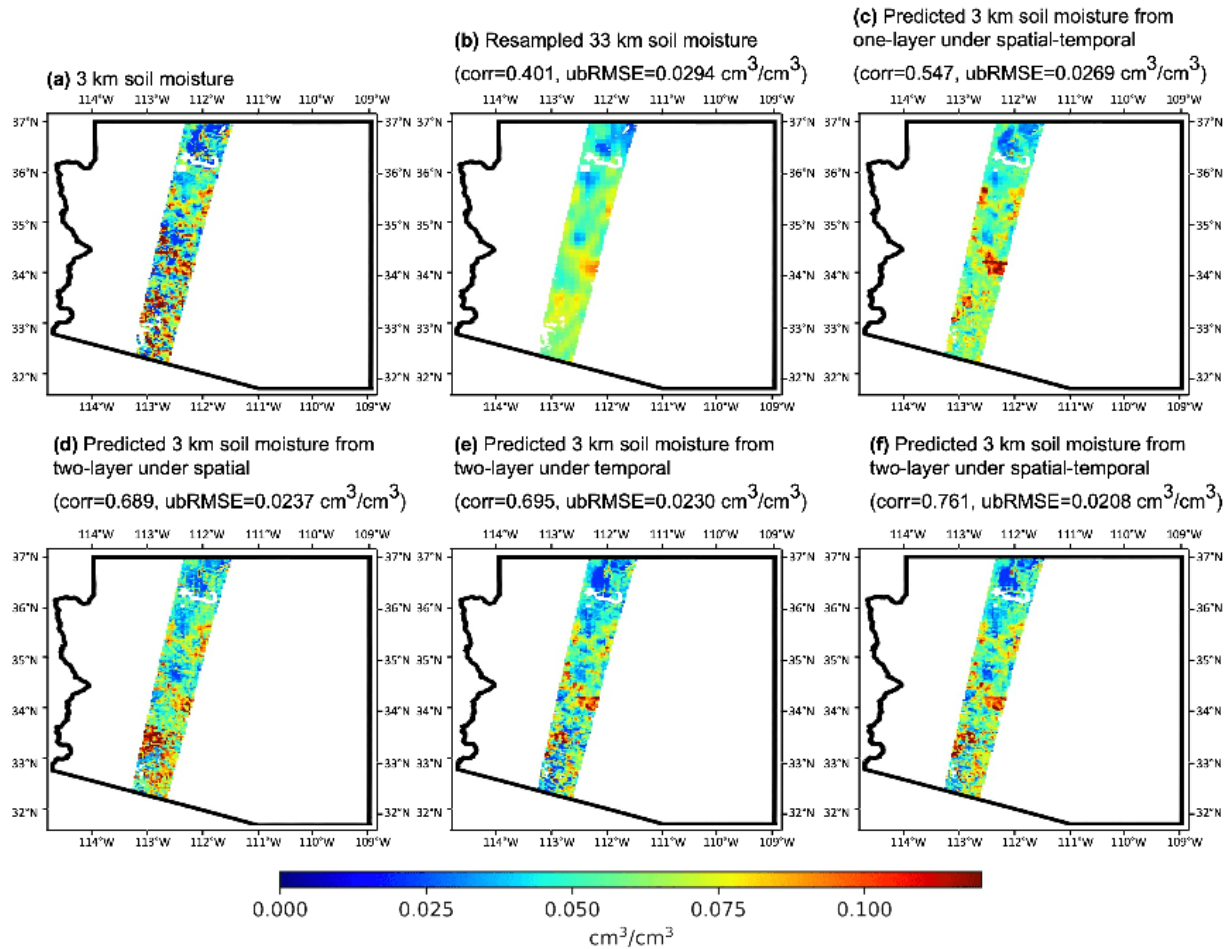
Figure 8: Unbiased RMSE (ubRMSE) comparisons of the resampled 33 km, one-layer model and our proposed two-layer framework for the selected study areas under different regional learning settings.

Figure 9: Performance comparisons for test region at Arizona on June 13th, 2018. (a) 3 km soil moisture from SMAP/Sentinel-1 active-passive Product. (b) resampled 33 km soil moisture from SMAP enhanced passive product. Predicted 3 km soil moisture from (c) the one-layer model under spatial-temporal regional learning, (d) the two-layer framework under spatial regional learning, (e) the two-layer framework under temporal regional learning and (f) the two-layer framework under spatial-temporal regional learning.

moisture predictions. We can also see that the "two-layer" outperforms both the "one-layer+ (SM)" and "one-layer+ (BT)" across all four study regions which underscores the importance of using a two-layer machine learning framework for soil moisture downscaling rather than directly downscaling 33 km soil moisture or 33 km brightness temperature to high-resolution soil moisture. Note that, as mentioned earlier, the Random Forest forms the "machine-learning model" for all the analyzed frameworks.

5.3 Feature importance of input data in two-layer framework

Another advantage of using the Random-Forest model in the two layer-framework is that it has the ability to assess the relative importance of the covariates used as inputs to the model. Note that the sum of the relative feature-importance of all inputs always sum to one [53]. Figure 13 shows the relative feature-importance of the input data (Table 1) in the first layer of the framework. It can be seen that the 30-day historical radar-backscatter data (co-pol hist and cross-pol hist in Figure 13) is the most important input of the first-layer of the framework followed by a moderate effect of the rest of the parameters. This implies that incorporating the historical radar backscatter data from both spatial and temporal regions can significantly improve the prediction capabilities of the machine-learning model to predict downscaled brightness temperature. For the second-layer of the model, as expected, the predicted 3 km brightness temperature from layer-1 is the most important input followed by surface temperature for Arizona and landcover for Arkansas, Iowa and Oklahoma (Figure 14). For Arkansas, the effect of landcover is especially significant because, as mentioned earlier in Section 2.4, Arkansas consists of diverse land-cover classes compared to the other study regions which comprise predominantly
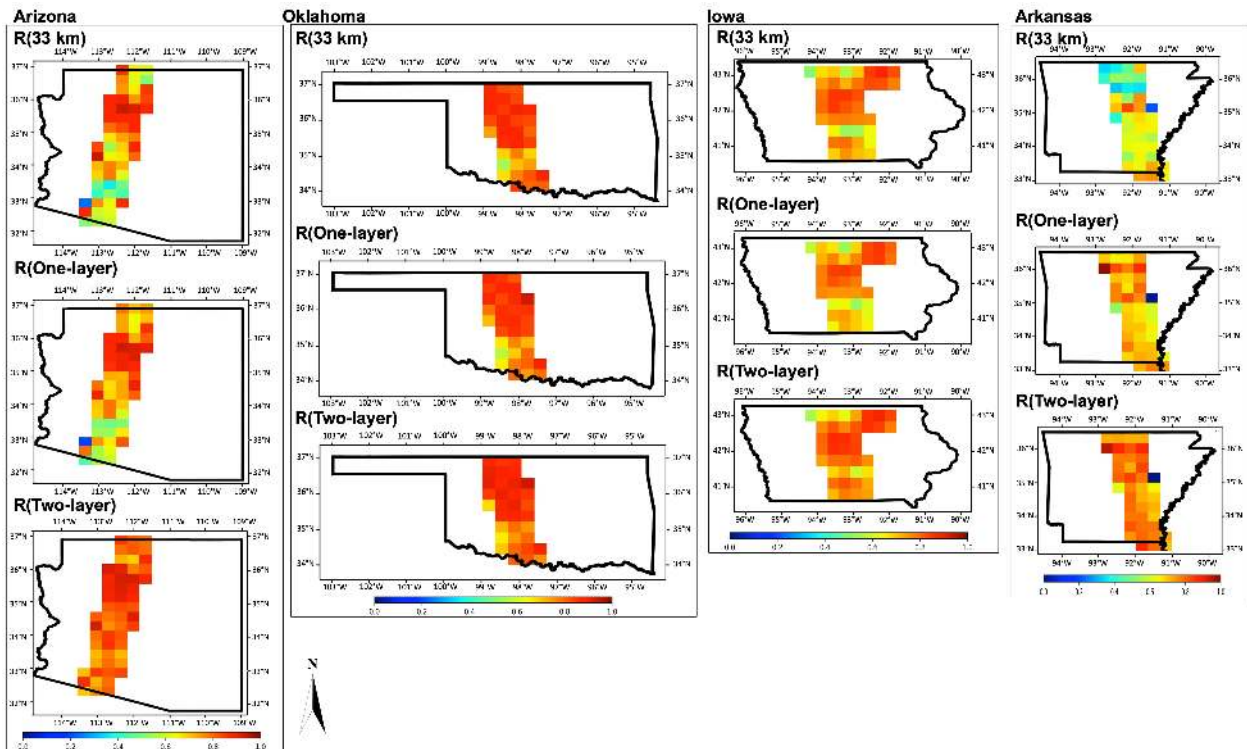
15

