DeepSoil: Physics-Guided Learning for High-Resolution Soil Moisture Mapping from Surface to Root-Zone via Multimodal Sensing and Process-Based Modeling

PAAHUNI KHANDELWAL, ANDREI BACHININ, TANJIM BIN FARUK, EVERETT LEWARK, SANGMI LEE PALLICKARA, SHRIDEEP PALLICKARA, Computer Science Department, Colorado State University, USA

Soil moisture plays a critical role in several domains and can be used to inform decision-making in agricultural settings, drought forecasting, forest fire predictions, and water conservation. Soil moisture is measured using in-situ and remote-sensing equipments. Depending on the type of equipment that is used, some challenges must be reconciled, including the density of observations, the measurement precision, and the resolutions at which these measurements are available. In general, in-situ measurements are high-precision but sparse, while remote sensing measurements benefit from spatial coverage, albeit at lower precision and coarser resolutions. The crux of this study is to produce higher-precision soil moisture estimates at the top soil and root zone and generate high resolution maps (30m). Our methodology combines scientific models, deep neural networks, topographical characteristics, and information about ambient conditions alongside both in-situ and remote sensing data to accomplish this. Domain science infuses several aspects of our methodology. In particular, we (1) construct science-infused training targets from physics-based soil moisture products, (2) regulate learning using multipart loss functions that enforce consistency with soil hydraulic behavior, and (3) scale training through region-aware clustering to balance tractability and spatial specialization. Our empirical benchmarks profile several aspects and demonstrate that our methodology accounts for spatial variability while accounting for both static (soil properties and elevation) and dynamically varying phenomena to generate accurate, high-precision 30m resolution soil moisture (SMC) maps. Finally, we extend DeepSoil to root-zone estimation by training a dedicated 20 cm model using supervision from mechanistic soil-water simulations and a physics-guided composite loss that enforces hydraulic bounds and matric-potential consistency, enabling high-resolution mapping beyond the surface layer.

Additional Key Words and Phrases: Big data, spatiotemporally evolving phenomena, science-guided learning, soil moisture, and deep neural networks

ACM Reference Format:
Paahuni Khandelwal, Andrei Bachinin, Tanjim Bin Faruk, Everett Lewark, Sangmi Lee Pallickara, Shrideep Pallickara. 2026. DeepSoil: Physics-Guided Learning for High-Resolution Soil Moisture Mapping from Surface to Root-Zone via Multimodal Sensing and Process-Based Modeling. 1, 1 (May 2026), 36 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

There has been a proliferation of in-situ and remote sensing equipment in several geospatial domains such as atmospheric sciences, agriculture, environmental and ecological monitoring, etc. These systems are used extensively to monitor and understand diverse phenomena [59, 83]. The devices report measurements at different frequencies, precision, and spatial resolutions.

Remote sensing systems provide excellent spatial coverage encompassing vast spatial extents (often earth scale). However, the data are available at coarse resolutions, and the reported measurements are commensurately low-precision

Author's address: Paahuni Khandelwal, Andrei Bachinin, Tanjim Bin Faruk, Everett Lewark, Sangmi Lee Pallickara, Shrideep Pallickara, Computer Science Department, Colorado State University, Fort Collins, Colorado, USA, paahuni@colostate.edu,andrei.bachinin@colostate.edu,tanjim.faruk@colostate.edu, everett.lewark@colostate.edu,sangmi.pallickara@colostate.edu,shrideep.pallickara@colostate.edu.

2026. Manuscript submitted to ACM

and contribute to averaging effects across the spatial extent. The coarse measurements coupled with the lower temporal revisit intervals preclude their effective use in decision-making.

In-situ sensing environments provide high-precision measurements and are typically equipped with networking capabilities and batteries allowing data generation to occur at higher frequencies. Given that in-situ devices provide point measurements their spatial coverage is poor. The costs for deploying in-situ sensors to ensure reasonably dense spatial coverage are often too prohibitive.

Scientific models encapsulate domain science and knowledge within the community to describe, analyze, and forecast phenomena. A challenge with scientific models is their parameterization to account for regional variations coupled with ancillary measurements to seed and calibrate the models.

