remote sensing

MDPI

Article

Enhancing the Spatial Applicability of SMAP Soil Moisture Using Multi-Stage Machine Learning Downscaling

Jeonghwan Park [1,2] and Jisung Geba Chang [3,*]

[1] Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA; jeonghwan.park@nasa.gov
[2] Global Science and Technology, Greenbelt, MD 20770, USA
[3] Geo-Informatics Environment & Agriculture Bridging AI, Vienna, VA 22181, USA
* Correspondence: cjisung@ieee.org

Highlights

What are the main findings?

- A multi-stage machine learning framework enhanced the spatial applicability of SMAP soil moisture across the CONUS through 500 m and 250 m downscaling, followed by 100 m regional refinement.
- RF-based regional refinement using ISMN observations reduced the five-fold cross-validated RMSE to approximately 0.052 m^3/m^3.

What are the implications of the main finding?

- The framework supports repeatable downscaling of newly available SMAP Level-3 observations for agricultural monitoring.
- Downscaled SMAP soil moisture can serve as a practical baseline for regional refinement using in situ observations and local environmental controls.

Abstract

The Soil Moisture Active Passive (SMAP) mission provides global soil moisture (SM) observations with high temporal frequency and broad spatial coverage, valuable for hydrological and agricultural applications. However, its coarse spatial resolution can limit direct use in heterogeneous agricultural landscapes where soil moisture varies over short distances. This study developed a multi-stage machine learning framework to improve the spatial applicability of SMAP Level-3 SM across the conterminous United States (CONUS). SMAP brightness temperatures and nighttime land surface temperature (LST) were used as core retrieval-relevant predictors, while vegetation-related controls and soil–vegetation background interactions were incorporated as supplementary information during downscaling and as regional controls during refinement. The Random Forest (RF) based framework first generated 500 m and 250 m downscaled SMAP soil moisture across the CONUS. A 100 m regionally refined soil moisture field was then produced by combining the downscaled SMAP soil moisture with regional environmental drivers. After aggregation to the 9 km SMAP Level-3 product grid, the 500 m downscaled product showed strong consistency with the corresponding SMAP values, with an r^2 of 0.880 and an RMSE of 0.0292 m^3/m^3 across the CONUS. Direct comparison with International Soil Moisture Network (ISMN) in situ observations showed RMSE values of up to 0.099 m^3/m^3 depending on the site network. Following regional refinement based on hydro-topographic, soil, and vegetation-related controls, the overall five-fold cross-validated RMSE was 0.052 m^3/m^3, indicating improved predictive refinement within the available ISMN station network. These results indicate that multi-stage machine learning downscaling can enhance spatial interpretability

Check for updates

Academic Editor: Nicolas Baghdadi

Received: 15 July 2026
Revised: 21 August 2026
Accepted: 27 August 2026
Published: 31 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

and strengthen the applicability of SMAP SM for agricultural monitoring, while preserving broad consistency with the original passive microwave retrieval and providing a practical baseline for regional refinement.

Keywords: SMAP; soil moisture; downscaling; Random Forest; CONUS; ISMN

1. Introduction

Soil moisture (SM) is a fundamental land-surface variable that regulates water, energy, and carbon exchanges between the land and atmosphere and strongly influences hydrological processes, ecosystem functioning, and agricultural productivity [1,2]. Because it controls infiltration, evapotranspiration, plant water availability, and drought development, reliable SM information is essential for hydrology, meteorology, climatology, and water resource management, and is particularly valuable for agricultural decision-making [3]. Although ground-based measurements can provide accurate observations, they remain spatially limited because SM often varies markedly over short distances due to differences in soil properties, topography, vegetation, and meteorological forcing [4,5]. Consequently, point-scale observations alone are not sufficient to characterize spatially distributed SM over heterogeneous landscapes [4,6,7]. For regional hydrological analysis and agricultural applications, SM information is needed not only with frequent temporal coverage but also with spatial detail fine enough to capture subpixel variability relevant to hydrological modeling and agricultural decision-making [3,7].

To address this need, satellite remote sensing has become an indispensable means of monitoring soil moisture over broad regions with repeated coverage [3]. Among the available approaches, microwave remote sensing is especially valuable because it is sensitive to near-surface soil water conditions and can support soil moisture information under a wide range of atmospheric conditions [8]. Passive microwave observations, in particular, have played a central role in operational large-scale soil moisture because they provide measurements that are temporally consistent and spatially extensive [1]. The Soil Moisture Active Passive (SMAP) mission demonstrates this capability, providing frequent near-global observations of surface soil moisture that support climate, hydrological, and regional water-cycle studies [6]. The native SMAP radiometer observations have an approximate spatial support of 36 km, whereas the Enhanced Level-3 soil moisture product is posted on a 9 km EASE-Grid. The 9 km grid spacing represents the posting of the enhanced product and should not be interpreted as independent soil moisture information at a 9 km effective resolution [9]. For agricultural applications, however, the effective information content of SMAP remains coarse relative to field-scale soil moisture variability shaped by irrigation, crop type, terrain, soil properties, and management practices. Therefore, while passive microwave products provide valuable and frequent observations at regional to continental scales, spatial downscaling is needed to make them suitable for field-scale agricultural monitoring and other fine-scale applications [3,10].

To improve the spatial applicability of coarse satellite soil moisture products, numerous downscaling approaches have been developed using different types of auxiliary information [3,11–14]. Optical and thermal remote sensing variables, such as land surface temperature and vegetation indices, have been widely used because they provide fine spatial detail and are closely related to land surface moisture and energy conditions [15,16]. Active–passive microwave fusion approaches have also been explored to combine the higher spatial detail of radar observations with the stronger soil moisture sensitivity of passive microwave measurements [3,17]. In addition, geoinformation-based and model-based

methods have incorporated terrain, soil, and land surface modeling variables to represent spatial controls on soil moisture redistribution [3,18]. Machine learning approaches have also been adopted to integrate these diverse data sources and capture nonlinear relationships among them [12,13,19–22]. These approaches have improved soil moisture downscaling, but they still depend on the quality, timing, and transferability of their input data. For example, optical and thermal data can be affected by cloud contamination and temporal mismatch, radar-based approaches can be limited by revisit frequency and sensitivity to surface conditions, and static geoinformation or model-derived variables may introduce transferability and uncertainty issues across regions [3]. Beyond these data-related constraints, a key challenge is how machine learning models incorporate multi-source auxiliary variables for soil moisture downscaling properly [13,19,21,22]. When many fine-resolution auxiliary variables are introduced simultaneously, the learned model may become disproportionately influenced by features that contain stronger spatial variability or finer-scale texture, even when those variables are not the most physically relevant to the original passive microwave soil moisture retrieval. In such cases, variables more directly linked to retrieval-consistent soil moisture information may contribute less to the final prediction than spatially detailed ancillary inputs. As a result, the downscaled product may appear visually refined or spatially sharpened, while the extent to which it preserves the physical meaning of the original microwave-derived soil moisture signal remains uncertain. This issue becomes more important when the timing of feature datasets is not handled carefully, because the physical relevance of thermal, vegetation, and microwave variables depends strongly on their timing relative to the soil moisture observation [3,21].

Importantly, passive microwave soil moisture retrievals, including SMAP, are grounded in a microwave radiative transfer framework commonly represented by the tau–omega model [8,17,23]. This physical basis suggests that downscaling should not be treated simply as an image-sharpening problem. Rather, it should be designed as a redistribution process that remains consistent with retrieval-relevant soil moisture information. Accordingly, the downscaling model should be built around variables directly observed by passive microwave sensors, such as brightness temperature, together with surface variables that physically influence the microwave soil moisture signal, such as temperature, vegetation, soil, and terrain conditions. Therefore, these variables should be incorporated carefully, considering both their physical roles and their timing relative to the SMAP observation. In this study, we develop and evaluate a multi-stage machine learning downscaling framework for SMAP soil moisture across agricultural regions of the conterminous United States (CONUS). The framework consists of retrieval-relevant RF downscaling, stratified incorporation of vegetation-related background conditions, and regional refinement. Within this framework, passive microwave brightness temperatures, thermal and optical variables, topographic information, and soil-related variables are assessed according to their contributions to the final prediction. The objective is to enhance the spatial applicability of SMAP soil moisture for agricultural monitoring while preserving consistency with the physical meaning of the original coarse-resolution passive microwave product. The initial downscaling stages are intentionally designed to remain constrained by the SMAP passive microwave signal. Rather than allowing high-resolution ancillary variables to independently define the output pattern, the framework uses SMAP SM, TbV, TbH, and LST to preserve the retrieval-relevant microwave–thermal structure before introducing local environmental refinement.

2. Materials and Methods

2.1. Study Area and ISMN In Situ Soil Moisture Networks

This study focused on the conterminous United States (CONUS; approximate center: 39.8°N, 98.6°W) as the primary domain for developing and applying the SMAP soil moisture

downscaling framework. The CONUS provides a broad environmental gradient for examining soil moisture variability because it includes contrasting climate, ecological, land-cover, and soil conditions within a single continental-scale domain. Precipitation generally ranges from arid and semi-arid conditions in the western and interior regions to more humid conditions in the eastern and southeastern United States, while temperature conditions vary from cooler northern and high-elevation regions to warmer southern and southwestern regions [24]. These climate gradients influence evapotranspiration demand, antecedent wetness, infiltration, and seasonal soil moisture persistence. The domain also spans major ecoregional settings, including western drylands, the Great Plains, humid temperate regions, and intensively managed agricultural landscapes, which represent different combinations of vegetation structure, terrain, climate, and soil water-storage conditions. Land-cover heterogeneity was characterized using the USDA National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL), which represents major crop, grassland/pasture, and non-agricultural land-cover patterns across the CONUS [25]. These land-cover differences are relevant to soil moisture because crop type, canopy cover, rooting characteristics, surface roughness, and vegetation water use can affect infiltration, evapotranspiration, and near-surface moisture dynamics. In addition, soil texture and hydraulic properties influence drainage, water-holding capacity, and near-surface moisture retention [26,27]. Therefore, the CONUS is not only a large geographic domain, but also a representative testbed for evaluating whether the proposed downscaling framework can preserve SMAP-consistent wet–dry patterns while resolving finer-scale variability across diverse environmental and agricultural conditions.

Figure 1 shows the study configuration, including the CONUS agricultural domain, the selected agricultural evaluation sites, and the distribution of ground-based monitoring stations used in this study. Ground-based soil moisture observations were obtained from the International Soil Moisture Network (ISMN), which provides harmonized in situ measurements from multiple monitoring networks [28]. The datasets used in this study included the Soil Climate Analysis Network (SCAN; 146 stations) and the U.S. Climate Reference Network (USCRN; 66 stations), both distributed across the CONUS [29,30]. Surface soil moisture observations at a 5 cm depth were extracted for the growing season period from April to October 2025. These stations were used to evaluate the spatial and temporal consistency of the refined version of the downscaled soil moisture product.

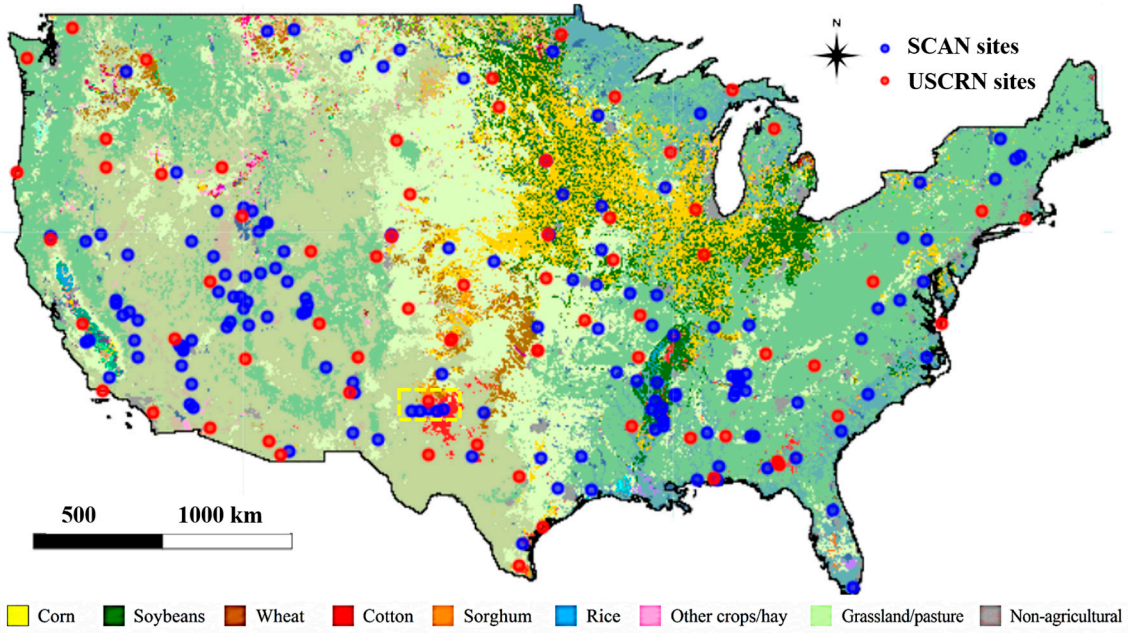


Figure 1. Study area across the conterminous United States (CONUS), showing the 2024 Cropland Data Layer (CDL) from USDA National Agricultural Statistics Service (NASS) and the distribution of

evaluation stations from the Soil Climate Analysis Network (SCAN; 146 sites) and the U.S. Climate Reference Network (USCRN; 66 sites) obtained from the International Soil Moisture Network (ISMN). The simplified major CDL legend at the bottom represents dominant crop and land-cover categories, including corn, soybeans, wheat, cotton, sorghum, rice, other crops/hay, grassland/pasture, and non-agricultural land. Yellow dashed boxes indicate the regions selected for further analysis in the Results and Discussion sections.