Figure 10: Correlation of the resampled 33 km, one-layer model, and our proposed two-layer model with the SMAP/Sentinel-1 observations under the spatial and temporal regional learning. Each pixel in this figure is a patch of 11×11 3 km EASE2 grids.
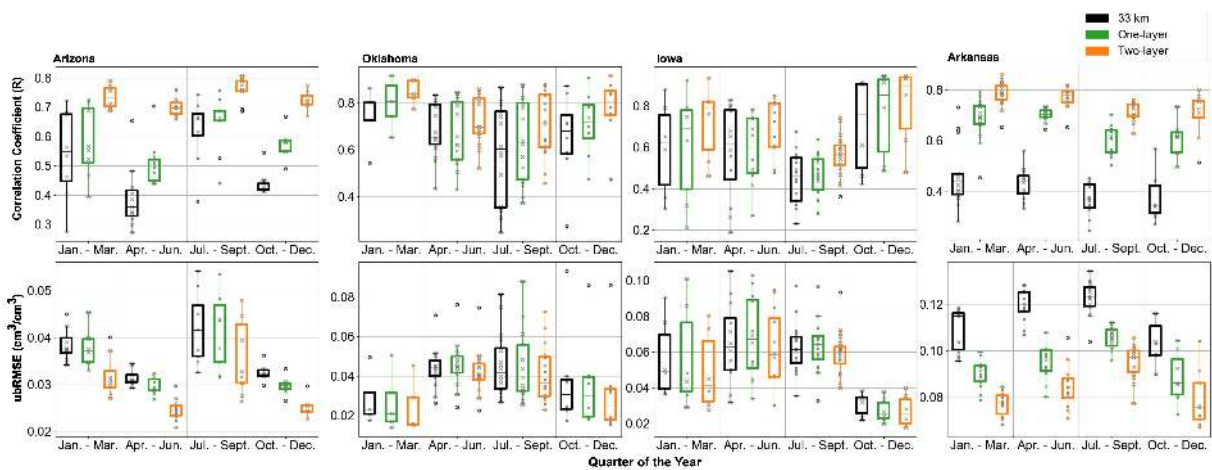

Figure 11: Seasonal performance metrics of the two-layer framework, one-layer framework, and 33-km soil moisture. The two layer framework outperforms the other two consistently across seasons.
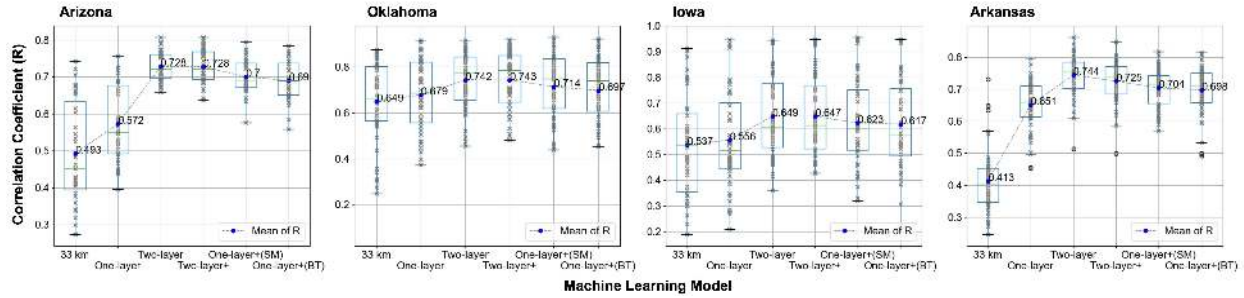
16

Figure 12: Performance comparisons of the our proposed two-layer machine learning models with three other machine learning models with different feature designs. Two-layer+ adds auxiliary environmental variables to the two-layer model. One-layer +(SM) is a one-layer model which uses all the features that are used in the two-layer models but directly downscale soil moisture from 33 km to 3 km. One-layer +(BT) is a one-layer model which directly downscales 33 km brightness temperature to 3 km soil moisture.

(more than 70 %) of one landcover only. It is interesting to note that, for the second-layer, the rest of the inputs are not significant for any study-region.

5.4 Comparison with airborne and in-situ soil moisture data

We also compare the proposed machine-learning framework (under spatio-temporal learning) using publicly available high-resolution (airborne) soil moisture dataset retrieved during the SMAPVEx 2016 campaign at South Fork, Iowa Figure 5(e). Due to the coarse temporal latency of the SMAP-Sentinel-1 product, only one day (August 6th, 2016) of the airborne campaign coincided with the SMAP-Sentinel-1 overpass. Figure 15(a) shows the PALS airborne soil moisture data upscaled to the 3 km EASE2 grid. Figure 15(b) shows the retrieved 3 km soil moisture from the SMAP/Sentinel product and Figure 15(c) shows the soil moisture predictions from the two-layer framework. The correlation of the two-layer predictions with the airborne soil moisture have similar correlation (0.23) and ubRMSE (0.032 cm^3/cm^3) compared to the actual SMAP-Sentinel-1 product which has a correlation of 0.09 and ubRMSE of 0.038 cm^3/cm^3 with airborne data which serves as a good valdidation of the framework.

To explore the temporal-stability of the proposed framework, we further validate the soil-moisture predictions with in-situ data from SCAN and USCRN. Figure 5 lists the available SCAN and USCRN stations falling into the test regions at the four selected areas (one each for Arizona and Oklahoma, two for Iowa and three for Arkansas). Figure 16 shows the time-series of the in-situ data, our soil moisture predictions and the actual SMAP-Sentinel soil moisture values. Note that, for predictions, the entire test region, following Figure 5 was held-out. It can be seen that the predictions from our two-layer machine learning framework mimic the majority of the wetting and drying phases of the observed SMAP-Sentinel-1 product and in-situ data. The performance of the predictions follow a similar trend as that of the actual SMAP-Sentinel-1 product (Table 3.). For instance, compared with USCRN data in Arizona, the correlation of the predictions is 0.71 (ubRMSE = 0.047 cm^3/cm^3) when the correlation of the actual SMAP-Sentinel-1 product is 0.65 (ubRMSE = 0.043 cm^3/cm^3) while for Iowa SCAN data, the correlation of predictions drops down to 0.10 (ubRMSE = 0.126 cm^3/cm^3) when the SMAP-Sentinel-1 soil moisture correlation is 0.04 (ubRMSE = 0.125 cm^3/cm^3). This is a good validation of the proposed framework as our aim is to extend the established SMAP/Sentinel-1 product to unobserved areas (and not improve the SMAP-Seninel-1 retrieval). The efficacy of the predictions for a given surface heterogeneity and hydroclimate is therefore limited by the accuracy of SMAP-Sentinel-1 retrievals in those conditions and future improvements in the retrieval algorithm will enhance the predictive capability of the machine learning algorithm as well.

5.5 Regional Learning Limitations

5.5.1 Spatial Regional Learning

To compare the downscaling performances of different regional learning settings, fixed-size test regions were adopted for the above experiments at each study area. It is worthwhile to explore whether our proposed two-layer framework can achieve satisfactory performance when the size of the test regions expands. To project how our models will perform when we fill real gaps, we split the training and test regions by shrinking the regions covered by the 3 km soil moisture. Taking the available swath at Arkansas on June 14th, 2018 as an example, as shown in Figure 17, we split the region covered by the SMAP/Sentinel-1 3 km soil moisture product (Figure 17 (a)) into training and test regions with various
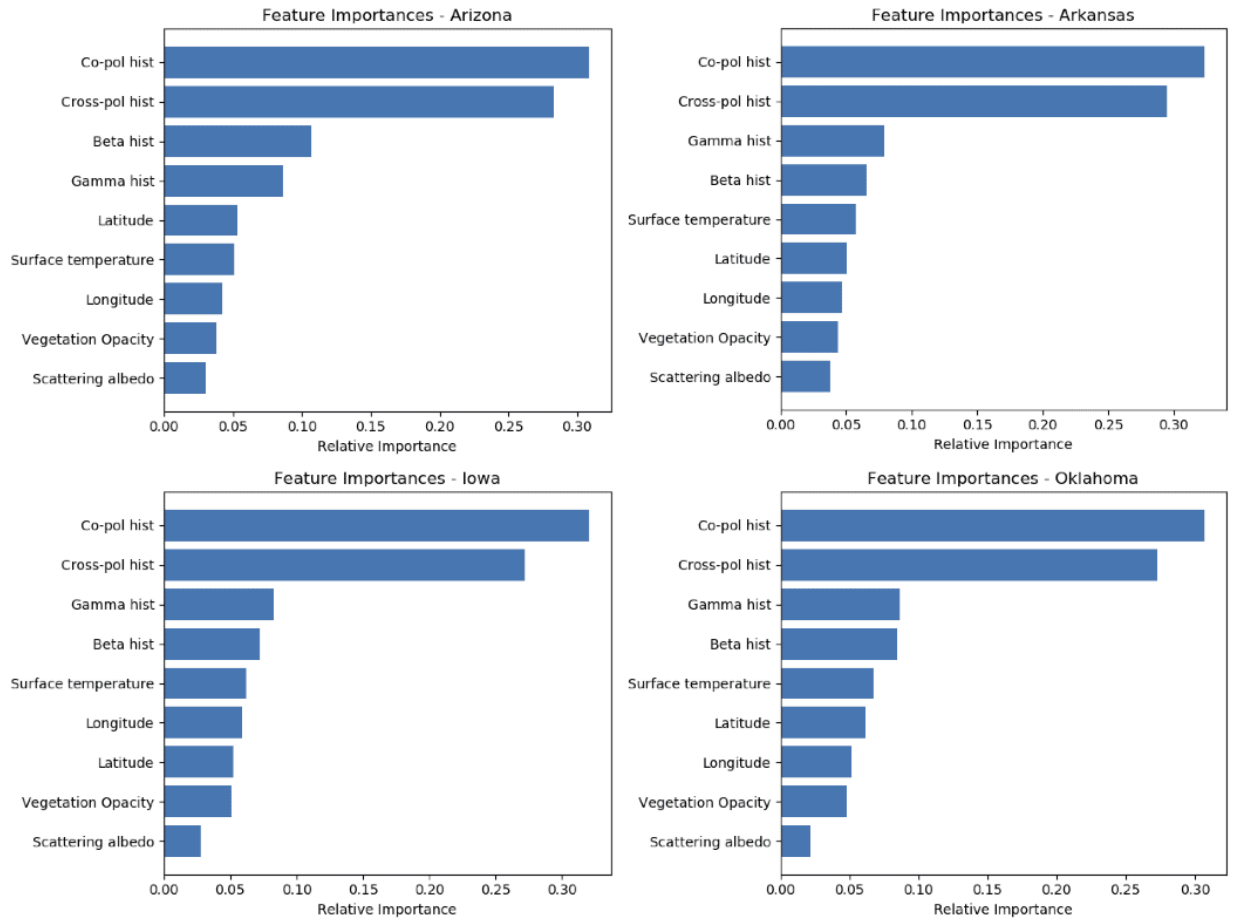
Figure 13: The 30-day average historical radar backscatter data (co-pol hist and cross-pol hist) is the most important feature for layer-1 of the machine learning framework with moderate importance of the rest of the features.