The phenomenon we target is soil moisture, which plays a critical role in several domains. The crux of this study is the generation of higher-precision soil moisture maps at high spatial resolutions (30m) daily that can facilitate decision-making. The broader science issue that we consider is how to combine remote sensing, in-situ data, and scientific models to produce high-precision (similar to in-situ) maps at higher resolutions (much higher than remote sensing), with substantial spatial coverage (similar to remote sensing), and at moderate temporal frequencies (higher than remote sensing but lower than frequencies in in-situ devices).

1.1 Challenges

Generation of soil moisture maps by combining in-situ, remote sensing, and scientific models face several challenges. These include:

(1) **Account for regional heterogeneity:** Soil moisture is subject to regional variations stemming from ambient conditions. As such, the methods should be designed to account for such heterogeneity that drives subtle variations.
(2) **Multimodal data:** Generation of such maps relies on the inclusion of ancillary data relating to topography, weather, and soil characteristics. These data are multimodal and often available at different resolutions.
(3) **Voluminous data:** Cumulatively, the data are voluminous and available at different dimensionality, and rates, and may have temporal and spatial sparsity.

1.2 Research Questions

Specific research questions that we explore in this study include:

RQ-1: How can we combine the benefits of remote sensing and in-situ sensing schemes at the 5cm top soil level? Specifically, how can we leverage the spatial coverage of satellite products while anchoring estimates to the precision of point-based stations at the near-surface layer where remote sensing is most available?

RQ-2: How can we leverage domain science to regulate model training and ensure scientific consistency in our multi-depth models? In particular, how can physics-based relationships (e.g., soil water retention behavior) be embedded into the objective so the learned maps remain physically plausible across heterogeneous soils and moisture regimes?

RQ-3: How can we reconcile the competing pulls of computational tractability vs. refining models that are tuned to the particular spatial extent? This includes identifying a scalable middle ground between a single global model and an impractical per-location model, while still respecting regional heterogeneity.

RQ-4: How can we combine scientific models and sparse in-situ sensing data to generate soil moisture maps at 20 cm root zone depths for sub-surface soil moisture mapping? We note that root-zone estimation introduces an additional

constraint: deeper layers are less directly observed and are governed by slower, vertically coupled soil-water processes, making mechanistic simulation a practical source of depth-consistent supervision when station coverage is insufficient.

1.3 Approach Summary

We combine data from in-situ, remote sensing, and scientific models alongside ancillary data to generate higher-precision soil moisture maps. To accomplish this, we rely on a novel mixture of - (1) data wrangling, (2) creation of science-infused training datasets, (3) leveraging multi-modal data from ancillary phenomena that contribute to variations in soil moisture such as meteorological, topological, and soil properties, (4) designing deep neural networks with carefully calibrated layers to account for spatiotemporal variations, (5) training deep networks using our multi-part, science-guided loss functions, and finally, (6) designing a novel transfer learning scheme to orchestrate model training. More specifically, domain science enters at three levels: physics-based target products and mechanistic simulations provide spatially complete or depth-consistent supervision; physics-guided loss terms enforce hydraulic consistency during learning; and region-aware clustering preserves hydroclimatic coherence while keeping training tractable.

Our data-wrangling schemes rely on leveraging data from diverse sensing equipment alongside related phenomena. These data are available at different resolutions and temporal frequencies. Furthermore, the data have a mix of *static* components – such as soil properties, hydraulic conductivity, and topographical information – and *dynamically varying* components such as measurements, weather, vegetation, etc. We spatiotemporally align these datasets, partition them along pre-configured boundaries, and disperse them over our cluster of machines to ensure data locality. Next, we generate fixed-size tiles from these datasets that are suited for being input to our models. This tiled representation also standardizes inputs for convolutional prediction, enabling the model to learn spatial structure (edges, gradients, texture).

We supplement these with data from a well-known scientific model, HydroBlocks [13], to curate training datasets. HydroBlocks outputs are generated at 30m resolutions. While HydroBlocks encapsulates domain science, a disadvantage of HydroBlocks is that there are errors in its estimations. We reconcile errors in the HydroBlocks data using ground-truth data from in-situ sensors and remote sensing in the loss function during model training. This provides a physically grounded, spatially complete supervision signal while allowing the training objective to "pull" predictions toward station-consistent values where high-precision measurements exist.