2.2. Soil Moisture Data and Feature Variables

2.2.1. SMAP Soil Moisture Product and Microwave Retrieval Basis

The primary target variable in the downscaling framework was surface soil moisture (SM) from the SMAP Enhanced Level-3 passive radiometer product (SPL3SMP_E), specifically the 'soil_moisture_am' band. SPL3SMP_E is posted on a 9 km EASE-Grid, whereas the underlying radiometer information retains an approximate native spatial support of 36 km. The selected band corresponds to the descending morning overpass closest to 6:00 a.m. local solar time [31]. Accordingly, the 9 km grid spacing represents the posting of the input product rather than an independent effective soil moisture resolution of 9 km. The SPL3SMP_E soil-moisture field served as the input target for statistical spatial disaggregation and should be distinguished from the in situ soil moisture observations used for site-level evaluation and regional refinement. The purpose of the initial downscaling stages was to redistribute the SMAP-consistent soil moisture signal onto finer output grids using retrieval-relevant and environmentally meaningful auxiliary information. The AM retrieval was selected because morning observations provide a more temporally stable reference for matching auxiliary remote sensing datasets and are less affected by strong daytime heating, rapidly changing surface conditions, and large thermal contrasts. The vertically and horizontally polarized corrected brightness temperatures (tb_v_corrected_am, hereafter TbV, and tb_h_corrected_am, hereafter TbH) from the same SMAP Level-3 product were included as microwave feature inputs [31]. Because TbV and TbH retain direct sensitivity to the microwave emission signal from which passive soil moisture is retrieved, they were treated as core feature variables to preserve retrieval-relevant microwave information and to reduce the risk that the learned model would be dominated solely by finer-resolution ancillary inputs.

The selection of downscaling inputs was conceptually guided by the physical basis underlying SMAP passive soil moisture retrieval, which is commonly formulated using the τ–ω (tau–omega) radiative transfer model for microwave interactions with soil and vegetation. In this framework, the observed brightness temperature (T_b) is represented as a combination of soil emission attenuated by three main components: soil emission attenuated by vegetation, direct vegetation emission, and soil–vegetation interaction terms [8,17,23]:

$$T_b = \gamma e_s T_s + (1-\omega)(1-\gamma)T_c + (1-\omega)(1-\gamma)\gamma(1-e_s)T_c \quad (1)$$

where *Ts* and *Tc* denote the effective soil and canopy temperatures, respectively, e_s is the soil emissivity, ω is the single scattering albedo of the vegetation layer, and γ is the vegetation transmissivity given by $\gamma = e^{-\tau/cos\theta}$, τ is the vegetation optical depth, and θ is the sensor incidence angle. In this formulation, soil emissivity is strongly controlled by the dielectric properties of the soil, which are highly sensitive to soil moisture. The thermal terms, represented by soil and canopy temperatures, indicate the importance of surface temperature conditions in interpreting microwave emission. Vegetation transmissivity and optical depth represent canopy attenuation and are related to vegetation structure and water status. Thus, the main physical components of the microwave retrieval framework are associated with soil dielectric behavior, surface thermal conditions, vegetation attenuation, and soil–vegetation interactions. Although the operational SMAP retrieval includes

algorithm-specific ancillary inputs and implementation details, its physical basis is rooted in this tau–omega microwave emission framework. In this study, the τ–ω model was not used to directly retrieve soil moisture at finer resolution. Instead, it provided the physical rationale for selecting downscaling features that are related to microwave emission, surface temperature, vegetation status, and background soil–terrain controls.

2.2.2. Remote Sensing and Geoinformation Variables

Based on the soil moisture-related physical factors described in Section 2.2.1, auxiliary remote sensing and geoinformation variables were selected to represent surface thermal conditions, vegetation status, and background soil–terrain controls on moisture redistribution. At the coarse spatial resolution of SMAP radiometer products, the observed microwave signal represents an aggregated response over heterogeneous land surface conditions. Therefore, these auxiliary variables were not selected simply to add finer spatial detail, but to provide physically meaningful information for spatially disaggregating and redistributing the SMAP-consistent soil moisture signal. Table 1 summarizes the principal input datasets, variables, spatial resolutions, and temporal characteristics used in the proposed framework. These include dynamic microwave, thermal, and vegetation observations together with static terrain, hydrologic, and soil-property information.

Table 1. Summary of input datasets used in the soil moisture downscaling framework.

Dataset/Product	Variable Used	Spatial Resolution	Temporal Resolution
SMAP Level-3 (SPL3SMP_E)	Soil moisture (SM), TbV, TbH	9 km grid posting	Daily
MODIS MYD21A1N	Nighttime LST	1 km	Daily
MODIS MOD13Q1	NDVI, EVI	250 m	16-day
Landsat Collection 2	NDVI, EVI	30 m	32-day and Annual
NASADEM	Elevation, slope, aspect	30 m	Static
HYDROGRAPHY90	Flow accumulation	90 m	Static
POLARIS	Saturated water content, bulk density, clay fraction	30 m	Static

Among these variables, surface thermal conditions were first represented using land surface temperature (LST). LST is closely related to the effective surface temperature component of passive microwave emission and is therefore an important dynamic feature for SMAP soil moisture downscaling. In this study, LST was derived from the MODIS Aqua MYD21A1N Version 6.1 product, which provides daily nighttime land surface temperature observations at 1 km spatial resolution. Nighttime LST was used because it is more temporally consistent with the SMAP AM retrieval than daytime thermal observations and is less affected by strong daytime heating and large daytime thermal contrasts. Same-date LST was used whenever available; when an exact-date match was unavailable, the most recent available LST observation within the specified temporal-matching window was used to minimize temporal mismatch.

Vegetation status was represented using vegetation-index variables because vegetation affects the passive microwave soil moisture signal through canopy attenuation, canopy structure, and vegetation water condition. NDVI and EVI were derived from the MODIS MOD13Q1 Version 6.1 product, which provides 16-day vegetation index composites at 250 m spatial resolution [32]. For each SMAP observation date, the nearest MODIS vegetation-index composite within the corresponding 16-day compositing period was selected to represent temporally consistent vegetation conditions. In addition, NDWI was incorporated as a supplementary vegetation- and moisture-sensitive feature to capture canopy water condition and surface wetness variability relevant to soil moisture redistribu-

tion [33]. Vegetation indices provide information on vegetation condition, while thermal infrared LST provides a physically relevant approximation of effective surface temperature for soil moisture downscaling [34,35]. Because vegetation conditions generally change more slowly than soil moisture and LST, the 16-day MODIS vegetation composites were considered suitable for representing vegetation status near each SMAP observation date. Landsat Collection 2 Tier 1 annual vegetation index composites at 30 m spatial resolution were also examined as supplementary vegetation indicators to assess finer-scale vegetation patterns for the same analysis year as the SMAP observations with the 2025 composite used for the 2025 implementation, although they were not used as the primary temporally consistent vegetation inputs in the main framework.

Among the soil moisture-related physical factors described above, background soil–terrain controls were represented using geoinformation variables. These variables were included because terrain and soil variability can strongly influence the spatial redistribution of soil moisture across heterogeneous landscapes, particularly in agricultural regions [7,18,36]. Elevation was obtained from the NASADEM dataset at 30 m spatial resolution, and terrain derivatives including slope and aspect were calculated from the elevation surface [37]. Hydrologic redistribution potential was represented using the HYDROGRAPHY90 flow accumulation layer [38]. In addition, soil-related features were incorporated from the POLARIS database, including saturated water content, bulk density, and clay fraction, to account for spatial variation in soil hydraulic and textural properties relevant to soil moisture dynamics [26]. Figure 2 presents an example visualization of these datasets on May 3, 2025, including SMAP Level-3 soil moisture (SM), TbV, LST, NDVI, elevation, and saturated soil water content (θ_S) across the CONUS.

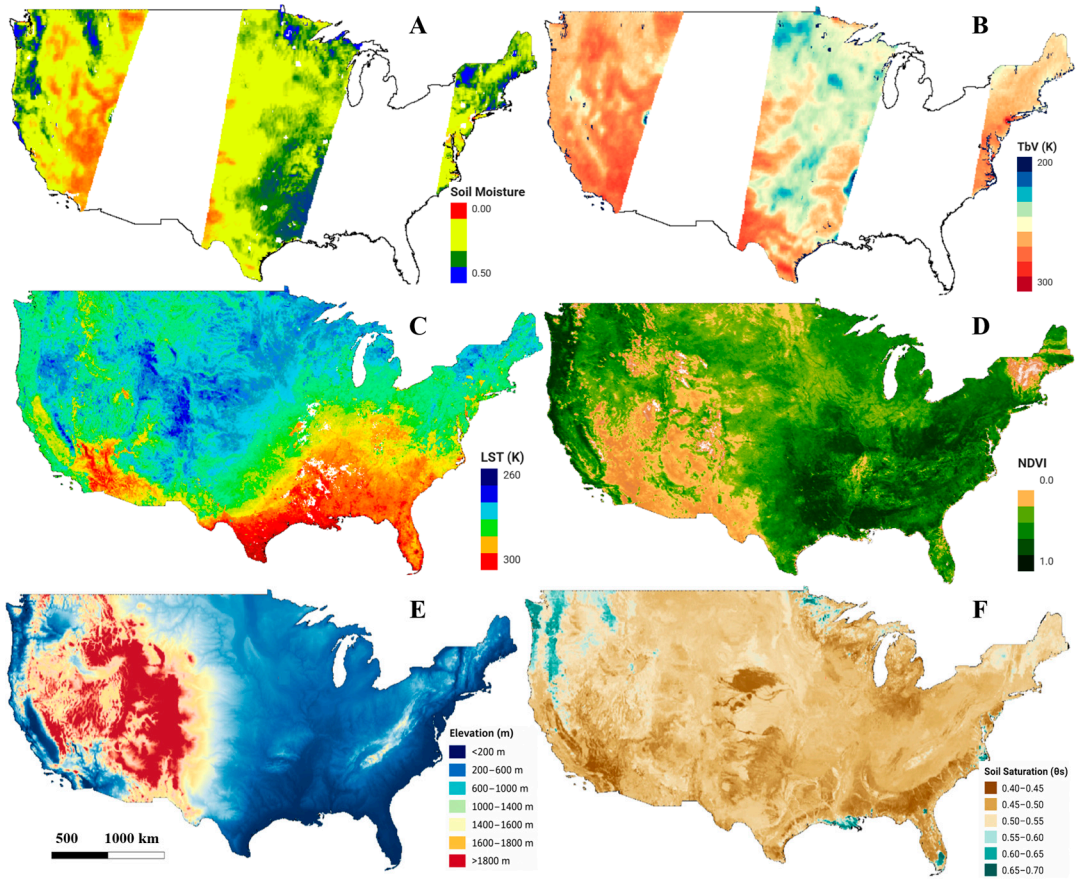


Figure 2. Example feature datasets across the CONUS on May 3, 2025, including (**A**) SMAP Level-3 soil moisture, (**B**) SMAP TbV, (**C**) MODIS Aqua LST, (**D**) MODIS Aqua NDVI, (**E**) the NASADEM digital elevation model (DEM), and (**F**) saturated soil water content (θ_S) from the POLARIS dataset.

2.3. Multi-Stage Machine Learning Downscaling Framework

The proposed downscaling framework was designed as a multi-stage machine learning approach for spatially redistributing the SMAP passive microwave soil moisture signal, rather than as a purely data-driven image-sharpening procedure. Physical considerations were incorporated into the framework in three main ways. First, retrieval-relevant microwave and thermal variables, including TbV, TbH, and LST, were treated as core inputs, and feature contribution analysis was used to examine whether their contributions remained central in the machine learning model. Second, physically meaningful ancillary variables with strong spatial gradients or high feature contributions were incorporated through condition- or range-based stratified modeling, rather than allowing them to dominate a single global model. Third, after the SMAP-consistent soil moisture field was downscaled, ISMN observations were used for site-level evaluation and regional refinement with local hydro-topographic, vegetation, and soil-related variables. In this refinement step, Multiple Linear Regression (MLR) was used to provide physical interpretability by quantifying the influence of direct physical controls, whereas RF-based refinement was used to evaluate the nonlinear predictive potential of the same physically meaningful variables. Within this framework, Random Forest (RF) regression model was used as the base algorithm for the main downscaling step because it can capture nonlinear relationships among heterogeneous remote sensing and geoinformation variables and showed stable performance for CONUS-wide implementation [12,13,21]. The workflow therefore included contribution-guided feature selection, adaptive stratified RF downscaling, generation of spatially explicit soil moisture products, ISMN-based evaluation, and post-downscaling regional refinement. Figure 3 illustrates the overall methodological workflow for soil moisture downscaling and the subsequent evaluation and refinement framework.