Table 3: Comparison of the SMAP-Sentinel-1 observations and the two-layer predictions with in-situ soil moisture.

| R/ubRMSE(cm^3/cm^3)		SMAP-Sentinel-1 v.s. in-situ	Two-layer predictions v.s. in-situ	Two-layer predictions v.s. SMAP-Sentinel-1
SCAN	Oklahoma	0.55/0.071	0.61/0.068	0.81/0.042
	Iowa	0.04/0.125	0.10/0.126	0.92/0.037
	Arkansas-1	0.65/0.069	0.47/0.085	0.76/0.053
	Arkansas-2	0.79/0.056	0.68/0.045	0.49/0.075
USCRN	Arizona	0.65/0.047	0.71/0.043	0.93/0.018
	Iowa	0.06/0.077	0.08/0.079	0.38/0.060
	Arkansas	0.52/0.056	0.42/0.062	0.76/0.032

18
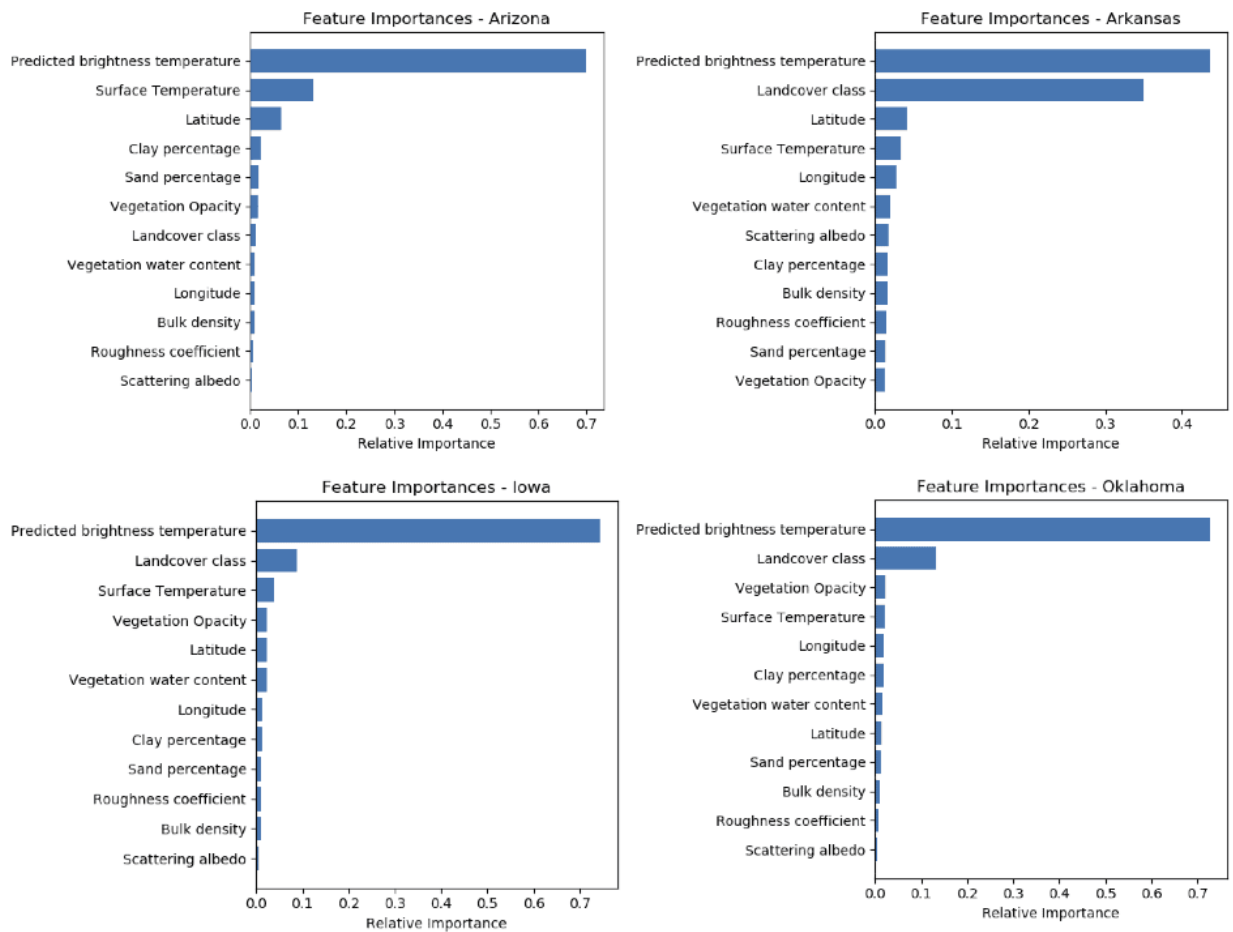
Figure 14: For layer-2 of the machine learning framework, the predicted 3 km brightness temperature from layer-1 is the most important input followed by surface temperature for Arizona and landcover for Arkansas, Iowa and Oklahoma. For Arkansas, the effect of land-cover is more significant due to the presence of multiple land-classes in the state.
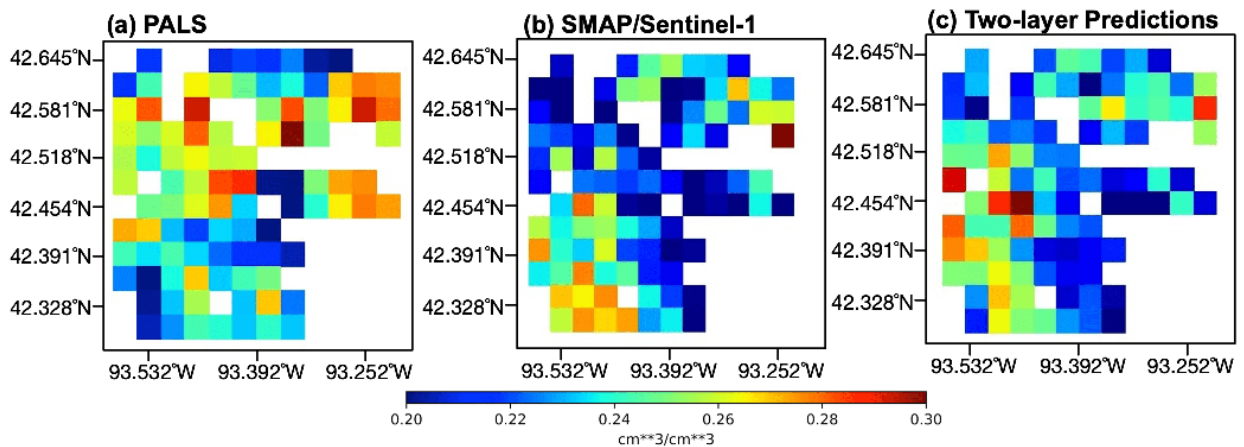

Figure 15: Soil moisture comparisons at South Fork, Iowa on August 6th, 2016. (a) airborne data upscaled to 3 km EASE2 grid. (b) 3 km soil moisture from SMAP/Sentinel-1 active-passive product. (c) predictions from two-layer model under spatial-temporal regional learning.
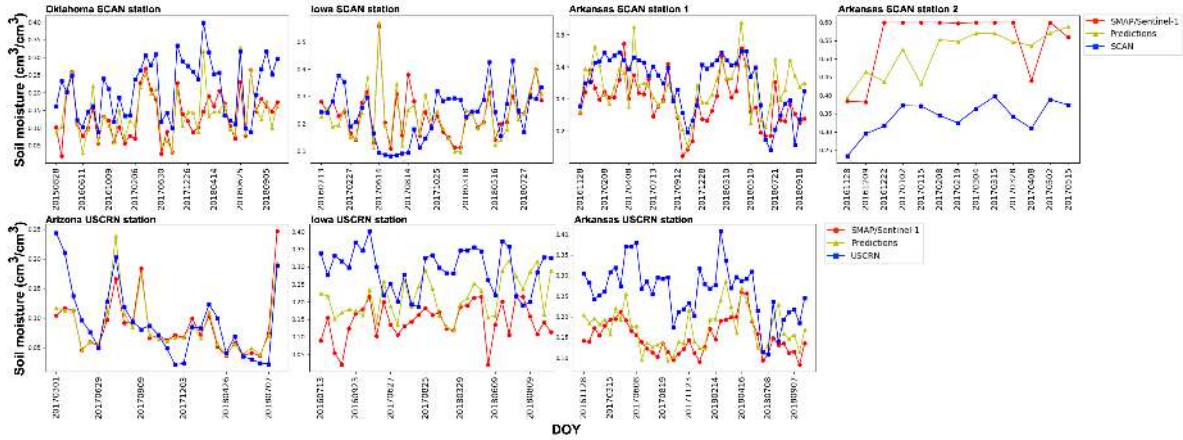
Figure 16: Time series matchup of the in-situ soil moisture collected by two sparse soil moisture networks, SCAN and USCRN, with the SMAP/Sentinel-1 retrievals and predictions from two-layer machine learning framework under the spatial-temporal regional learning.

size ratio (Figure 17 (b), (c), (d) and (e)). Here, four different size ratios, 1: 4, 1: 2, 1: 1 and 2: 1, are chosen. A higher ratio indicates that less training samples are used during the training process and more samples are used to validate predictions. Models learned using data from the training region are used to downscale soil moisture at test region. Figure 18 shows the corresponding results at the three selected study areas. While generally deteriorating performances are observed with the growing size ratio, at Arizona and Arkansas our proposed two-layer framework outperforms the resampled 33 km soil moisture and the one-layer framework. Especially at Arkansas, the R values are approximately 0.7 when the test region is twice the size of the training region compared with ∼ 0.4 R from the 33 km soil moisture and ∼ 0.5 R from the one-layer framework.
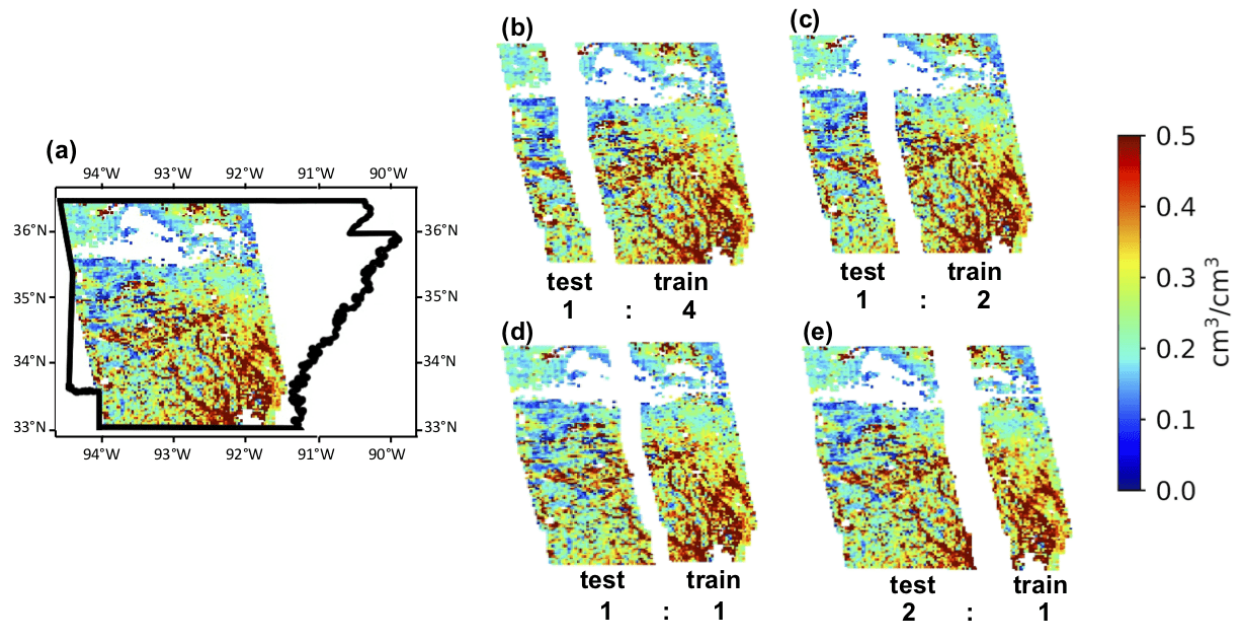

Figure 17: Vertical split of the whole swath at Arkansas on June 14th, 2018 with various size ratios of test region v.s. training region.
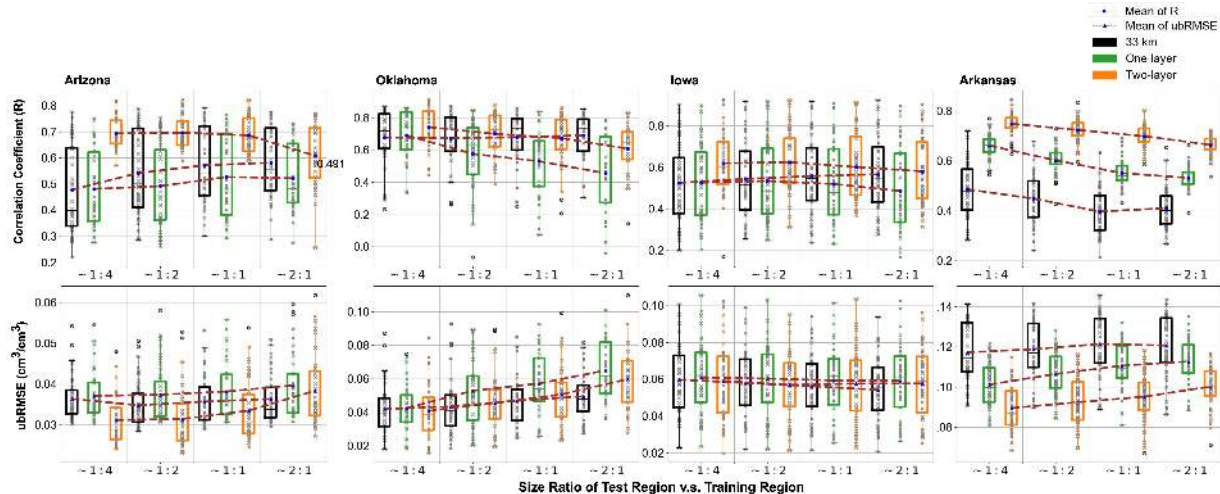
Figure 18: Performance comparisons of the resampled 33 km soil moisture, one-layer model and our proposed two-layer framework under different size ratios of the test region v.s. the training region.

5.5.2 Temporal Regional Learning

To obtain the temporal neighboring region for the temporal regional learning and spatio-temporal regional learning, we search back from the current day to find the most recent historical temporal neighboring region from another swath pattern within a 30-day window. In this section, we explore how the time span between the current day test region and the obtained temporal neighboring region influences the predictive performance. For the Sentinel-1 radars, the temporal resolution of one swath pattern is ~ 12 days while the timespan between two different but consecutive swath patterns is ~ 6 days. Considering that the swaths may be missing due to various reasons, the timespan between the current-day test region and the obtained temporal neighboring region can be ~ 6 days, ~ 18 days or ~ 30 days. To explore how these different timespans influence the downscaling performances of the one-layer framework and our two-layer framework, as shown in Figure 19, we continue searching within the 30-day window to find all available temporal neighboring regions and compare the results aggregated by different timespans. Figure 20 shows the comparisons at the three study areas. All the three study areas observe deteriorating performances, to different extents, as the timespan increases between the test region and the temporal neighboring region. Again, our two-layer framework returns better downscaling performances, especially at Arizona and Arkansas, compared with the one-layer framework and the resampled 33 km soil moisture.

5.6 Prediction at Gap Areas

In this section, we predict and compare the downscaled soil moisture produced by the models at real gap areas where the 3 km soil moisture estimates are indeed missing for the SMAP-Sentinel-1 product. As we have discussed in Section 4.5, we first train models using data from spatial-temporal neighboring regions of the gap regions. The learned models are then fitted to predict the 3 km brightness temperature and soil moisture at real gap areas. Figure 21 shows the predicted high resolution soil moisture at Arizona on June 13th, 2018, as an example, from the resampled 33 km soil moisture, the one-layer framework and the two-layer framework under different regional learning settings. Again, our two-layer framework reconstructs more subgrid soil moisture heterogeneities compared with the one-layer framework.