Before model construction, all feature layers were temporally matched and spatially aligned before model training and prediction. Temporal matching was performed using the SMAP AM observation date as the reference date. MODIS Aqua nighttime LST was selected to maintain consistency with the SMAP AM retrieval; same-date LST was used when available, and the most recent available LST observation within approximately ± 7 days was used when exact matching was not possible. Vegetation-related variables, including NDVI, EVI, and NDWI, were treated as temporally smoother features than soil moisture and LST, and the nearest available vegetation-index observation was selected relative to each SMAP observation date. After temporal matching, all datasets were spatially harmonized to the target analysis grids. The resampling described here refers to spatial resampling and grid alignment, not temporal resampling. SMAP soil moisture, brightness temperatures, MODIS LST, vegetation indices, hydro-topographic variables, terrain variables, and soil-property layers were reprojected and aligned to the 500 m and 250 m grids for CONUS-wide downscaling. After the 500 m and 250 m downscaled SMAP soil moisture fields were generated, a 100 m refined soil moisture field was produced by integrating the downscaled SMAP soil moisture with regional environmental drivers, including 30 m EVI, 30 m DEM-derived terrain information, 90 m flow accumulation, and 30 m soil-related variables. During reprojection and grid alignment, the default nearest-neighbor resampling method was used unless otherwise specified. Continuous predictor layers were aligned to the target grids using this resampling scheme, whereas categorical land-cover information was not interpolated and was used primarily to define the cultivated-land analysis mask, thereby preserving the original class information. The 500 m and 250 m grids were used for the main CONUS-wide downscaling analysis, while the 100 m field was generated for finer-scale regional environmental refinement and agricultural interpretation. A 100 m regional refinement was also implemented over regional subsets, whereas routine CONUS-wide processing at that scale was not computationally practical within Google Earth

Engine (GEE). These variables were selected not simply to enhance spatial detail, but to provide physically and environmentally relevant fine-scale information that could support downscaling while remaining consistent with the original passive microwave soil moisture signal. For data-quality screening, the analysis was restricted to the valid cultivated-land domain, and only samples containing valid values for the SMAP soil-moisture target and all predictors required by the corresponding model configuration were retained. Samples containing missing or masked values after temporal matching and spatial harmonization were excluded from subsequent model analysis.

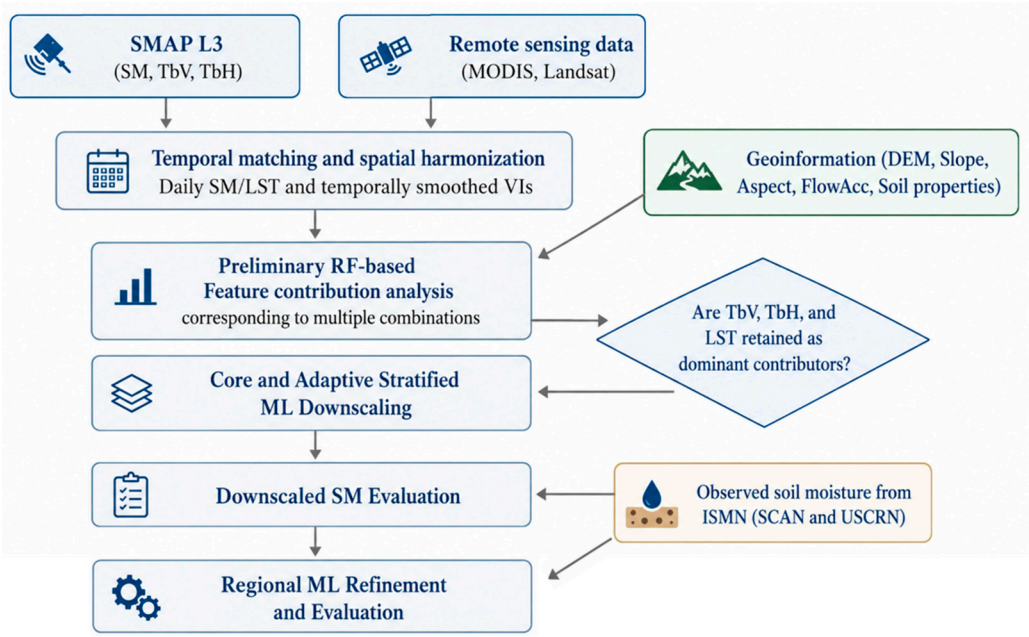


Figure 3. Overview of the multi-stage machine learning soil moisture downscaling workflow. The framework integrates SMAP Level-3 soil moisture and brightness temperature data (TbV, TbH), temporally matched LST, vegetation-related variables, and geoinformation variables through temporal matching and spatial harmonization. The multi-stage design includes three main components: (1) contribution-guided feature selection to preserve strong contributions from the core retrieval-relevant variables, TbV, TbH, and LST; (2) Core RF Downscaling with adaptive stratified RF modeling to incorporate high-contribution ancillary variables by condition or range without allowing them to dominate a single global model; and (3) site-level evaluation and regional ML refinement using ISMN observations and local environmental controls.

Following this preprocessing, the temporally matched and spatially harmonized feature layers were used as inputs to the multi-stage machine learning framework. Microwave, thermal, vegetation, hydrologic, terrain, and soil-related variables were merged into stacked feature images for model construction. During model training, the SMAP Level-3 soil moisture retrieval was used as the coarse-resolution target for the main downscaling model. Therefore, the proposed framework should not be interpreted as an independent field-scale soil moisture retrieval, but as a spatial disaggregation of the SMAP passive microwave soil moisture signal. Because TbV and TbH retain the spatial support of the native SMAP passive microwave observations, their alignment to the 500 m and 250 m grids does not create independent fine-scale microwave information. Instead, fine-scale spatial contrasts in the downscaled product are introduced primarily through LST, hydro-topographic variables, soil properties, and vegetation-based stratification, while TbV and TbH help preserve consistency with the original SMAP retrieval signal. For the feature-combination experiments, samples were distributed approximately equally across the available observation dates and randomly selected within the valid CONUS cultivated-land domain at the target analysis

scale. Sampling was not explicitly stratified by geographic region, soil-moisture range, or SMAP product grid; consequently, multiple fine-grid samples could share the same SMAP soil-moisture target while having different predictor values.

A preliminary RF-based feature contribution analysis was then conducted to examine how different combinations of feature variables influenced model behavior and variable importance. Multiple feature sets were tested using SMAP soil moisture as the response variable, with particular attention to whether TbV, TbH, and LST retained strong contributions in the learned model. RF regression was implemented using 100 trees, and the same model configuration was maintained across the feature-combination experiments so that differences in model fit and feature contribution primarily reflected changes in the predictor sets. This step was used to identify a feature structure that balanced comparative model fit with retrieval consistency, rather than allowing the model to be dominated by fine-resolution ancillary variables with strong spatial texture but weaker physical linkage to the original passive microwave retrieval. Thus, feature contribution analysis was used as a diagnostic screening step, not as direct proof that the physical mechanism was preserved. Even physically meaningful ancillary variables, such as vegetation indices, terrain variables, and soil hydraulic properties, can dominate a flexible RF model when they contain strong spatial gradients or high-resolution texture. If their contributions become larger than those of TbV, TbH, and LST, the downscaled product may increasingly reflect ancillary spatial patterns rather than the original SMAP microwave sensing signal. Through this iterative process, the feature structure was refined before the final downscaling model was established.

To preserve retrieval-consistent microwave and thermal contributions, the main downscaling step was implemented using an adaptive RF design. TbV, TbH, and LST were treated as the core features, whereas supplementary variables were incorporated when they improved comparative model fit without substantially weakening the contribution of the core microwave–thermal variables. Variables with strong spatial gradients or high contributions were not regarded as physically invalid; rather, their effects were incorporated through stratified RF modeling or subsequent regional refinement to reduce the risk that a single global model would be driven primarily by high-spatial-resolution ancillary texture. In the stratified design, separate RF models were trained for different vegetation-background conditions or ranges, allowing vegetation-related information to support downscaling while maintaining a stronger connection to the SMAP retrieval-consistent microwave–thermal structure.

The trained RF models were subsequently applied to the stacked feature images to generate downscaled soil moisture fields on 500 m and 250 m output grids across the CONUS. A 100 m regional refinement was also demonstrated over regional subsets to examine finer-scale applicability, although routine CONUS-wide processing at this resolution was limited by memory and computational constraints. Input-product consistency was assessed by aggregating the 250 m and 500 m estimates to the 9-km SMAP product grid and comparing the aggregated estimates with the corresponding SMAP Level-3 soil-moisture values. Importantly, the downscaled SMAP soil moisture was retained as the baseline predictor during regional refinement so that local environmental variables refined, rather than replaced, the SMAP-consistent moisture signal. The resulting downscaled soil moisture products were then compared with ISMN observations to examine their site-level consistency relative to local measurements. In addition, because the downscaled products provided more spatially resolved moisture patterns than the original SMAP product grid, they served as intermediate baseline soil moisture fields for post-downscaling regional refinement using local hydro-topographic, soil, and vegetation-related information. In this refinement step, in situ soil moisture was used as the target variable, while the downscaled

soil moisture and regional environmental variables were used as predictors. Both interpretable region-specific multiple linear regression and more flexible RF-based refinement were examined. This refinement step allowed additional physical controls to be incorporated after the SMAP passive microwave signal had been spatially disaggregated. The MLR-based analysis was included primarily to provide physical interpretability of these regional controls, whereas the RF-based refinement was used to examine predictive flexibility. Overall, the framework linked contribution-guided input selection, adaptive stratified RF downscaling, spatially explicit soil moisture generation, ISMN-based evaluation, and regional refinement. This post-downscaling regional refinement was particularly relevant in agricultural landscapes, where hydro-topographic and soil controls can influence local moisture redistribution at scales finer than the 9-km SMAP product grid [7,18].

2.4. Feature Contribution Analysis and Model Evaluation

To assess the relative importance of each feature, feature contribution analysis was conducted within the Random Forest framework. Variable importance was quantified using the mean decrease in node impurity, which enabled the identification of features that contributed most strongly to soil moisture variability [39,40]. This analysis was used to compare the relative roles of microwave, thermal, vegetation, topographic, and soil-related variables across different feature combinations and to diagnose whether the core retrieval-relevant variables remained important within the learned RF model. These feature-combination experiments were treated as contribution-guided diagnostic comparisons rather than as a conventional accuracy-maximizing ablation. Model selection considered comparative model fit together with the contribution structure of the retrieval-relevant microwave–thermal variables.

Comparative model fit across different feature combinations was examined using the root mean squared error (RMSE), unbiased root mean squared error (ubRMSE), and the coefficient of determination (r^2). RMSE was used to quantify overall error magnitude, ubRMSE was used to assess the random error component after removing mean bias, and r^2 was used as a comparative indicator of model agreement across experiments with the same model structure but different feature sets. Although Random Forest is a nonlinear model and r^2 does not directly represent explained variance in a strict linear sense, it remains useful as a relative performance indicator for comparing alternative feature combinations within the same modeling framework [41]. The evaluation metrics were calculated as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \quad (2)$$

$$Bias = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i) \quad (3)$$

$$ubRMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[(\hat{y}_i - y_i) - Bias]^2} \quad (4)$$

$$r^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \quad (5)$$

where N is the total number of observations, and y_i is the observed value for the ith observation, $\hat{y}_i$ is the predicted value for the ith observation, and $\overline{y}$ is the mean of the actual observed values.

In addition to the model-based performance metrics, ISMN observations were used for site-level comparisons of the downscaled soil moisture product and for evaluating the

regional refinement step. For the regional refinement model, five-fold cross-validation was applied using the available ISMN observations, with four folds used for training and the remaining fold used for evaluation in each iteration. The reported refinement performance was calculated from the cross-validated predictions. Because the cross-validation was conducted at the observation level, repeated observations from the same monitoring station could occur in different folds; therefore, the resulting performance was interpreted as predictive refinement within the available ISMN station network rather than as evidence of transferability to entirely unseen stations or fully independent spatial–temporal generalization. In addition, because of the inherent scale mismatch between point-scale in situ measurements and gridded satellite soil moisture products, the ISMN-based comparisons were interpreted as assessments of site-level consistency and predictive refinement performance rather than strict one-to-one validation of the satellite-derived product.

3. Results

3.1. Contribution-Guided Input Selection for Multi-Stage Machine Learning Downscaling

To evaluate the input structure of the downscaling framework, SMAP brightness temperatures (TbH and TbV), LST, vegetation-related variables, and selected geoinformation variables were temporally and spatially aligned and merged into stacked feature datasets. Approximately 60,000 samples were extracted across the CONUS during the growing season (April–October 2025) at the 500 m/250 m analysis scale. The total sampling target was distributed approximately equally across the available observation dates, with samples randomly selected within the valid CONUS cultivated-land domain. Random Forest regression was repeatedly trained using different combinations of feature variables. These experiments were designed not only to compare relative model fit across feature configurations, but also to examine whether the selected input structure retained the central roles of the microwave and thermal variables that are most directly linked to the original SMAP retrieval.

Table 2 and Figure 4 summarize representative feature combinations and their corresponding model-fit statistics and feature contribution patterns. The feature-combination experiments revealed substantial differences in relative model fit depending on the chosen input variables. Although the full feature set yielded the highest comparative model fit (r^2 = 0.916), removing the microwave brightness temperatures (TbH and TbV) reduced comparative model fit markedly, and models relying mainly on optical, terrain, or soil-related features showed considerably weaker model fit. These results indicate that passive microwave brightness temperatures provide important retrieval-relevant information for downscaling soil moisture across the CONUS. However, model fit alone was not used as the sole criterion for final model input selection. Instead, the feature-combination experiments were treated as contribution-guided diagnostic comparisons of comparative model fit and feature-contribution structure. Particular attention was given to whether retrieval-relevant variables, especially TbV, TbH, and LST, remained central in the learned model. The r^2 values were therefore used as comparative diagnostic metrics across feature configurations rather than as independent validation of downscaling accuracy.

Among the tested combinations, the baseline model using only LST, TbV, and TbH produced balanced contributions from the three core variables, with each accounting for approximately one-third of the total importance. The addition of flow accumulation improved comparative model fit while preserving the strong contributions of TbV, TbH, and LST, indicating that hydrologic redistribution information could be meaningfully incorporated without substantially weakening the core microwave–thermal structure. Based on this result, the final direct input structure retained TbV, TbH, and LST as the core features and included flow accumulation as a supplementary input variable. In contrast, when

variables such as DEM, NDVI, or NDWI were added to the core feature set, they often gained relatively large feature contributions because of their strong spatial variability and fine-scale explanatory power. In several cases, these variables contributed more strongly than LST, and in some combinations even approached or exceeded the contribution of TbH. This pattern suggests that, although such features can enhance spatial detail, the resulting downscaled product may increasingly reflect terrain- or vegetation-driven image structure rather than the SMAP-consistent redistribution based on the original SMAP microwave signal.