Gaps of SMAP/Sentinel-1 active-passive product for CONUS can be filled adopting the same methodology. The produced gap-filled soil moisture has the same 3 km resolution as the SMAP/Sentinel-1 active-passive product and the same spatial and temporal coverage as the SMAP enhanced passive product. Figure 22 (a) and (b) show the daily 3 km soil moisture product from the SMAP/Sentinel-1 active passive product and daily resampled 33 km soil moisture product from the SMAP enhanced passive product, on June 16th, 2018, respectively. Large areas that are not covered by the fine-scale soil moisture can be observed. Again, this is due to missing high resolution radar backscatter measurements. Figure 22 (c) and (d) show the comparison of the gap-filled soil moisture produced by the one-layer framework and our proposed two-layer framework. We were able to gap-fill most of the missing areas except those that have missing inputs, mainly the historical radar scatter measurements. Additionally, while the one-layer framework is able to reconstruct the 3 km soil moisture (under the spatial-temporal regional learning), it produces much less fine-scale soil moisture heterogeneity than proposed two-layer framework.
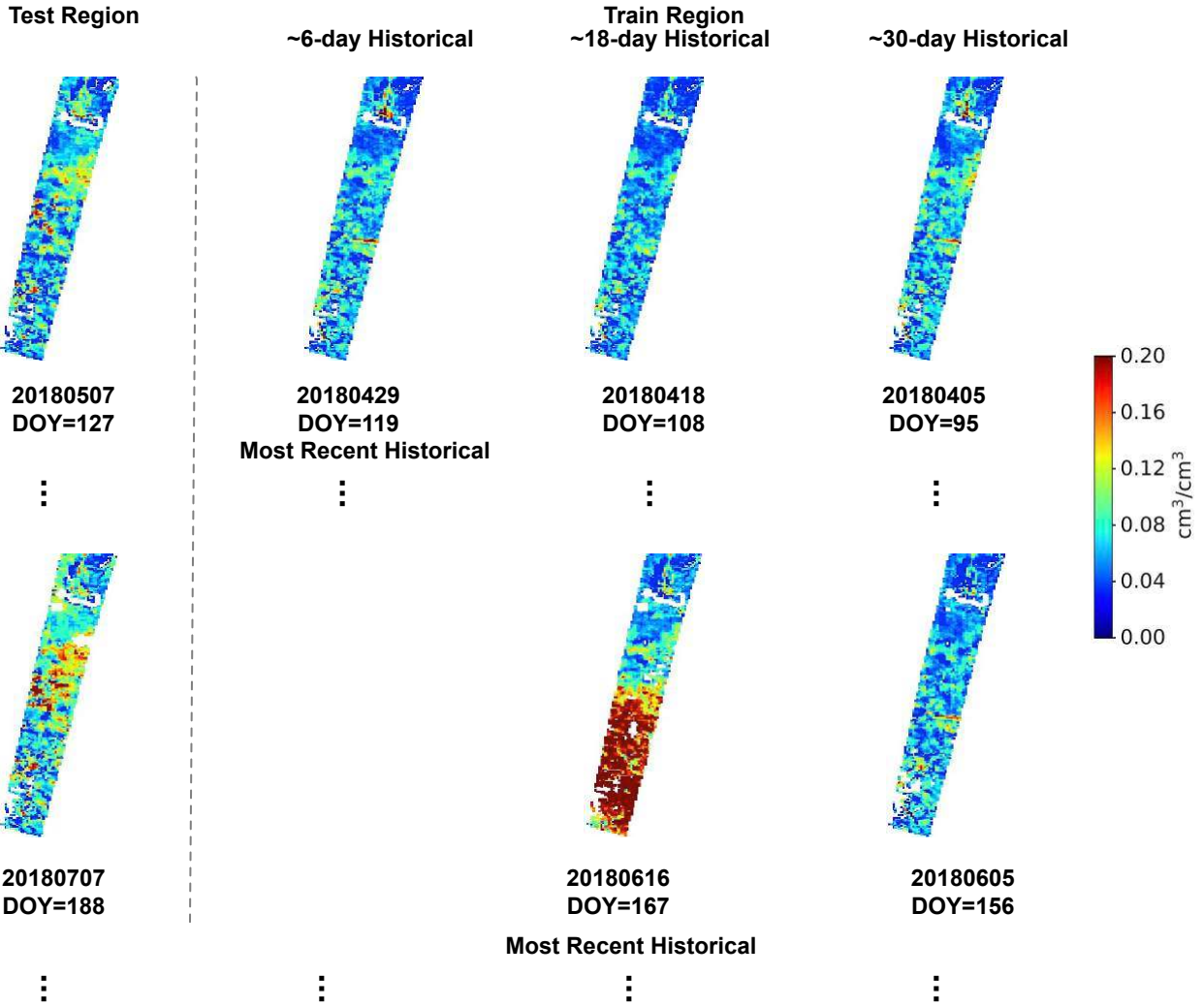
Figure 19: Search for temporal neighboring regions within a 30-day window for test regions at Arizona on May 7th, 2018 and July 7th, 2018.
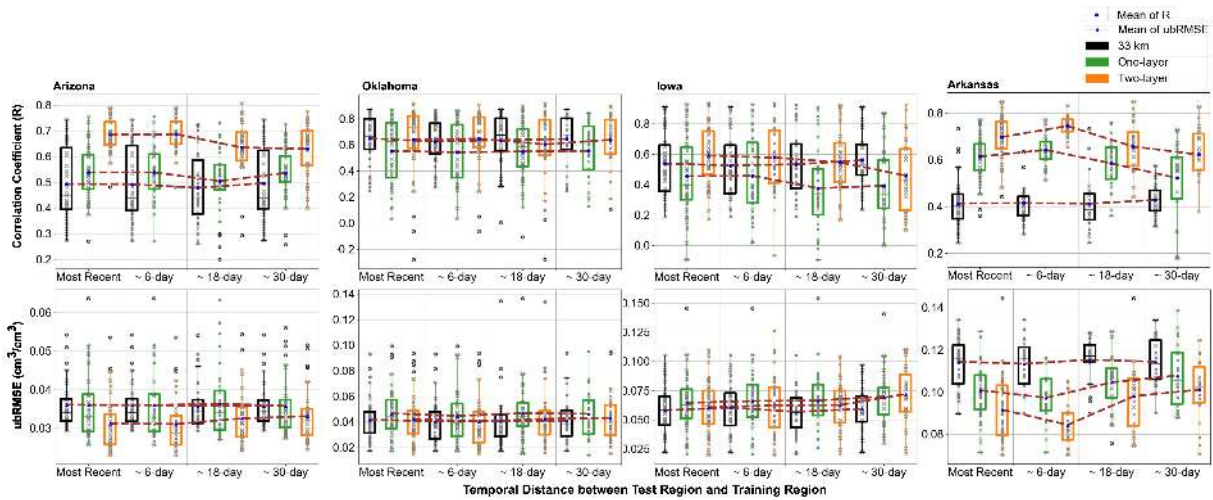

Figure 20: Performance comparisons of the resampled 33 km soil moisture, one-layer model and our proposed two-layer framework under different temporal distances between the test region and the temporal neighboring region.
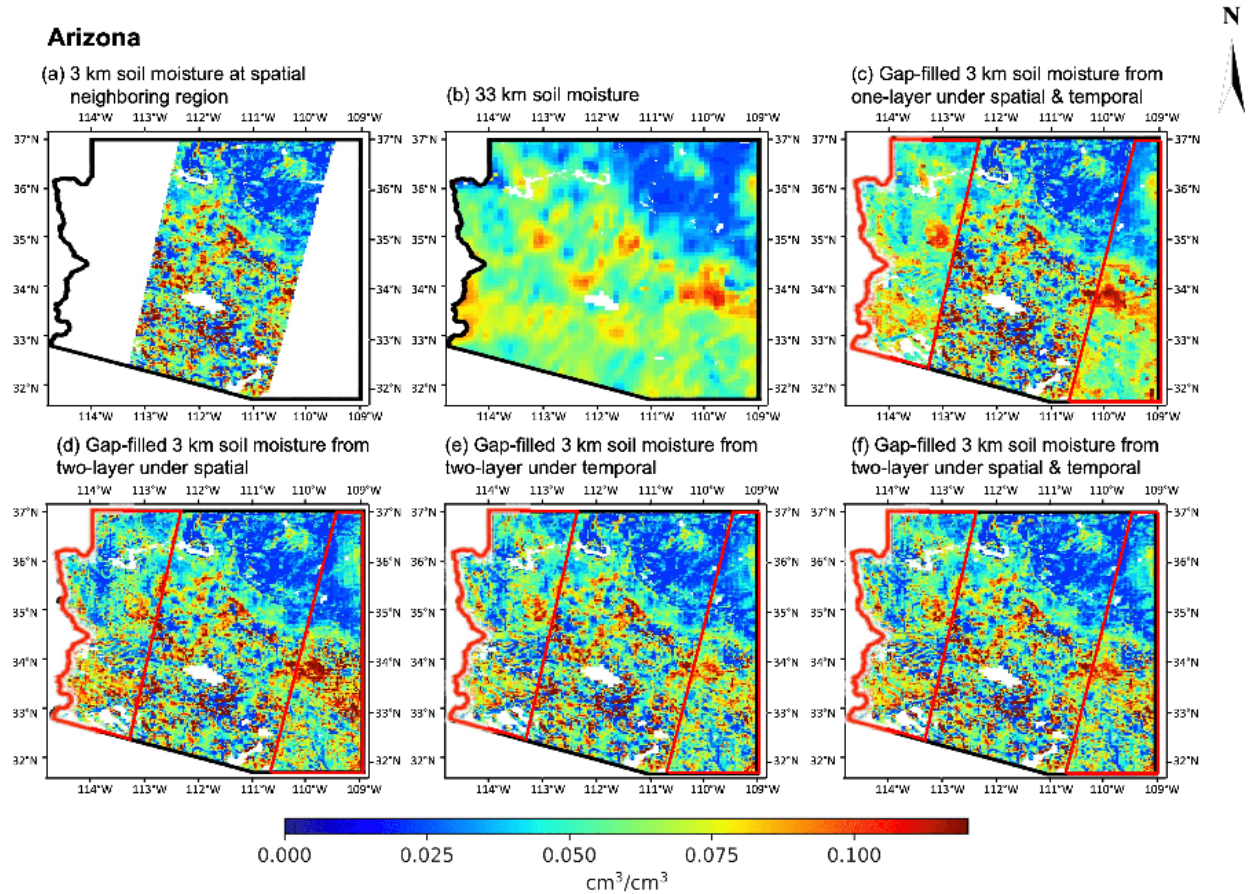
Figure 21: Real gap filling at Arizona on June 13th, 2018. (a) 3 km soil moisture from SMAP/Sentinel Product for spatial neighboring region on the same day. (b) Resampled 33 km soil moisture from SMAP L3_SM_P_E. Gap-filled 3 km soil moisture from (c) the one-layer model using spatial-temporal regional learning, (d) the hierarchical learning models using spatial regional learning, (e) the hierarchical learning models using temporal regional learning and (f) the hierarchical learning models using spatial-temporal regional learning.

6   Conclusions

Recognizing the importance of global-scale soil moisture estimates with fine spatial and temporal resolution, this paper introduces the regional learning strategy to enable regional gap-filled high-resolution soil moisture that maintains high consistency with the SMAP/Sentinel-1 3 km soil moisture product. It also presents a novel two-layer machine learning-based framework motivated by the SMAP active-passive retrieval algorithm. The two-layer framework can downscale/gap-fill soil moisture with more heterogeneity captured at areas where the SMAP/Sentinel-1 active-passive 3 km soil moisture product is missing. The average of historical radar backscatter measurements is introduced as one of the key inputs for soil moisture downscaling. We compare the two-layer framework with a conventional one-layer framework and the resampled 33 km soil moisture product under three different regional learning settings against hold-out SMAP/Sentinel-1 active-passive soil moisture estimates. We show that our two-layer framework outperforms the one-layer model and the resampled 33 km soil moisture, in terms of R and ubRMSE, under all scenarios. The most significant performance improvements of the proposed two-layer framework were seen on the R under the spatial-temporal regional learning where it is increased to 0.728 at Arizona compared to 0.493 for the resampled 33 km and 0.572 for the one-layer, 0.742 at Oklahoma compared to 0.649 for the resampled 33 km and 0.679 for the one-layer, 0.649 at Iowa compared to 0.537 for the resampled 33 km and 0.556 for the one-layer, and 0.744 at Arkansas compared to 0.413 for the resampled 33 km and 0.651 for the one-layer.

One limitation of our gap-filled soil moisture product is that its availability is still restricted by the availability and accuracy of the inputs, mainly the 33 km soil moisture and 33 km brightness temperature from the SMAP enhanced passive product. The best temporal resolution produced is therefore ∼2-3 days. This issue will hopefully be addressed

(a) 3 km soil moisture from SMAP/Sentinel-1 active-passive

(b) 33 km soil moisture from SMAP enhanced passive

(c) Gap-filled 3 km soil moisture from one-layer under spatial-temporal

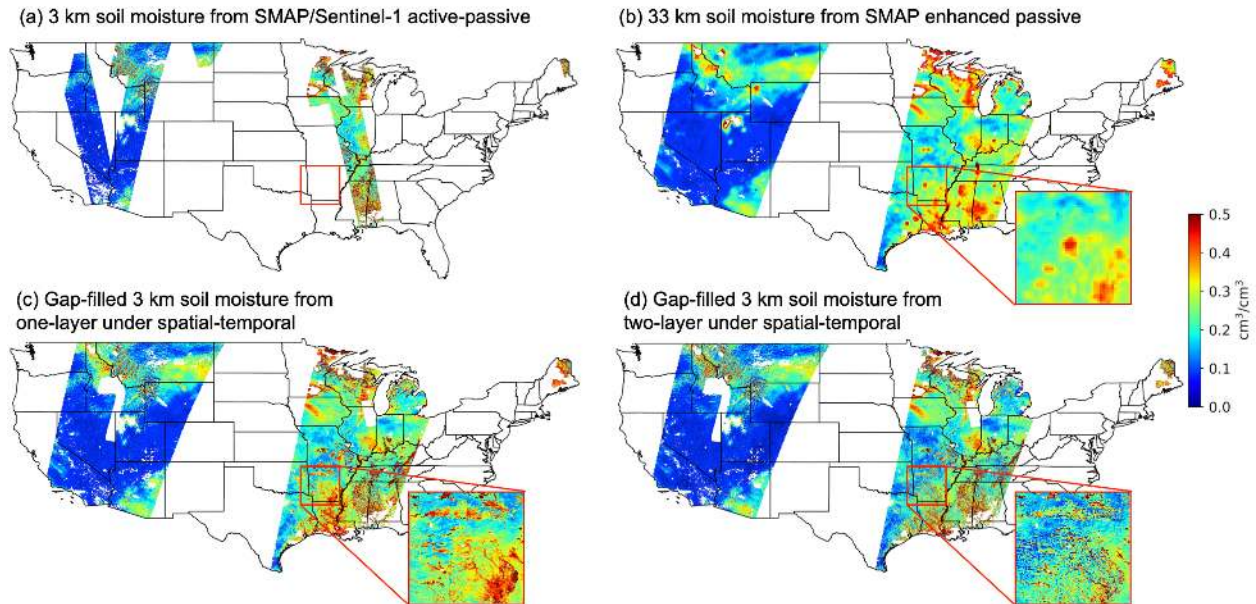(d) Gap-filled 3 km soil moisture from two-layer under spatial-temporal

Figure 22: Real gap filling of the Continental United States on June 16th, 2018. (a) Observed 3 km soil moisture from SMAP L2_SM_SP. (b) Resampled 33 km soil moisture from SMAP L3_SM_P_E. (c) Gap-filled 3 km soil moisture from the one-layer model under the spatial-temporal regional learning setting. (d) Gap-filled 3 km soil moisture from the proposed theory-guided hierarchical framework under the spatial-temporal regional learning setting.

by the generation of seamlessly covered coarse-scale soil moisture and/or brightness temperature data in the future. This can be achieved by the development of another gap-filling technique to fill the gaps of the 33 km soil moisture product or a fusion of multiple passive microwave soil moisture products.

References

[1] John D. Bolten, Wade T. Crow, Xiwu Zhan, Thomas J. Jackson, and Curt A. Reynolds. Evaluating the utility of remotely sensed soil moisture retrievals for operational agricultural drought monitoring. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3(1):57–66, mar 2010.

[2] M. Drusch. Initializing numerical weather prediction models with satellite-derived surface soil moisture: Data assimilation experiments with ECMWF's integrated forecast system and the TMI soil moisture data set. *Journal of Geophysical Research*, 112(D3), feb 2007.

[3] Sonia I. Seneviratne, Thierry Corti, Edouard L. Davin, Martin Hirschi, Eric B. Jaeger, Irene Lehner, Boris Orlowsky, and Adriaan J. Teuling. Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, 99(3-4):125–161, may 2010.

[4] John Pastor and W. M. Post. Influence of climate, soil moisture, and succession on forest carbon and nitrogen cycles. *Biogeochemistry*, 2(1):3–27, mar 1986.

[5] Wenju Cai, Tim Cowan, Peter Briggs, and Michael Raupach. Rising temperature depletes soil moisture and exacerbates severe drought conditions across southeast australia. *Geophysical Research Letters*, 36(21), nov 2009.