Table 2. Representative feature combinations, corresponding comparative model-fit statistics, and relative feature contributions at the 500 m analysis scale across the CONUS during the growing season. Feature variables include LST, TbH, TbV, flow accumulation (flowAcc), slope, DEM, saturated soil water content (theta_s), NDVI, and NDWI.

Feature Variables	r^2	1st	2nd	3rd	4th	Etc.
LST, TbH, TbV	0.82	TbV (33.5)	LST (33.4)	TbH (33.1)	-	0.0
LST, TbH, TbV, Slope	0.85	TbV (35.2)	TbH (29.4)	LST (25.4)	Slope (9.9)	0.0
LST, TbH, TbV, flowAcc	0.86	TbV (33.5)	TbH (27.1)	LST (23.1)	flowAcc (16.3)	0.0
LST, TbH, TbV, DEM	0.89	TbV (31.7)	DEM (25.8)	TbH (24.7)	LST (17.9)	0.0
LST, TbH, TbV, theta_s	0.87	TbV (32.1)	TbH (24.4)	theta_s (22.3)	LST (21.2)	0.0
LST, TbH, TbV, NDVI	0.89	TbV (32.5)	TbH (25.0)	NDVI (24.1)	LST (18.3)	0.0
LST, TbH, TbV, flowAcc, Slope	0.88	TbV (29.1)	TbH (24.7)	LST (19.9)	flowAcc (17.8)	8.6
LST, TbH, TbV, flowAcc, DEM	0.91	TbV (27.5)	DEM (20.9)	TbH (20.9)	LST (15.7)	14.9
LST, TbH, TbV, flowAcc, theta_s	0.89	TbV (27.8)	TbH (21.4)	theta_s (20.0)	LST (18.1)	12.8
LST, TbH, TbV, flowAcc, theta_s, DEM, Slope	0.92	TbV (21.7)	TbH (17.5)	DEM (16.2)	theta_s (13.9)	30.6
LST, TbH, TbV, flowAcc, theta_s, DEM, Slope, NDWI	0.91	TbV (19.2)	TbH (14.9)	DEM (14.6)	NDWI (13.2)	38.1
LST, TbH, TbV, flowAcc, theta_s, DEM, Slope, NDVI, NDWI	0.92	TbV (21.4)	DEM (13.6)	TbH (13.2)	NDWI (11.2)	40.5

This contribution-based interpretation was especially important for vegetation-related features. Their direct inclusion as standard input features was not preferred because vegetation indices could exert contributions large enough to weaken the dominance of the core microwave–thermal features. At the same time, vegetation information could not be excluded entirely because it remains physically relevant within the tau–omega retrieval framework. Therefore, rather than allowing a vegetation-related variable to act as a directly dominant feature in a single global RF model, vegetation influences were incorporated adaptively through stratified RF modeling based on vegetation background conditions. Specifically, the vegetation conditions were stratified into five empirically defined ranges based on the statistical distribution of EVI across the study domain, and separate RF models were trained within each range. This data-distribution-guided stratification allowed vegetation-related physical context to be incorporated while reducing the risk that a single vegetation feature would dominate the overall downscaling model.

Overall, the feature contribution analysis indicated that the final feature structure for the main downscaling framework consisted of the core microwave–thermal variables (TbV, TbH, and LST), supplemented by flow accumulation (flowAcc) as the selected hydrologic redistribution feature, while vegetation-related influences were represented through adaptive stratified RF modeling rather than through a uniformly dominant vegetation feature. Flow accumulation was retained as an effective supplementary feature because it improved model performance without substantially weakening the central contributions of the core microwave–thermal variables. For the final vegetation-based stratification, EVI was used instead of NDVI because it showed similar overall contribution behavior while providing

slightly improved sensitivity under moderate to high vegetation cover, where NDVI tends to saturate. This final input structure was intended to retain the central contributions of the retrieval-relevant microwave and thermal variables while incorporating supplementary spatial information in a controlled way, thereby avoiding a downscaling result controlled primarily by high-spatial-resolution ancillary features.

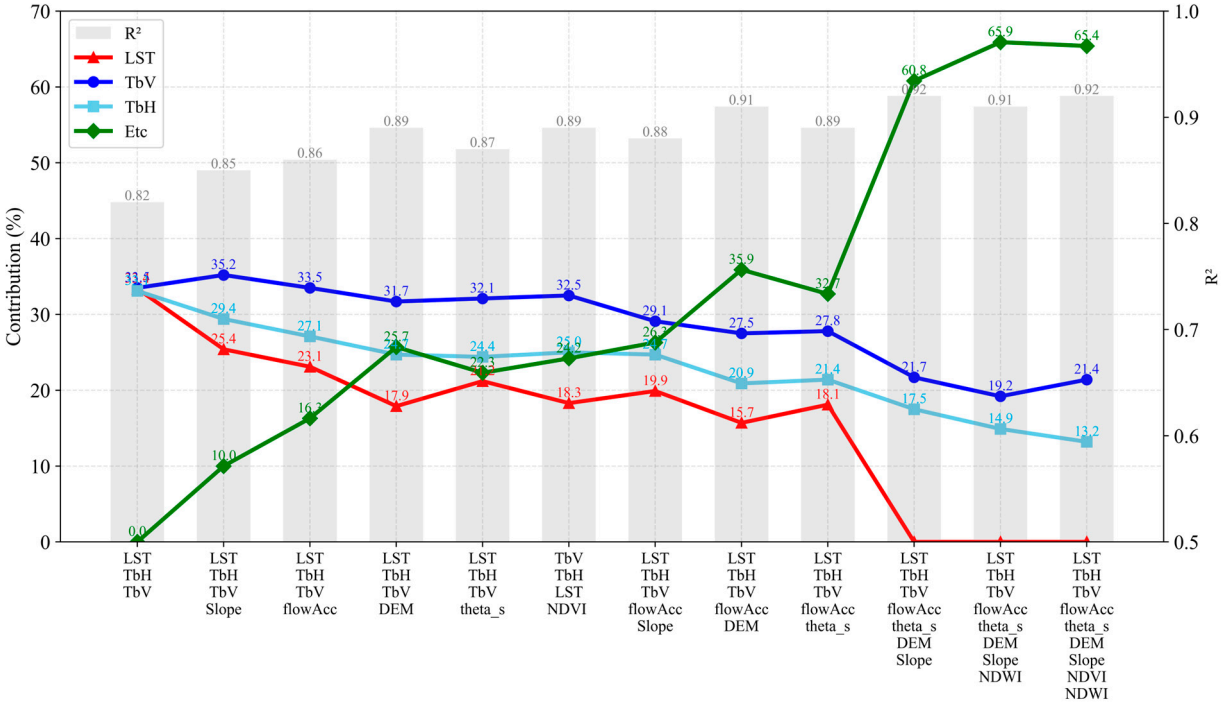


Figure 4. Feature contribution patterns across representative feature combinations, highlighting the relative importance of LST, TbV, TbH, and supplementary environmental features. In the x-axis labels, Core refers to the baseline feature set consisting of LST, TbH, and TbV. Higher contributions from supplementary variables indicate a greater risk that the model may be driven by ancillary spatial texture rather than retrieval-relevant microwave–thermal information.

These results demonstrate the added value of the multi-stage design, in which contribution-guided input selection retained the retrieval-relevant microwave and thermal variables as core predictors while ancillary information was incorporated in a controlled manner through adaptive stratification. The feature-contribution analysis is interpreted as a diagnostic of the learned model structure rather than as direct evidence of physical signal preservation; consistency with the SMAP Level-3 input signal is evaluated separately in Section 3.2.

3.2. Downscaled Soil Moisture Across CONUS

Using the final feature framework identified in Section 3.1, which retained TbV, TbH, and LST as core inputs and incorporated flow accumulation and EVI-based stratification, the first two stages of the downscaling framework were implemented across agricultural regions of the CONUS at 500 m and 250 m spatial resolutions. These stages were designed to enhance the spatial applicability of SMAP soil moisture while preserving consistency with the passive microwave retrieval signal. The 250 m and 500 m downscaled estimates were spatially aggregated to the 9 km SMAP Level-3 product grid and compared with the corresponding SMAP soil moisture. During April–October 2025, the aggregated 250 m product showed strong agreement with SMAP, with an r^2 of 0.885, an RMSE of 0.0286 m^3/m^3, an ubRMSE of 0.0282 m^3/m^3, a mean bias of +0.0050 m^3/m^3, and a correlation coefficient of 0.943. Comparable results were obtained for the aggregated 500 m product, with an

r^2 of 0.880, an RMSE of 0.0292 m^3/m^3, an ubRMSE of 0.0285 m^3/m^3, a mean bias of +0.0061 m^3/m^3, and a correlation coefficient of 0.942. These aggregation-based comparisons assess consistency with the input SMAP product rather than independent validation. Because the 250 m product did not provide a meaningful improvement in consistency with the input SMAP product relative to the 500 m product, the 500 m product was selected as the primary baseline for subsequent station-based evaluation and regional refinement. The 250 m product was retained to demonstrate implementation of the downscaling framework on a finer output grid.

Figure 5 presents representative examples of the original SMAP Level-3 soil moisture and the corresponding downscaled soil moisture on May 3 and May 4, 2025. Panels (A) and (B) show the original SMAP soil moisture for May 3 and May 4, respectively, whereas panels (C) and (D) show the corresponding estimated downscaled soil moisture. Compared with the original SMAP product, which presents broad regional moisture patterns because of its coarse spatial resolution, the downscaled product revealed finer spatial variability within agricultural landscapes. Finer-scale contrasts became more evident within SMAP product grid, particularly in areas where terrain, hydrologic redistribution, and vegetation background conditions varied over relatively short distances. These results indicate that the proposed multi-stage machine learning framework enhanced the spatial interpretability of soil moisture patterns while maintaining consistency with the broad wet–dry structure of the original SMAP observations. Although the overall agreement was strong, the downscaled product showed a tendency to underestimate the original SMAP soil moisture under wetter conditions above approximately 0.4 m^3/m^3. This pattern may reflect the combined effects of Random Forest smoothing behavior and the relatively limited representation of very wet conditions in the training samples. Nevertheless, the downscaled product more clearly resolved spatial variability under relatively dry conditions, which is particularly relevant for identifying drought-sensitive conditions in agricultural regions. Figure 6 shows scatterplots comparing SMAP soil moisture and downscaled soil moisture at the ISMN station locations for the SCAN and USCRN networks during April–October 2025. These comparisons were used to assess product-to-product consistency between the original SMAP and downscaled soil moisture products at the station locations, whereas direct comparison with in situ observations is presented in the following section. Overall, the downscaled product showed strong agreement with the original SMAP soil moisture, with coefficients of determination (r^2) of 0.773 and 0.687, RMSE values of 0.045 and 0.056 m^3/m^3, and unbiased RMSE (ubRMSE) values of 0.044 and 0.052 m^3/m^3 for the SCAN and USCRN networks, respectively. The slightly smaller sample sizes in Figure 7 than in Figure 6 result from the requirement for coincident valid in situ observations; Figure 6 requires only paired SMAP and downscaled values at the station locations. The slightly lower ubRMSE values indicate that part of the total error was associated with systematic bias, although the overall agreement remained strong. The scatterplots also suggest a slight tendency toward underestimation under very dry conditions below approximately 0.2 m^3/m^3 and at higher moisture levels above approximately 0.4 m^3/m^3. Despite these local deviations, the downscaled estimates preserved the regional-scale gradients present in the original SMAP product while redistributing soil moisture at finer spatial scales according to retrieval-relevant microwave–thermal information and supplementary hydrologic and vegetation-conditioned spatial controls. Overall, these results indicate that the proposed framework enhanced spatial detail for agricultural monitoring while maintaining broad consistency with the original passive microwave retrieval.

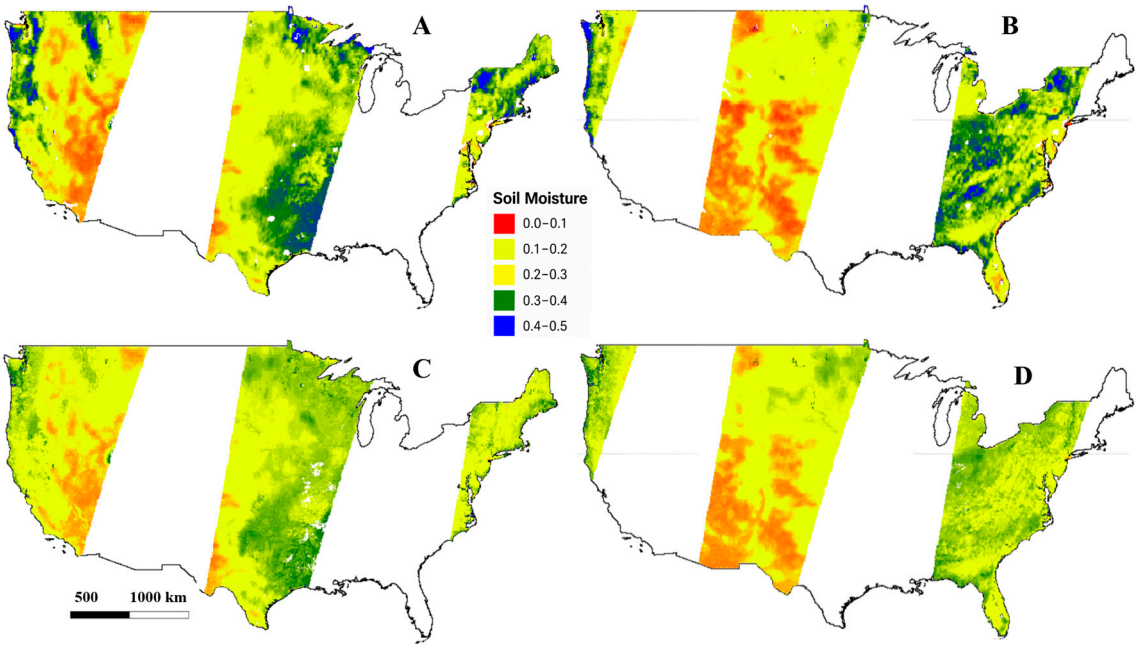


Figure 5. Example comparison of SMAP Level-3 soil moisture and downscaled soil moisture estimates. Panels show SMAP Level-3 soil moisture on (**A**) 3 May and (**B**) 4 May 2025, and downscaled soil moisture on (**C**) 3 May and (**D**) 4 May 2025. Spatial gaps mainly reflect the daily SMAP orbital coverage, as near-complete CONUS coverage is typically obtained over approximately two consecutive days.