[6] N. Wanders, D. Karssenberg, A. de Roo, S. M. de Jong, and M. F. P. Bierkens. The suitability of remotely sensed soil moisture for improving operational flood forecasting. *Hydrology and Earth System Sciences*, 18(6):2343–2357, jun 2014.

[7] H.R. Bogena, J.A. Huisman, C. Oberdörster, and H. Vereecken. Evaluation of a low-cost soil water content sensor for wireless network applications. *Journal of hydrology*, 344(1-2):32–42, sep 2007.

[8] D. A. Robinson, S. B. Jones, J. M. Wraith, D. Or, and S. P. Friedman. A review of advances in dielectric and electrical conductivity measurement in soils using time domain reflectometry. *Vadose Zone Journal*, 2(4):444, 2003.

[9] Alan Robock, Konstantin Y. Vinnikov, Govindarajalu Srinivasan, Jared K. Entin, Steven E. Hollinger, Nina A. Speranskaya, Suxia Liu, and A. Namkhai. The global soil moisture data bank. *Bulletin of the American Meteorological Society*, 81(6):1281–1299, 2000.

[10] Wolfgang Wagner, Sebastian Hahn, Richard Kidd, Thomas Melzer, Zoltan Bartalis, Stefan Hasenauer, Julia Figa-Saldaña, Patricia de Rosnay, Alexander Jann, Stefan Schneider, Jürgen Komma, Gerhard Kubu, Katharina Brugger, Christoph Aubrecht, Johann Züger, Ute Gangkofner, Stefan Kienberger, Luca Brocca, Yong Wang, Günter Blöschl, Josef Eitzinger, and Kla Steinnocher. The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications. *Meteorologische Zeitschrift*, 22(1):5–33, feb 2013.

[11] E G Njoku, T J Jackson, V Lakshmi, T K Chan, and S V Nghiem. Soil moisture retrieval from AMSR-e. *IEEE Transactions on Geoscience and Remote Sensing*, 41(2):215–229, feb 2003.

[12] Yann H Kerr, Philippe Waldteufel, Jean-Pierre Wigneron, Steven Delwart, François Cabot, Jacqueline Boutin, Maria-José Escorihuela, Jordi Font, Nicolas Reul, Claire Gruhier, Silvia Enache Juglea, Mark R Drinkwater, Achim Hahne, Manuel Martín-Neira, and Susanne Mecklenburg. The SMOS mission: New tool for monitoring key elements ofthe global water cycle. *Proceedings of the IEEE*, 98(5):666–687, may 2010.

[13] Dara Entekhabi, Eni G. Njoku, Peggy E. O'Neill, Kent H. Kellogg, Wade T. Crow, Wendy N. Edelstein, Jared K. Entin, Shawn D. Goodman, Thomas J. Jackson, Joel Johnson, John Kimball, Jeffrey R. Piepmeier, Randal D. Koster, Neil Martin, Kyle C. McDonald, Mahta Moghaddam, Susan Moran, Rolf Reichle, J. C. Shi, Michael W. Spencer, Samuel W. Thurman, Leung Tsang, and Jakob Van Zyl. The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, 98(5):704–716, may 2010.

[14] Ramon Torres, Paul Snoeij, Dirk Geudtner, David Bibby, Malcolm Davidson, Evert Attema, Pierre Potin, BjÖrn Rommen, Nicolas Floury, Mike Brown, Ignacio Navas Traver, Patrick Deghaye, Berthyl Duesmann, Betlem Rosich, Nuno Miranda, Claudio Bruno, Michelangelo L'Abbate, Renato Croci, Andrea Pietropaolo, Markus Huchler, and Friedhelm Rostan. GMES sentinel-1 mission. *Remote Sensing of Environment*, 120:9–24, may 2012.

[15] A. AghaKouchak, A. Farahmand, F. S. Melton, J. Teixeira, M. C. Anderson, B. D. Wardlow, and C. R. Hain. Remote sensing of drought: Progress, challenges and opportunities. *Reviews of Geophysics*, 53(2):452–480, jun 2015.

[16] Randal D Koster, Luca Brocca, Wade T Crow, Mariko S Burgin, and Gabrielle J M De Lannoy. Precipitation estimation using l-band and c-band soil moisture retrievals. *Water resources research*, 52(9):7213–7225, sep 2016.

[17] Siyuan Tian, Paul Tregoning, Luigi J. Renzullo, Albert I. J. M. van Dijk, Jeffrey P. Walker, Valentijn R. N. Pauwels, and Sébastien Allgeyer. Improved water balance component estimates through joint assimilation of GRACE water storage and SMOS soil moisture retrievals. *Water resources research*, 53(3):1820–1840, mar 2017.

[18] E G Njoku, W J Wilson, S H Yueh, S J Dinardo, F K Li, T.J. Jackson, V. Lakshmi, and J. Bolten. Observations of soil moisture using a passive and active low-frequency microwave airborne sensor during SGP99. *IEEE Transactions on Geoscience and Remote Sensing*, 40(12):2659–2673, dec 2002.

[19] Fawwaz T. Ulaby, Pascale C. Dubois, and Jakob van Zyl. Radar mapping of surface soil moisture. *Journal of hydrology*, 184(1-2):57–84, oct 1996.

[20] E G Njoku and Dara Entekhabi. Passive microwave remote sensing of soil moisture. *Journal of hydrology*, 184(1-2):101–129, oct 1996.

[21] P E O'Neill, S Chan, E G Njoku, T Jackson, and R Bindlish. SMAP L3 radiometer global daily 36 km EASE-grid soil moisture, version 5, 2018a.

[22] Jianzhi Dong and Wade T. Crow. The added value of assimilating remotely sensed soil moisture for estimating summertime soil moisture - air temperature coupling strength. *Water resources research*, jul 2018.

[23] Yang Lu, Susan C. Steele-Dunne, Leila Farhadi, and Nick van de Giesen. Mapping surface heat fluxes by assimilating SMAP soil moisture and GOES land surface temperature data. *Water resources research*, dec 2017.

[24] R. C. Nijzink, S. Almeida, I. G. Pechlivanidis, R. Capell, D. Gustafssons, B. Arheimer, J. Parajka, J. Freer, D. Han, T. Wagener, R. R. P. Nooijen, H. H. G. Savenije, and M. Hrachowitz. Constraining conceptual hydrological models with multiple information sources. *Water resources research*, 54(10):8332–8362, oct 2018.

[25] Jian Peng, Alexander Loew, Olivier Merlin, and Niko E. C. Verhoest. A review of spatial downscaling of satellite remotely sensed soil moisture. *Reviews of Geophysics*, 55(2):341–366, jun 2017.

[26] Sabah Sabaghy, Jeffrey P. Walker, Luigi J. Renzullo, and Thomas J. Jackson. Spatially enhanced passive microwave derived soil moisture: Capabilities and opportunities. *Remote Sensing of Environment*, 209:551–580, may 2018.

[27] N N Das, Dara Entekhabi, Eni G. Njoku, Jiancheng J. C. Shi, Joel T. Johnson, and Andreas Colliander. 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, apr 2014.

[28] Thomas J. Jackson. III. measuring surface soil moisture using passive microwave remote sensing. *Hydrological processes*, 7(2):139–152, apr 1993.

[29] D Entekhabi, N Das, E Njoku, J Johnson, and J Shi. SMAP L3 radar/radiometer global daily 9 km EASE-grid soil moisture, version 3, 2016.

[30] N. N. Das, D. Entekhabi, S. Kim, S. Yueh, and P. O'Neill. Combining SMAP and sentinel data for high-resolution soil moisture product. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 129–131. IEEE, jul 2016.

[31] Simon Yueh, Dara Entekhabi, Peggy O'Neill, Eni Njoku, and Jared Entin. NASA soil moisture active passive mission status and science performance. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 116–119. IEEE, jul 2016.

[32] H Lievens, R H Reichle, Q Liu, G J M De Lannoy, R S Dunbar, S B Kim, N N Das, M Cosh, J P Walker, and W Wagner. Joint sentinel-1 and SMAP data assimilation to improve soil moisture estimates. *Geophysical research letters*, 44(12):6145–6153, jun 2017.

[33] E. Santi, S. Paloscia, S. Pettinato, D. Entekhabi, S. H. Alemohammad, and A. G. Konings. Integration of passive and active microwave data from SMAP, AMSR2 and sentinel-1 for soil moisture monitoring. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 5252–5255. IEEE, jul 2016.

[34] C. Rudiger, C.-H. Su, D. Ryu, and W. Wagner. Disaggregation of low-resolution l-band radiometry using c-band radar data. *IEEE Geoscience and Remote Sensing Letters*, 13(10):1425–1429, oct 2016.

[35] N N Das, D Entekhabi, R S Dunbar, S Kim, S Yueh, A Colliander, P E O'Neill, and T Jackson. SMAP/Sentinel-1 L2 radiometer/radar 30-second scene 3 km EASE-grid soil moisture, version 2, 2018.

[36] Thomas Jagdhuber, Martin Baur, Ruzbeh Akbar, Narendra N. Das, Moritz Link, Lian He, and Dara Entekhabi. Estimation of active-passive microwave covariation using SMAP and sentinel-1 data. *Remote Sensing of Environment*, 225:458–468, may 2019.