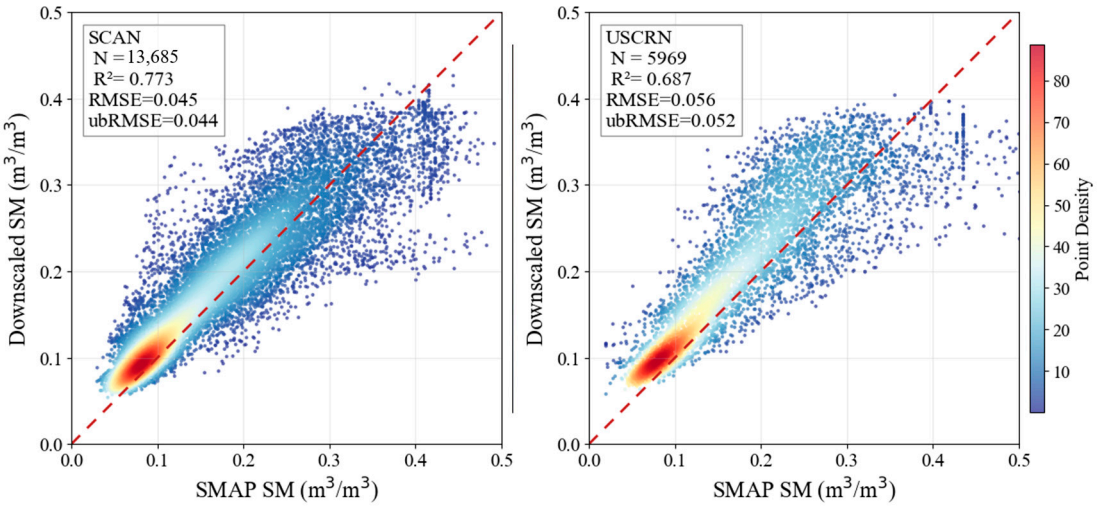


Figure 6. Scatterplots comparing SMAP soil moisture and downscaled soil moisture at the ISMN SCAN and USCRN stations during April–October 2025.

3.3. Evaluation with In Situ Observations and Regional Refinement

After the 500 m and 250 m SMAP downscaling stages, ISMN observations were used to evaluate site-level consistency and to support regional environmental refinement. Based on the comparable consistency of the two products with the 9 km SMAP Level-3 product, the 500 m product was used as the primary baseline for the subsequent station-based evaluation. First, scatterplots were constructed comparing the 500 m downscaled soil moisture with ISMN SCAN and USCRN observations during April–October 2025. Figure 7 is intended to evaluate the improvement from the baseline downscaled product to regional refinement within each network, rather than to characterize performance variability among individual stations; accordingly, the reported metrics represent pooled network-level

performance. The RMSE values were 0.093 and 0.099 m^3/m^3, and the ubRMSE values were 0.091 and 0.089 m^3/m^3 for the SCAN and USCRN networks, respectively. The lower ubRMSE values indicate that part of the total error was associated with systematic bias, particularly for the USCRN network. These results indicate moderate agreement with point-scale observations. This outcome is expected because the downscaled product remains a gridded satellite-based estimate, whereas ISMN measurements represent point-scale soil moisture conditions. Therefore, although the downscaled product provides enhanced spatial detail, it still represents a redistribution of the coarse SMAP passive microwave signal rather than an independent field-scale soil moisture measurement. The ISMN-based comparison should therefore be interpreted as a site-level consistency assessment rather than strict point-scale validation. Nevertheless, the downscaled soil moisture product on the finer grid provides a spatially explicit baseline field that can support further regional refinement using local environmental controls. In this role, the downscaled product is important not because it fully removes the limitations of the original SMAP retrieval, but because it converts the coarse passive microwave signal into a finer-resolution soil moisture field that can be further adjusted using regional hydro-topographic, soil, and vegetation-related information.

Using the direct comparison in Figure 7A,B as the baseline relationship, regional environmental refinement was evaluated to examine whether finer-scale local controls could improve the spatial applicability of the downscaled product for site-level agricultural monitoring. Using the 500 m downscaled SMAP soil moisture as the baseline, a 100 m refined soil moisture field was produced by combining the downscaled soil moisture with regional environmental drivers. Observed in situ soil moisture from ISMN was used as the target variable, while the downscaled soil moisture and regional environmental drivers were used as predictors. Accordingly, the 100 m product represents a station-supervised regional refinement of the SMAP-consistent downscaled soil-moisture field, rather than an independent field-data model or an additional SMAP downscaling level. A linear regional refinement model was first developed to retain physical interpretability. DEM and flow accumulation represented hydro-topographic effects on drainage, redistribution, and terrain-related conditions; EVI represented vegetation background; and θs represented soil water-holding capacity and near-surface moisture storage. These regional predictors provide additional hydro-topographic, soil-hydraulic, and vegetation-cover information. Static or temporally aggregated predictors do not generate temporal soil moisture dynamics but represent persistent spatial controls on drainage, storage, and moisture redistribution; temporal variability remains primarily inherited from the downscaled SMAP baseline. This refinement corresponds to Figure 7C,D and was used to evaluate whether physically meaningful regional variables could improve the site-level agreement while retaining interpretability. The slightly smaller sample sizes in panels C and D reflect complete-case filtering for the additional regional predictor variables used in the MLR analysis.

Compared with the baseline downscaled-SM relationship, the linear regional refinement improved model performance for both station networks. For SCAN, r^2 increased from 0.388 to 0.481, and RMSE decreased from 0.093 to 0.086 m^3/m^3. For USCRN, r^2 increased from 0.172 to 0.406, and RMSE decreased from 0.099 to 0.084 m^3/m^3. These improvements indicate that the SMAP-consistent downscaled SM product retained the primary SMAP-related moisture signal, while the additional regional physical variables helped explain local deviations between the downscaled SMAP-based product and point-scale observations. An additional interaction term between downscaled soil moisture and annual average EVI was also tested to examine whether vegetation condition modulated the relationship between the SMAP-based downscaled signal and in situ soil moisture. However, this interaction term was not statistically significant and did not provide meaningful im-

provement, indicating that the regional physical variables were more effective as additive refinement terms than as a vegetation-modulated interaction structure in this analysis.

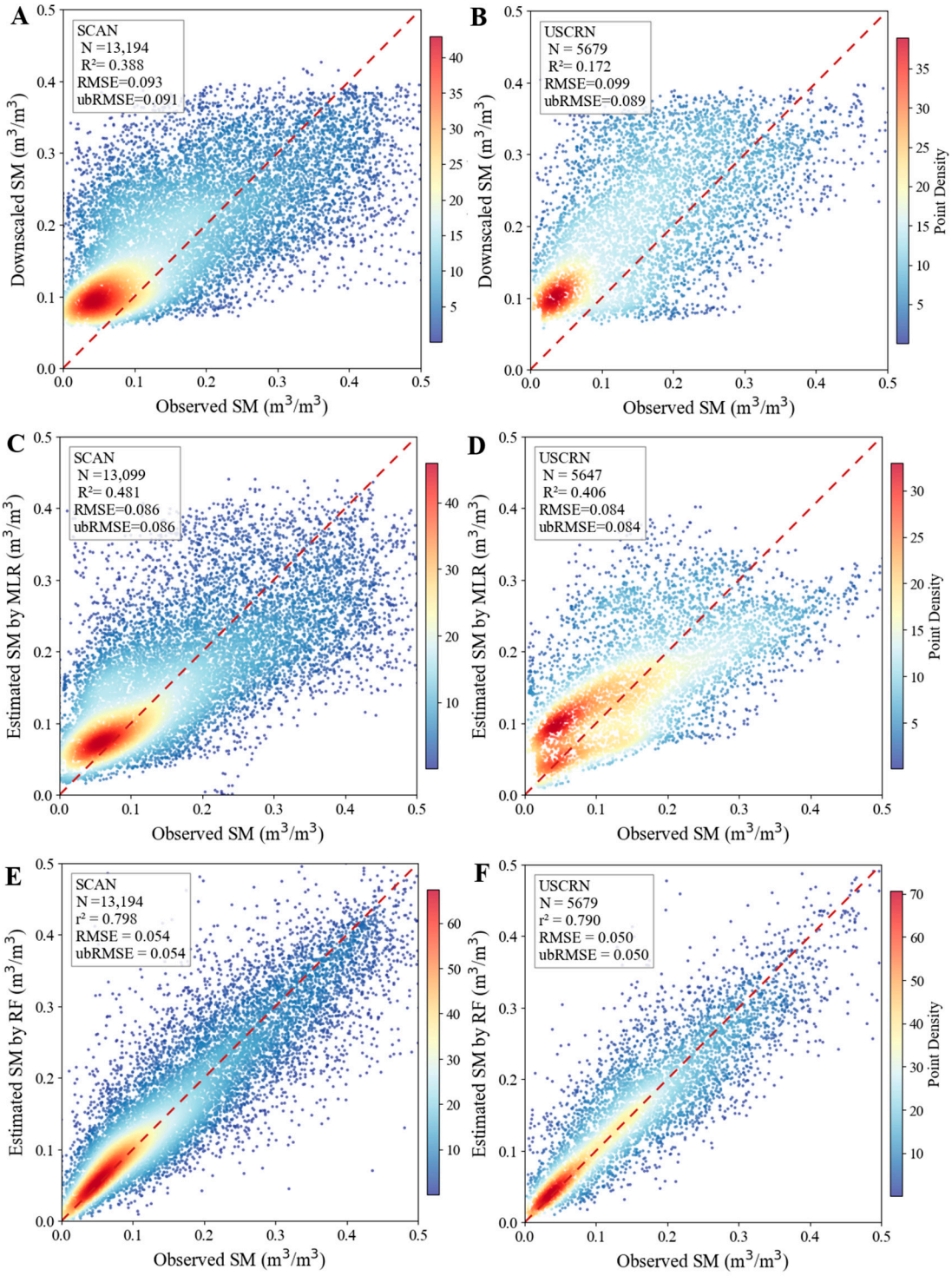


Figure 7. Scatterplots comparing soil moisture at the ISMN station locations during April–October 2025. Panels (**A**,**B**) show comparisons between observed in situ soil moisture and downscaled soil moisture for the SCAN and USCRN networks, respectively. Panels (**C**,**D**) show region-specific refinement results using multiple linear regression (MLR) for the corresponding station datasets. Panels (**E**,**F**) show Random Forest (RF)-based regional refinement results evaluated using five-fold cross-validation.

The standardized coefficients of the linear regional refinement provide physically interpretable information on the regional controls of field-scale soil moisture. The positive coefficient of downscaled SM was highly significant for both SCAN and USCRN ($p < 0.001$), confirming that the SMAP-consistent downscaled product retained the baseline SMAP-related moisture signal observed by the ISMN stations. Among the regional geoinformation variables, DEM showed significant negative coefficients in both SCAN and USCRN ($p < 0.001$), suggesting that higher-elevation areas were generally associated with lower observed soil moisture. This pattern may reflect elevation-related differences in drainage, regional climate, and terrain-controlled surface conditions. The positive coefficient of saturated soil water content (θs) was also highly significant for both networks ($p < 0.001$), indicating that soil water-holding capacity was an important control on local soil moisture. The effect of annual average EVI was network-dependent, with a significant negative coefficient for SCAN ($p < 0.001$) and a weak but significant positive coefficient for USCRN ($p = 0.026$). This suggests that vegetation condition may represent both moisture availability and vegetation water use depending on regional and network-specific conditions. Flow accumulation showed weaker or network-dependent behavior; it was not significant for SCAN ($p = 0.374$), but was significant for USCRN ($p < 0.001$), indicating that its effect may depend on local drainage representation, station location, and spatial scale. The additional interaction term between downscaled SM and annual EVI was not significant for either SCAN ($p = 0.280$) or USCRN ($p = 0.300$), and did not improve model performance. Therefore, the regional physical variables were more effective as additive refinement terms than as vegetation-modulated interaction terms in this analysis. Overall, these coefficient patterns indicate that the regional refinement did not simply apply a statistical correction to the downscaled product. Instead, it incorporated physically meaningful controls related to terrain-mediated thermal and drainage conditions, vegetation background, and soil water-holding capacity to adjust the SMAP-consistent downscaled soil moisture field toward site-level observations.

Using the same set of physically meaningful predictors, a Random Forest-based regional refinement was also implemented to examine whether nonlinear relationships among the downscaled soil moisture, terrain, vegetation, hydrologic redistribution, and soil water-holding variables could further improve predictive performance. This RF-based refinement is shown in Figure 7E,F. Unlike the linear refinement, which was mainly used for physical interpretation through standardized coefficients, the RF-based refinement was used to evaluate the predictive potential of the same regional environmental controls under a more flexible nonlinear model structure. The RF-based regional refinement achieved an overall five-fold cross-validated RMSE of approximately 0.052 m^3/m^3, with network-specific RMSE values of 0.054 m^3/m^3 for SCAN and 0.050 m^3/m^3 for USCRN. After regional refinement, the corresponding ubRMSE values were close to the RMSE values, indicating that the refinement reduced much of the systematic bias component observed in the direct downscaled-SM comparison. This suggests that regional refinement can partially mitigate broad-scale bias inherited from the original SMAP-based signal by incorporating region-specific hydro-topographic, soil, and vegetation-related controls. These results suggest that the main value of the multi-stage machine learning framework lies not in producing a direct point-scale soil moisture product, but in providing SMAP-consistent, spatially enhanced soil moisture baseline that can be further refined for site-level and regional agricultural applications.

Figure 8 presents an enlarged regional example corresponding to the yellow dashed box in Figure 1, showing the CDL, SMAP Level-3 soil moisture, downscaled soil moisture, and RF-based regionally refined soil moisture on 1 May 2025. While the original SMAP product represents broad moisture gradients at coarse spatial resolution, the downscaled

soil moisture resolves finer spatial heterogeneity across agricultural landscapes. The left and right zoomed views further illustrate finer-scale variability within local areas corresponding to approximately two SMAP product grids and one SMAP product grid, respectively. These regional examples indicate that the RF-based refinement more clearly incorporates region-specific hydro-topographic, soil, and vegetation-related controls within the downscaled SMAP soil moisture field. Overall, Figure 8 illustrates how the multi-stage machine learning framework can improve the spatial representation of SMAP-consistent soil moisture across heterogeneous agricultural landscapes.

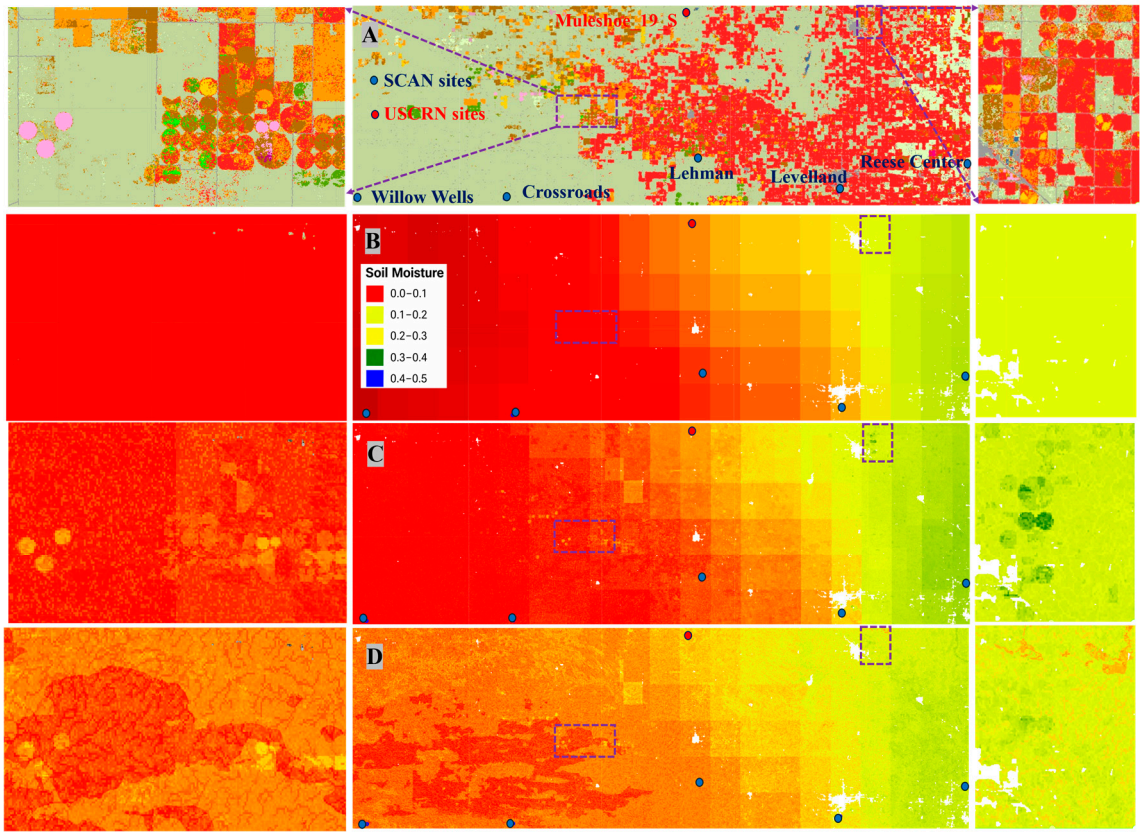


Figure 8. Enlarged regional view corresponding to the yellow dashed box in Figure 1, with an approximate center location of 33.70°N, 102.81°W, showing (**A**) the USDA NASS Cropland Data Layer (CDL), (**B**) SMAP Level-3 soil moisture, (**C**) downscaled soil moisture, and (**D**) RF-based regionally refined soil moisture on 1 May 2025. The central panels show the enlarged regional view, while the left and right zoomed views further enlarge local areas corresponding to approximately two SMAP Level-3 product-grid cells and one SMAP product-grid cell, respectively. The dashed boxes indicate the local areas shown in the corresponding zoomed panels, and the dashed connecting lines/arrows link each box to its enlarged view. Colors in panel (**A**) represent CDL land-cover classes, while colors in panels (**B**–**D**) represent soil moisture magnitude. Blue and red circles indicate SCAN and USCRN sites, respectively. The dashed boxes indicate the local areas shown in the corresponding zoomed panels, and the dashed connecting lines/arrows link each box to its enlarged view. Colors in panel (**A**) represent CDL land-cover classes, while colors in panels (**B**–**D**) represent soil moisture magnitude. Blue and red circles indicate SCAN and USCRN sites, respectively.

4. Discussion

4.1. Agricultural Applicability and Operational Potential

The novelty of the framework does not lie in the use of MODIS variables or Random Forest alone. Rather, it lies in diagnosing and limiting ancillary-variable dominance, retaining retrieval-relevant microwave–thermal variables as the core information source, incorporating vegetation information through adaptive stratification, and separating SMAP-

consistent disaggregation from ISMN-targeted regional refinement. Previous reviews have shown that many downscaling approaches gain apparent spatial detail by introducing high-resolution ancillary variables, but such gains can be sensitive to input assumptions, scale mismatch, and the physical relevance of auxiliary predictors [3]. In this context, the framework proposed here was designed so that retrieval-relevant microwave and thermal variables remained central, while supplementary variables were introduced in a controlled and interpretable way. From this perspective, the methodological contribution of this study is less about visual sharpening and more about establishing a retrieval-consistent pathway for spatial refinement. The added value of the proposed framework lies in explicitly controlling ancillary-variable dominance while retaining retrieval-relevant microwave and thermal variables as the core information source. By combining contribution-guided input selection, adaptive stratification, and post-downscaling regional refinement, the multi-stage strategy is designed to balance finer-grid spatial representation, predictive performance, and retrieval-consistent physical interpretability rather than optimizing predictive accuracy alone.

This distinction is especially important for agricultural applications. SMAP is valuable because it provides near-global surface soil moisture observations at frequent revisit intervals, and the original mission concept itself was intended to improve both temporal and spatial utility through passive–active microwave synergy [6,17]. For managed agricultural landscapes, however, temporal frequency alone is insufficient; the product must also become spatially relevant to field-to-field and within-region variability. Previous work on CONUS crop field sizes and field-scale agricultural soil moisture has shown that meaningful management-relevant heterogeneity often occurs at scales far below the SMAP product grid [10,42–44]. The practical agricultural value of the present framework, therefore, is not that it produces a direct field-truth product, but that it converts each temporally frequent SMAP retrieval into a more spatially interpretable moisture surface across heterogeneous agricultural regions. The preferred 500 m and 250 m implementations should also be interpreted as more than computational compromises. Their value lies in the fact that they provide a practical and retrieval-consistent balance between finer spatial representation and the scale support of the predictor variables, particularly the thermal information used in the framework. In this sense, these resolutions are computationally practical across the CONUS-wide analysis and remain conceptually appropriate for retaining a meaningful connection to the original microwave–thermal signal. Although the 100 m product was generated and was useful for finer-scale regional refinement, increasingly fine output grids do not automatically imply more meaningful soil moisture estimates when the scale support of key predictor variables remains coarser than the target resolution [45,46]. Accordingly, the goal of the present framework is not unlimited sharpening, but repeatable and retrieval-consistent refinement.

A further practical strength of the framework is its operational repeatability. Once the multi-stage downscaling structure has been established, each newly available SMAP Level-3 product can be converted into an updated, spatially enhanced soil moisture field using the same workflow. The resulting downscaled product can then serve as the spatial baseline for additional regional refinement using local hydro-topographic, soil, and vegetation information. This regional refinement is important because environmental controls on soil moisture and agricultural productivity vary substantially across the CONUS, and regionalized modeling can improve the representation of heterogeneous hydroclimatic, topographic, soil, and vegetation conditions [47]. In this way, the framework functions not only as a downscaling method, but also as an operational pathway for repeatedly updated and regionally improved soil moisture estimation across agricultural landscapes. Although the present implementation was demonstrated over the CONUS, the overall architecture

is not intrinsically limited to that region. Because it is built on routinely available SMAP and ancillary remote-sensing products, the same general framework could, in principle, be extended more broadly if sufficient storage, memory, and cloud-computing resources are available [31].

The results of the regional refinement further support this operational interpretation. The downscaled SMAP soil moisture first provided a SMAP-consistent baseline moisture field, and additional regional environmental controls were then used to adjust this baseline toward site-level observations. The linear refinement showed that adding regional geoinformation variables representing hydro-topographic controls, vegetation background, and soil water-holding capacity improved agreement with ISMN observations compared with using the downscaled soil moisture alone. This result indicates that the downscaled product can serve as a practical intermediate soil moisture layer rather than a final independent field-scale retrieval. The standardized coefficients also supported physical interpretability by showing that terrain, vegetation background, and soil water-holding capacity contributed to local deviations from the downscaled SMAP-based signal. In contrast, the RF-based refinement demonstrated the predictive potential of the same regional environmental controls under a nonlinear model structure. Thus, the framework should be understood as a multi-stage process consisting of core RF downscaling, adaptive stratified modeling, and regional ML refinement. It is important to clarify that this study does not directly validate specific agricultural decision-support applications, such as irrigation scheduling, drought monitoring, or crop water stress assessment. Rather, the main contribution is to provide a finer-resolution, SMAP-consistent soil moisture baseline that can be integrated with other agricultural and environmental datasets. In practical agricultural monitoring, soil moisture is commonly interpreted together with precipitation, evapotranspiration, evaporative stress index, vegetation condition, crop type, and long-term monitoring records. Therefore, the downscaled and regionally refined SMAP soil moisture product should be regarded as an improved input layer for future agricultural monitoring, while application-specific validation remains necessary.

Figure 9 further illustrates the operational behavior of the proposed framework through temporal comparisons between measured ISMN soil moisture and the downscaled soil moisture at three representative ISMN SCAN stations during April–October 2025. The selected stations represent contrasting agricultural and environmental settings, including Busby_Farm in central Missouri (approximately 183 m above sea level; humid continental–humid subtropical transition), Reese_Center in the Texas High Plains (approximately 1015 m; semi-arid), and Princeton_#1 in western Kentucky (approximately 183 m; humid subtropical). Across these contrasting settings, the downscaled product generally followed the temporal behavior of the measured ISMN observations while providing a more spatially interpretable soil moisture surface that may support agricultural monitoring when combined with other environmental information. In particular, the temporal trend at Reese_Center was captured relatively well except during autumn, whereas Busby_Farm and Princeton_#1 showed a tendency toward slight underestimation from spring to summer. These examples suggest that the proposed framework can be repeatedly applied as new SMAP Level-3 data become available, thereby supporting regionally relevant agricultural monitoring rather than serving only as a one-time downscaling experiment. At the same time, these examples also highlight several remaining limitations. The comparisons shown in Figure 9 should not be interpreted as strict point-scale validation, because both SMAP and the derived downscaled product originate from a coarse passive microwave footprint and therefore remain affected by spatial scale mismatch relative to local in situ conditions [4,42]. In addition, temporal discrepancies under certain seasonal or wetting conditions indicate that short-term precipitation effects, within-grid spatial heterogeneity,

and locally varying land-surface controls are not yet fully resolved within the present framework. Residual temporal offsets between SMAP and LST may further contribute to uncertainty during rapid wetting or drying events. Although these limitations do not preclude the operational application of the proposed framework, they indicate the need for further refinement to improve robustness across varying environmental and temporal conditions. These limitations and the corresponding future directions are discussed in the following section.