[37] European Space Agency. Sentinel-1 User Handbook. Technical report, 2013.

[38] Jeffrey P. Walker and Paul R. Houser. Requirements of a global near-surface soil moisture satellite mission: accuracy, repeat time, and spatial resolution. *Advances in water resources*, 27(8):785–801, aug 2004.

[39] N. S. Chauhan, S. Miller, and P. Ardanuy. Spaceborne soil moisture estimation at high resolution: a microwave-optical/IR synergistic approach. *International journal of remote sensing*, 24(22):4599–4622, jan 2003.

[40] Olivier Merlin, Christoph Rudiger, Ahmad Al Bitar, Philippe Richaume, Jeffrey P. Walker, and Yann H. Kerr. Disaggregation of SMOS soil moisture in southeastern australia. *IEEE Transactions on Geoscience and Remote Sensing*, 50(5):1556–1571, may 2012.

[41] María Piles, Adriano Camps, Mercè Vall-llossera, Ignasi Corbella, Rocco Panciera, Christoph Rudiger, Yann H. Kerr, and Jeffrey Walker. Downscaling SMOS-derived soil moisture using MODIS visible/infrared data. *IEEE Transactions on Geoscience and Remote Sensing*, 49(9):3156–3166, sep 2011.

[42] H. Lievens, S.K. Tomer, A. Al Bitar, G.J.M. De Lannoy, M. Drusch, G. Dumedah, H.-J. Hendricks Franssen, Y.H. Kerr, B. Martens, M. Pan, J.K. Roundy, H. Vereecken, J.P. Walker, E.F. Wood, N.E.C. Verhoest, and V.R.N. Pauwels. SMOS soil moisture assimilation for improved hydrologic simulation in the murray darling basin, australia. *Remote Sensing of Environment*, 168:146–162, oct 2015.

[43] Rolf H. Reichle, Dara Entekhabi, and Dennis B. McLaughlin. Downscaling of radio brightness measurements for soil moisture estimation: A four-dimensional variational data assimilation approach. *Water resources research*, 37(9):2353–2364, sep 2001.

[44] Peyman Abbaszadeh, Hamid Moradkhani, and Xiwu Zhan. Downscaling SMAP radiometer soil moisture over the CONUS using an ensemble learning method. *Water resources research*, dec 2018.

[45] S. Chai, J. Walker, Bert Veenendaal, and Geoff West. An artificial neural network model for downscaling of passive microwave soil moisture. *Recent Researches in Hydrology, Geology and Continuum Mechanics*, 2011.

[46] Jungho Im, Seonyoung Park, Jinyoung Rhee, Jongjin Baik, and Minha Choi. Downscaling of AMSR-e soil moisture with MODIS products using machine learning approaches. *Environmental earth sciences*, 75(15):1120, aug 2016.

[47] J Kolassa, R H Reichle, Q Liu, S H Alemohammad, P Gentine, K Aida, J Asanuma, S Bircher, T Caldwell, A Colliander, M Cosh, C Holifield Collins, T J Jackson, J Martínez-Fernández, H McNairn, A Pacheco, M Thibeault, and J P Walker. Estimating surface soil moisture from SMAP observations using a neural network technique. *Remote sensing of environment*, 204:43–59, jan 2018.

[48] Prashant K. Srivastava, Dawei Han, Miguel Rico Ramirez, and Tanvir Islam. Machine learning techniques for downscaling SMOS satellite soil moisture using MODIS land surface temperature for hydrological application. *Water Resources Management*, 27(8):3127–3144, jun 2013.

[49] Paulin Coulibaly, Yonas B. Dibike, and François Anctil. Downscaling precipitation and temperature with temporal neural networks. *Journal of Hydrometeorology*, 6(4):483–496, aug 2005.

[50] Xiaogang He, Nathaniel W. Chaney, Marc Schleiss, and Justin Sheffield. Spatial downscaling of precipitation using adaptable random forests. *Water resources research*, 52(10):8217–8237, oct 2016.

[51] Mitra Basu and Tin Kam Ho, editors. *Data Complexity in Pattern Recognition*. Springer Science & Business Media, illustrated edition, 2006.

[52] Kuai Fang, Chaopeng Shen, Daniel Kifer, and Xiao Yang. Prolongation of SMAP to spatiotemporally seamless coverage of continental u.s. using a deep learning neural network. *Geophysical Research Letters*, 44(21):11,030–11,039, nov 2017.

[53] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[54] Seonyoung Park, Jungho Im, Sumin Park, and Jinyoung Rhee. AMSR2 soil moisture downscaling using multisensor products through machine learning approach. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1984–1987. IEEE, jul 2015.

[55] Wei Zhao, Nilda Sánchez, Hui Lu, and Ainong Li. A spatial downscaling approach for the SMAP passive surface soil moisture product using random forest regression. *Journal of hydrology*, 563:1009–1024, aug 2018.

[56] Peggy O'Neill, Steven Chan, Eni Njoku, Tom Jackson, and Rajat Bindlish. Algorithm theoretical basis document L2 & L3 soil moisture (passive) data products, 2015.

[57] Mary J. Brodzik, Brendan Billingsley, Terry Haran, Bruce Raup, and Matthew H. Savoie. EASE-grid 2.0: Incremental but significant improvements for earth-gridded data sets. *ISPRS international journal of geo-information*, 1(1):32–45, mar 2012.

[58] Andreas Colliander, Thomas J. Jackson, Mike Cosh, Sid Misra, Rajat Bindlish, Jarrett Powers, Heather McNairn, Paul Bullock, Aaron Berg, Ramata Magagi, Peggy O'Neill, and Simon Yueh. Soil moisture retrieval with airborne PALS instrument over agricultural areas in SMAPVEX16. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 3949–3952. IEEE, jul 2017.

[59] Garry L. Schaefer, Michael H. Cosh, and Thomas J. Jackson. The USDA natural resources conservation service soil climate analysis network (SCAN). *Journal of Atmospheric and Oceanic Technology*, 24(12):2073–2077, dec 2007.

[60] Jesse E. Bell, Michael A. Palecki, C. Bruce Baker, William G. Collins, Jay H. Lawrimore, Ronald D. Leeper, Mark E. Hall, John Kochendorfer, Tilden P. Meyers, Tim Wilson, and Howard J. Diamond. U.s. climate reference network soil moisture and temperature observations. *Journal of Hydrometeorology*, 14(3):977–988, jun 2013.

[61] D A Miller and R A White. A conterminous united states multi-layer soil characteristics data set for regional climate and hydrology modeling. *Earth Interact*, 2, 1998.

[62] S Kim. SMAP Ancillary Data Report on Landcover Classification, 2013.

[63] NASA JPL. NASA shuttle radar topography mission global 3 arc second, 2013.

[64] Arthur Y. Hou, Ramesh K. Kakar, Steven Neeck, Ardeshir A. Azarbarzin, Christian D. Kummerow, Masahiro Kojima, Riko Oki, Kenji Nakamura, and Toshio Iguchi. The global precipitation measurement mission. *Bulletin of the American Meteorological Society*, 95(5):701–722, may 2014.

[65] Zhengming Wan and J. Dozier. A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on Geoscience and Remote Sensing*, 34(4):892–905, jul 1996.

[66] M. C. Peel, B. L. Finlayson, and T. A. McMahon. Updated world map of the köppen-geiger climate classification. *Hydrology and Earth System Sciences Discussions*, 4(2):439–473, mar 2007.

[67] D Entekhabi, N Das, E Njoku, J Johnson, and J Shi. Algorithm theoretical basis document L2 & L3 radar/radiometer soil moisture (active/passive) data products, 2012.

[68] P E O'Neill, Steven Chan, Eni Njoku, Tom Jackson, and Rajat Bindlish. Algorithm theoretical basis document L2 & L3 soil moisture (passive) data products, 2015.

[69] Nandita Gaur and Binayak P. Mohanty. Land-surface controls on near-surface soil moisture dynamics: Traversing remote sensing footprints. *Water resources research*, 52(8):6365–6385, aug 2016.

[70] Nandita Gaur and Binayak P. Mohanty. A nomograph to incorporate geo-physical heterogeneity in soil moisture downscaling. *Water resources research*, oct 2018.

[71] Dhruva Kathuria, Binayak P. Mohanty, and Matthias Katzfuss. A non-stationary geostatistical framework for soil moisture prediction in the presence of surface heterogeneity. *Water resources research*, jan 2019.

[72] A. Loew and W. Mauser. On the disaggregation of passive microwave soil moisture data using a priori knowledge of temporally persistent soil moisture fields. *IEEE Transactions on Geoscience and Remote Sensing*, 46(3):819–834, mar 2008.

[73] Tiejun Wang, Trenton E. Franz, Ruopu Li, Jinsheng You, Martha D. Shulski, and Chittaranjan Ray. Evaluating climate and soil effects on regional soil moisture spatial variability using EOFs. *Water resources research*, 53(5):4022–4035, may 2017.