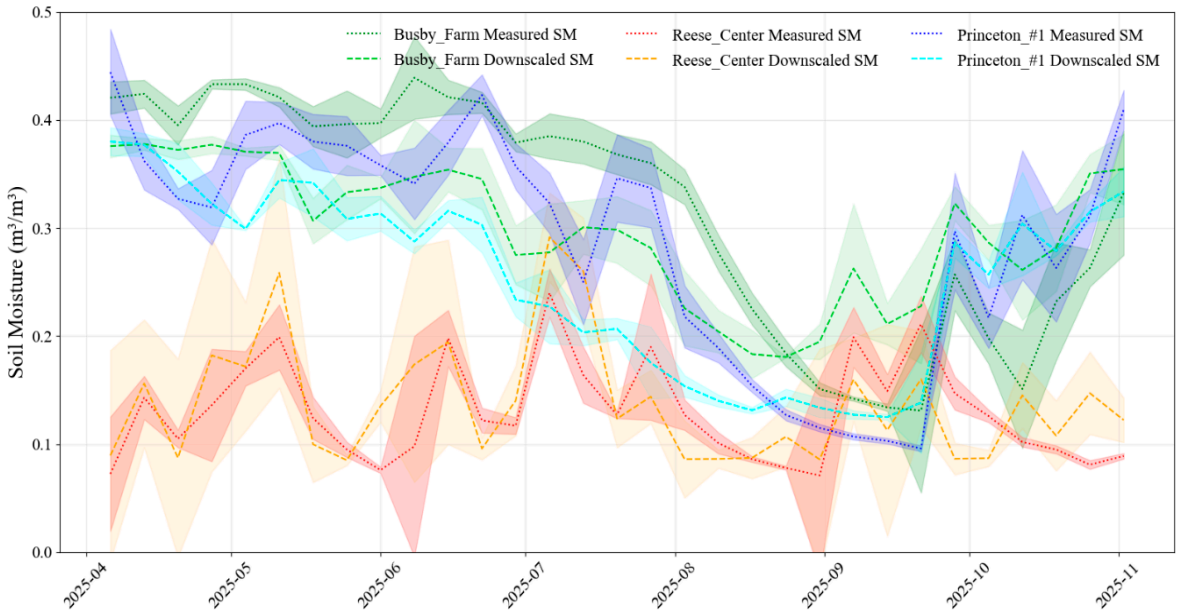


Figure 9. Temporal comparison between measured ISMN soil moisture and downscaled soil moisture estimates at three representative ISMN SCAN stations during April–October 2025: Busby_Farm (38.50°N, 92.25°W), Reese_Center (33.62°N, 102.04°W), and Princeton_#1 (37.10°N, 87.83°W). The stations represent contrasting environmental settings in central Missouri, the Texas High Plains, and western Kentucky, respectively. Dotted lines indicate weekly averages, and error bars represent the standard deviation within each weekly period.

4.2. Remaining Limitations and Future Directions

The proposed framework was effective in representing broad background soil moisture variability and in improving the spatial applicability of SMAP soil moisture across agricultural regions. At the same time, the results suggest that its sensitivity is relatively lower for rapid wetting signals associated with recent precipitation events or other short-term land-surface changes. In some cases, this appeared as a tendency to underestimate wet conditions or to smooth event-scale increases in soil moisture. This behavior is physically understandable because precipitation is among the most direct controls on short-term soil moisture variability, yet it was not included here as an explicit predictor. The SMAP target already contains the wetting response observed at the satellite overpass time, and this temporal signal is inherited by the downscaled products. However, without explicit precipitation information, the framework cannot resolve sub-footprint precipitation heterogeneity or fully represent rapid localized wetting events that are not adequately captured by the coarse SMAP observation. The framework was intentionally designed around sensing-based variables that remain closely linked to the original passive microwave retrieval. As a result, the current approach is better suited to retrieval-consistent spatial refinement of background moisture conditions than to fully resolving rapid event-driven wetting dynamics. Incorporating precipitation directly would shift the framework toward a more forcing-driven soil moisture modeling approach rather than retrieval-consistent

downscaling of SMAP Level-3 observations. This also means that the present product should not be interpreted as a standalone agricultural decision-support product for irrigation scheduling, drought monitoring, or crop water stress assessment without additional application-specific validation.

In this respect, future research will investigate precipitation-informed refinement to improve the representation of rapid wetting events while maintaining the downscaled SMAP product as the retrieval-consistent baseline. Model-based products such as SMAP Level-4 may provide useful guidance because they incorporate meteorological forcing and land-surface model dynamics in ways that are more responsive to precipitation-driven variability [9,48,49]. However, such products differ fundamentally from the retrieval-based SMAP Level-3 framework used in this study, because they depend more strongly on model structure, forcing inputs, and parameterization. A similar issue applies to other model-based variables, such as evapotranspiration and land-surface model outputs. Although these variables may improve predictive performance, they are not direct sensing-based observations and therefore introduce additional assumptions that may reduce physical interpretability and regional consistency [3,11]. For this reason, the present study focused intentionally on sensing-based predictors to preserve stronger consistency with the original passive microwave signal, while leaving precipitation-informed or model-assisted adjustment as an important direction for future work. For agricultural applications, future refinement should integrate the downscaled SMAP soil moisture with precipitation, evapotranspiration, evaporative stress index, vegetation condition, crop type, and long-term monitoring records to evaluate specific management outcomes.

A second limitation concerns the tradeoff between predictive gain and retrieval consistency in predictor selection. Geoinformation variables, including terrain and soil-related properties, are clearly important in agricultural landscapes, but their strong and often static spatial structure can dominate machine-learning models if they are introduced too directly. In such cases, the final product may increasingly reflect terrain-, soil-, or vegetation-driven image structure rather than the SMAP-consistent redistribution of the original microwave soil moisture signal. Vegetation-related variables pose a similar challenge. Because vegetation indices reflect canopy condition, phenology, and integrated response to water availability, they are physically relevant but can also become statistically dominant when used as direct predictors in a single global model. The present framework addressed this issue by using contribution-guided selection and by shifting vegetation-related information into an adaptive stratified RF design. More broadly, this result suggests that the main challenge in multi-stage machine learning downscaling is not simply identifying more predictors, but incorporating them in ways that do not overwhelm the retrieval-relevant microwave–thermal structure [3,23,50]. However, the physical considerations incorporated in this study should not be interpreted as a hard physical constraint imposed on the RF model structure, loss function, or output range. Rather, they were implemented through retrieval-relevant feature selection, temporal matching, contribution-guided screening, adaptive stratified RF modeling, and interpretable regional ML refinement. Therefore, the proposed framework is best described as a multi-stage machine learning downscaling approach that incorporates physical considerations, not as a fully physically constrained retrieval model.

These limitations point directly to future directions. One important direction is improved spatial integration at finer regional scales, especially where downscaled SMAP soil moisture can be combined with locally relevant hydro-topographic, soil, and vegetation information in a more adaptive way. Another important direction is interoperability with SAR observations. The original SMAP mission concept itself was based on the complementary use of passive and active microwave observations to improve both temporal and

spatial usefulness, and that broader idea remains highly relevant for agricultural applications [6,17]. In this context, C-band SAR, such as Sentinel-1, offers practical opportunities for regional monitoring, but its shorter wavelength makes it more sensitive to upper canopy and leaf-layer conditions, which can limit its direct sensitivity to subsurface moisture variability in densely vegetated agricultural settings [51–53]. In contrast, L-band systems are generally more favorable for agricultural soil moisture applications because their longer wavelength penetrates more deeply through the vegetation canopy, making them more responsive to broader vegetation structure as well as underlying soil moisture conditions. This makes future interoperability with missions such as NISAR particularly promising [54]. At the same time, SAR-based enhancement must still be applied with caution, because radar backscatter is influenced not only by soil moisture, but also by vegetation structure, surface roughness, incidence angle, local viewing geometry, and precipitation-induced surface wetness, including water films or droplets on vegetation and soil surfaces [53,55]. Temporal mismatch between passive and active microwave observations further complicates this issue, because the two systems may respond to different moisture and recovery conditions. Therefore, SAR integration should be developed as a retrieval-consistent complement to SMAP-based downscaling rather than as a simple high-resolution replacement.

A separate but directly relevant future direction is the use of thermal observations with higher spatial and temporal resolution. Because LST played a central role in this study as a retrieval-relevant thermal variable, forthcoming thermal observations from missions such as Landsat Next and the ESA Copernicus Land Surface Temperature Monitoring (LSTM) mission could directly improve the operational applicability of microwave–thermal soil moisture downscaling [56,57]. Landsat Next is being developed to improve the spatial, temporal, and spectral continuity of Landsat observations, while LSTM is designed to provide frequent field-scale land surface temperature and evapotranspiration information for applications such as agricultural water management. In addition, future agriculture-oriented thermal SmallSat concepts could provide targeted LST observations over priority agricultural regions, potentially improving the temporal consistency between SMAP retrievals and thermal inputs when observations are scheduled within comparable overpass windows. Rather than serving as a replacement for broad-coverage thermal sensors, such platforms could complement existing satellite products by providing more frequent or higher-resolution thermal information over drought-sensitive agricultural areas. However, such thermal integration would also require careful treatment of overpass-time differences, cloud contamination, calibration consistency, and scale mismatch between thermal and microwave observations. Future progress will therefore depend not simply on adding higher-resolution SAR or thermal information, but on developing physically consistent multi-sensor frameworks that preserve the retrieval-consistent structure of SMAP while improving sensitivity to local agricultural variability and enhancing the operational use of soil moisture information for drought-stress detection and climate-resilient agricultural decision-making.

5. Conclusions

This study demonstrated that SMAP soil moisture can be spatially refined using a multi-stage machine learning framework while preserving consistency with the original passive microwave retrieval. By maintaining the central role of retrieval-relevant microwave and thermal information, the proposed framework enhanced the spatial representation of SMAP soil moisture while preserving the broad wet–dry patterns of the original SMAP Level-3 passive microwave product. The downscaled soil moisture product provided a finer-resolution spatial baseline across the CONUS and supported further regional refinement using local hydro-topographic, soil, and vegetation-related information. This

multi-stage structure is important because it recognizes both the value and the limitation of satellite-based soil moisture: SMAP provides temporally consistent large-scale information, while regional environmental features can help improve site-level applicability. Overall, the proposed framework offers a practical pathway for enhancing the spatial interpretability and potential agricultural utility of SMAP soil moisture products. Future work should focus on incorporating precipitation effects, developing physically consistent multi-sensor approaches that integrate passive microwave observations with active microwave sensor data, and evaluating the application-specific value of the resulting SMAP-consistent soil moisture baseline using irrigation, drought, crop water stress, or yield-response information.

Author Contributions: Conceptualization, J.P. and J.G.C.; Methodology, J.P. and J.G.C.; Software, J.P. and J.G.C.; Validation, J.P. and J.G.C.; Formal analysis, J.P. and J.G.C.; Investigation, J.P. and J.G.C.; Resources, J.P. and J.G.C.; Data curation, J.P. and J.G.C.; Writing—original draft, J.P. and J.G.C.; Writing—review & editing, J.P. and J.G.C.; Visualization, J.G.C.; Supervision, J.P. and J.G.C.; Project administration, J.P. and J.G.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Geo-Informatics Environment & Agriculture Bridging AI under project number GEO-SM-2026-001.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ochsner, T.E.; Cosh, M.H.; Cuenca, R.H.; Dorigo, W.A.; Draper, C.S.; Hagimoto, Y.; Kerr, Y.H.; Larson, K.M.; Njoku, E.G.; Small, E.E. State of the Art in Large-scale Soil Moisture Monitoring. *Soil Sci. Soc. Am. J.* **2013**, *77*, 1888–1919. [CrossRef]
2. Seneviratne, S.I.; Corti, T.; Davin, E.L.; Hirschi, M.; Jaeger, E.B.; Lehner, I.; Orlowsky, B.; Teuling, A.J. Investigating Soil Moisture–Climate Interactions in a Changing Climate: A Review. *Earth-Sci. Rev.* **2010**, *99*, 125–161. [CrossRef]
3. Peng, J.; Loew, A.; Merlin, O.; Verhoest, N.E.C. A Review of Spatial Downscaling of Satellite Remotely Sensed Soil Moisture. *Rev. Geophys.* **2017**, *55*, 341–366. [CrossRef]
4. Crow, W.T.; Berg, A.A.; Cosh, M.H.; Loew, A.; Mohanty, B.P.; Panciera, R.; De Rosnay, P.; Ryu, D.; Walker, J.P. Upscaling Sparse Ground-based Soil Moisture Observations for the Validation of Coarse-resolution Satellite Soil Moisture Products. *Rev. Geophys.* **2012**, *50*, RG2002. [CrossRef]
5. Cosh, M.H.; Jackson, T.J.; Moran, S.; Bindlish, R. Temporal Persistence and Stability of Surface Soil Moisture in a Semi-Arid Watershed. *Remote Sens. Environ.* **2008**, *112*, 304–313. [CrossRef]
6. Entekhabi, D.; Njoku, E.G.; O'Neill, P.E.; Kellogg, K.H.; Crow, W.T.; Edelstein, W.N.; Entin, J.K.; Goodman, S.D.; Jackson, T.J.; Johnson, J. The Soil Moisture Active Passive (SMAP) Mission. *Proc. IEEE* **2010**, *98*, 704–716. [CrossRef]
7. Vereecken, H.; Huisman, J.A.; Pachepsky, Y.; Montzka, C.; Van Der Kruk, J.; Bogena, H.; Weihermüller, L.; Herbst, M.; Martinez, G.; Vanderborght, J. On the Spatio-Temporal Dynamics of Soil Moisture at the Field Scale. *J. Hydrol.* **2014**, *516*, 76–96. [CrossRef]
8. Njoku, E.G.; Entekhabi, D. Passive Microwave Remote Sensing of Soil Moisture. *J. Hydrol.* **1996**, *184*, 101–129. [CrossRef]
9. Enrekhabi, D.; Yueh, S.; O'Neil, P.E.; Kellogg, K.H.; Allen, A.; Bindlish, R.; Brown, M.E.; Chan, S.T.K.; Colliander, A.; Crow, W.T.; et al. *SMAP Handbook–Soil Moisture Active Passive: Mapping Soil Moisture and Freeze/Thaw from Space*; California Institute of Technology: Pasadena, CA, USA, 2014; p. 192.
10. Zappa, L.; Forkel, M.; Xaver, A.; Dorigo, W. Deriving Field Scale Soil Moisture from Satellite Observations and Ground Measurements in a Hilly Agricultural Region. *Remote Sens.* **2019**, *11*, 2596. [CrossRef]
11. Djamai, N.; Magagi, R.; Goïta, K.; Merlin, O.; Kerr, Y.; Roy, A. A Combination of DISPATCH Downscaling Algorithm with CLASS Land Surface Scheme for Soil Moisture Estimation at Fine Scale during Cloudy Days. *Remote Sens. Environ.* **2016**, *184*, 1–14. [CrossRef]
12. Im, J.; Park, S.; Rhee, J.; Baik, J.; Choi, M. Downscaling of AMSR-E Soil Moisture with MODIS Products Using Machine Learning Approaches. *Environ. Earth Sci.* **2016**, *75*, 1120. [CrossRef]
13. Abbaszadeh, P.; Moradkhani, H.; Zhan, X. Downscaling SMAP Radiometer Soil Moisture over the CONUS Using an Ensemble Learning Method. *Water Resour. Res.* **2019**, *55*, 324–344. [CrossRef]

14. Wu, X.; Walker, J.P.; Das, N.N.; Panciera, R.; Rudiger, C. Evaluation of the SMAP Brightness Temperature Downscaling Algorithm Using Active-Passive Microwave Observations. *Remote Sens. Environ.* **2014**, *155*, 210–221. [CrossRef]
15. Chauhan, N.S.; Miller, S.; Ardanuy, P. Spaceborne Soil Moisture Estimation at High Resolution: A Microwave-Optical/IR Synergistic Approach. *Int. J. Remote Sens.* **2003**, *24*, 4599–4622. [CrossRef]
16. Merlin, O.; Chehbouni, A.; Walker, J.P.; Panciera, R.; Kerr, Y.H. A Simple Method to Disaggregate Passive Microwave-Based Soil Moisture. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 786–796. [CrossRef]
17. Das, N.N.; Entekhabi, D.; Njoku, E.G. An Algorithm for Merging SMAP Radiometer and Radar Data for High-Resolution Soil-Moisture Retrieval. *IEEE Trans. Geosci. Remote Sens.* **2010**, *49*, 1504–1512. [CrossRef]
18. Busch, F.A.; Niemann, J.D.; Coleman, M. Evaluation of an Empirical Orthogonal Function–Based Method to Downscale Soil Moisture Patterns Based on Topographical Attributes. *Hydrol. Process.* **2012**, *26*, 2696–2709. [CrossRef]
19. Dandridge, C.; Fang, B.; Lakshmi, V. Downscaling of SMAP Soil Moisture in the Lower Mekong River Basin. *Water* **2019**, *12*, 56. [CrossRef]
20. Fang, L.; Zhan, X.; Yin, J.; Liu, J.; Schull, M.; Walker, J.P.; Wen, J.; Cosh, M.H.; Lakhankar, T.; Collins, C.H. An Intercomparison Study of Algorithms for Downscaling SMAP Radiometer Soil Moisture Retrievals. *J. Hydrometeorol.* **2020**, *21*, 1761–1775. [CrossRef]
21. Xu, J.; Su, Q.; Li, X.; Ma, J.; Song, W.; Zhang, L.; Su, X. A Spatial Downscaling Framework for SMAP Soil Moisture Based on Stacking Strategy. *Remote Sens.* **2024**, *16*, 200. [CrossRef]
22. Ghafari, E.; Walker, J.P.; Zhu, L.; Colliander, A.; Faridhosseini, A. Spatial Downscaling of SMAP Radiometer Soil Moisture Using Radar Data: Application of Machine Learning to the SMAPEx and SMAPVEX Campaigns. *Sci. Remote Sens.* **2024**, *9*, 100122. [CrossRef]
23. Jackson, T.J.; Schmugge, T.J. Vegetation Effects on the Microwave Emission of Soils. *Remote Sens. Environ.* **1991**, *36*, 203–212. [CrossRef]
24. Loveland, T.R.; Merchant, J.M. Ecoregions and Ecoregionalization: Geographical and Ecological Perspectives. *Environ. Manag.* **2004**, *34*, S1–S13. [CrossRef] [PubMed]
25. U.S. Department of Agriculture USDA National Agricultural Statistics Service. Available online: https://quickstats.nass.usda.gov/ (accessed on 4 April 2026).
26. Chaney, N.W.; Minasny, B.; Herman, J.D.; Nauman, T.W.; Brungard, C.W.; Morgan, C.L.S.; McBratney, A.B.; Wood, E.F.; Yimam, Y. POLARIS Soil Properties: 30-m Probabilistic Maps of Soil Properties over the Contiguous United States. *Water Resour. Res.* **2019**, *55*, 2916–2938. [CrossRef]
27. CEC. *Ecological Regions of North America: Toward a Common Perspective*; CEC: Montreal, QC, Canada, 1997; ISBN 292230518X.
28. Dorigo, W.; Himmelbauer, I.; Aberer, D.; Schremmer, L.; Petrakovic, I.; Zappa, L.; Preimesberger, W.; Xaver, A.; Annor, F.; Ardö, J. The International Soil Moisture Network: Serving Earth System Science for over a Decade. *Hydrol. Earth Syst. Sci. Discuss.* **2021**, *2021*, 1–83. [CrossRef]
29. Schaefer, G.L.; Cosh, M.H.; Jackson, T.J. The USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN). *J. Atmos. Ocean. Technol.* **2007**, *24*, 2073–2077. [CrossRef]
30. Bell, J.E.; Palecki, M.A.; Baker, C.B.; Collins, W.G.; Lawrimore, J.H.; Leeper, R.D.; Hall, M.E.; Kochendorfer, J.; Meyers, T.P.; Wilson, T. US Climate Reference Network Soil Moisture and Temperature Observations. *J. Hydrometeorol.* **2013**, *14*, 977–988. [CrossRef]
31. ONeill, P.; Chan, S.; Njoku, E.; Jackson, T.; Bindlish, R.; Chaubell, J.; Colliander, A. SMAP Enhanced L3 Radiometer Global and Polar Grid Daily 9 Km EASE-Grid Soil Moisture, Version 6. NASA National Snow and Ice Data Center Distributed Active Archive Center. Data Set. 2023 M20OXIZHY3RJ. Available online: https://data.nasa.gov/dataset/smap-enhanced-l3-radiometer-global-and-polar-grid-daily-9-km-ease-grid-soil-moisture-v006 (accessed on 4 April 2026).
32. Didan, K.; Munoz, A.B.; Solano, R.; Huete, A. MODIS Vegetation Index User's Guide (MOD13 Series). *Univ. Ariz. Veg. Index Phenol. Lab* **2015**, *35*, 2–33.
33. Gao, B.-C. NDWI—A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space. *Remote Sens. Environ.* **1996**, *58*, 257–266. [CrossRef]
34. Carlson, T. An Overview of the" Triangle Method" for Estimating Surface Evapotranspiration and Soil Moisture from Satellite Imagery. *Sensors* **2007**, *7*, 1612–1629. [CrossRef]
35. Carlson, T.N.; Gillies, R.R.; Perry, E.M. A Method to Make Use of Thermal Infrared Temperature and NDVI Measurements to Infer Surface Soil Water Content and Fractional Vegetation Cover. *Remote Sens. Rev.* **1994**, *9*, 161–173. [CrossRef]
36. Chang, J.G.; Anderson, M.; Gao, F.; Russ, A.; Zhao, H.; Cirone, R.; Pachepsky, Y.; Johnson, D.M. Hydro-Topographic Contribution to In-Field Crop Yield Variation Using High-Resolution Surface and GPR-Derived Subsurface DEMs. *Remote Sens.* **2025**, *17*, 3061. [CrossRef]
37. Crippen, R.; Buckley, S.; Agram, P.; Belz, E.; Gurrola, E.; Hensley, S.; Kobrick, M.; Lavalle, M.; Martin, J.; Neumann, M. NASADEM Global Elevation Model: Methods and Progress. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *41*, 125–128. [CrossRef]

38. Amatulli, G.; Garcia Marquez, J.; Sethi, T.; Kiesel, J.; Grigoropoulou, A.; Üblacker, M.M.; Shen, L.Q.; Domisch, S. Hydrography90m: A New High-Resolution Global Hydrographic Dataset. *Earth Syst. Sci. Data* **2022**, *14*, 4525–4550. [CrossRef]
39. Louppe, G. Understanding Random Forests: From Theory to Practice. *arXiv* **2014**, arXiv:1407.7502v3.
40. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
41. Spiess, A.-N.; Neumeyer, N. An Evaluation of R 2 as an Inadequate Measure for Nonlinear Models in Pharmacological and Biochemical Research: A Monte Carlo Approach. *BMC Pharmacol.* **2010**, *10*, 1–11. [CrossRef] [PubMed]
42. Colliander, A.; Jackson, T.J.; Chan, S.K.; O'Neill, P.; Bindlish, R.; Cosh, M.H.; Caldwell, T.; Walker, J.P.; Berg, A.; McNairn, H. An Assessment of the Differences between Spatial Resolution and Grid Size for the SMAP Enhanced Soil Moisture Product over Homogeneous Sites. *Remote Sens. Environ.* **2018**, *207*, 65–70. [CrossRef]
43. Vergopolan, N.; Xiong, S.; Estes, L.; Wanders, N.; Chaney, N.W.; Wood, E.F.; Konar, M.; Caylor, K.; Beck, H.E.; Gatti, N. Field-Scale Soil Moisture Bridges the Spatial-Scale Gap between Drought Monitoring and Agricultural Yields. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 1827–1847. [CrossRef]
44. Yan, L.; Roy, D.P. Conterminous United States Crop Field Size Quantification from Multi-Temporal Landsat Data. *Remote Sens. Environ.* **2016**, *172*, 67–86. [CrossRef]
45. Anderson, M.C.; Norman, J.M.; Kustas, W.P.; Houborg, R.; Starks, P.J.; Agam, N. A Thermal-Based Remote Sensing Technique for Routine Mapping of Land-Surface Carbon, Water and Energy Fluxes from Field to Regional Scales. *Remote Sens. Environ.* **2008**, *112*, 4227–4241. [CrossRef]
46. Peng, J.; Albergel, C.; Balenzano, A.; Brocca, L.; Cartus, O.; Cosh, M.H.; Crow, W.T.; Dabrowska-Zielinska, K.; Dadson, S.; Davidson, M.W.J. A Roadmap for High-Resolution Satellite Soil Moisture Applications–Confronting Product Characteristics with User Requirements. *Remote Sens. Environ.* **2021**, *252*, 112162. [CrossRef]
47. Chang, J.G.; Gao, F.; Anderson, M.; Cirone, R.; Zhao, H. Regionalization Analysis of Environmental Drivers of CONUS Grazing Land Biomass. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 12634–12644. [CrossRef]
48. Reichle, R.; De Lannoy, G.; Liu, Q.; Ardizzone, J.; Kimball, J.; Koster, R. SMAP Level 4 Surface and Root Zone Soil Moisture. In *Proceedings of the 2016 IEEE international Geoscience and Remote sensing Symposium (Igarss)*; IEEE: New York, NY, USA, 2016; pp. 136–138.
49. Colliander, A.; Reichle, R.H.; Crow, W.T.; Cosh, M.H.; Chen, F.; Chan, S.; Das, N.N.; Bindlish, R.; Chaubell, J.; Kim, S. Validation of Soil Moisture Data Products from the NASA SMAP Mission. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *15*, 364–392. [CrossRef]
50. Fang, B.; Lakshmi, V.; Bindlish, R.; Jackson, T.J.; Cosh, M.; Basara, J. Passive Microwave Soil Moisture Downscaling Using Vegetation Index and Skin Surface Temperature. *Vadose Zo. J.* **2013**, *12*, vzj2013-05. [CrossRef]
51. Das, N.N.; Entekhabi, D.; Dunbar, S.; Kim, S.-B.; Yueh, S.; Colliander, A.; O'Neill, P.; Jackson, T.J.; Jagdhuber, T.; Chen, F. SMAP/Sentinel-1 L2 Radiometer/Radar 30-Second Scene 3 Km EASE-Grid Soil Moisture-Global. National Snow and Ice Data Center. 2019. Available online: https://www.earthdata.nasa.gov/data/catalog/nsidc-cprd-spl2smap-s-003 (accessed on 4 April 2026).
52. Chang, J.G.; Kraatz, S.; Anderson, M.; Gao, F. Enhanced Polarimetric Radar Vegetation Index and Integration with Optical Index for Biomass Estimation in Grazing Lands Across the Contiguous United States. *Remote Sens.* **2024**, *16*, 4476. [CrossRef]
53. Ulaby, F.T.; Long, D. *Microwave Radar and Radiometric Remote Sensing*; University of Michigan Press: Ann Arbor, MI, USA, 2014.
54. Kellogg, K.; Hoffman, P.; Standley, S.; Shaffer, S.; Rosen, P.; Edelstein, W.; Dunn, C.; Baker, C.; Barela, P.; Shen, Y. NASA-ISRO Synthetic Aperture Radar (NISAR) Mission. In *Proceedings of the 2020 IEEE Aerospace Conference*; IEEE: New York, NY, USA, 2020; pp. 1–21.
55. Chang, J.G.; Kraatz, S.; Oh, Y.; Gao, F.; Anderson, M. Quality Assessment of Dual-Polarization C-Band SAR Data Influenced by Precipitation Based on Normalized Polarimetric Radar Vegetation Index. *Remote Sens.* **2025**, *17*, 2343. [CrossRef]
56. Roy, D.P.; Wulder, M.A.; Woodcock, C. Independent Oversight Crucial to Landsat Next. *Science* **2025**, *388*, 1033. [CrossRef] [PubMed]
57. Koutsantoni, P.; Kremezi, M.; Karathanassi, V.; Di Lauro, P.; Vargas-Solano, J.A.; Ceriola, G.; Aiello, A.; Lamboglia, E. A Multisensor Framework for Satellite Data Simulation: Generating Representative Datasets for Future ESA Missions—CHIME and LSTM. *Remote Sens.* **2026**, *18*, 1384. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.