Our deep networks consider observations (in-situ and remote) alongside ancillary information such as topography, elevation, weather conditions, etc., that, when collated together, get us closer to generating accurate, high-resolution estimates. Our deep network, code-named DeepSoil, is a variant of the U-Net architecture [68]. It is a convolutional neural network commonly used for image-to-image translation tasks. It features an encoder-decoder structure with skip connections. The encoder extracts features from the input image, the bottleneck refines them, and the decoder generates the output. U-Net is ideal for tasks like semantic segmentation, image colorization, and medical image analysis, as it preserves spatial details and contextual information through skip connections, making it highly effective in maintaining image quality during translation. We modify this architecture to generate high-resolution soil moisture maps by adding spatial feature maps to further enhance image quality.

We calibrate custom loss functions to ensure effective training of our deep networks. Our loss functions are multipart, science-guided, and account for several aspects such as the mean squared error (MSE), mean absolute error (MAE), incorporating SMC of in-situ observations, and enforcing fuzzy bounds on the expected variation of soil moisture within a tile. Different components of the loss function are weighted dynamically to account for their contributions and significance during model training. Our loss functions allow the training to be robust and numerically stable

while enhancing the model's capability for generalization. This is accomplished by regulating the model training using the van-Genuchten model [77], which characterizes the Water Retention Curve (WRC), capturing the relationship between water content and soil water properties. This physics-guided regularization is especially valuable under sparse supervision because it reduces the space of admissible solutions to those consistent with known hydraulic behavior, thereby improving generalization to unseen locations.

Our methodology involves a scaling component that relies on an ensemble of models to generate the soil moisture maps. Rather than train each of these constituent models exhaustively, we first partition the area of interest (CONUS or the continuous United States) into a set of clusters based on their land cover, climatic classifications, and soil characteristics. Clustering provides a practical bias–variance tradeoff: it reduces cross-regime averaging while avoiding the cost and instability of training a unique model per station or per small region.

We extend the core DeepSoil framework to enable the prediction of daily soil moisture at 20 cm, a depth critical for understanding root zone dynamics [24]. Estimating soil moisture at 20 cm is essential because it captures the root zone conditions that directly influence plant water uptake, crop stress, and ecosystem functioning. Unlike the near-surface layer, where satellite products and physics-based downscaling provide spatially complete supervision, the 20 cm layer is both less directly observed and more strongly governed by vertically coupled soil-water processes, making direct learning from sparse sensors undetermined. Though the core DeepSoil model produces spatially detailed maps at 5 cm, it cannot reliably represent conditions at deeper layers due to the scarcity and uneven distribution of in-situ measurements. This depth gap motivates the introduction of a mechanistic source of supervision that encodes physically consistent infiltration, redistribution, and drainage dynamics over time.

To address this limitation, we leverage the Agricultural Production Systems sIMulator (APSIM) [29], a well-established agricultural systems model that solves Richards' equation to simulate soil water dynamics. We execute APSIM at a large set of representative locations using locally parameterized soil hydraulic properties and meteorological forcing, producing depth-specific time series that can be aligned to our daily mapping cadence. We use APSIM outputs at 20 cm depth as spatially complete ground truth for training a dedicated U-Net model. During training, APSIM values serve as supervisory targets at the subset of pixels corresponding to simulation locations, and losses are evaluated only at these labeled pixels via a mask; the convolutional model then learns a mapping from gridded covariates to dense fields that generalizes beyond labeled points. This approach allows us to inherit the physical rigor of APSIM while producing predictions at 30m resolution across the state of Colorado. In effect, the mechanistic model supplies depth-consistent "anchors", while the learned model supplies the spatial continuity and scalability required for high-resolution statewide maps.

Our extended model incorporates physics-guided learning through a composite loss function. The loss combines a data-driven component (mean squared error against APSIM outputs) with a science-guided component based on the van Genuchten equation [77]. This composite objective is designed to balance fidelity to mechanistic supervision with hydraulic plausibility across heterogeneous soils, especially in regimes (very dry or near saturation) where small moisture errors can imply large changes in soil suction. We derive soil hydraulic parameters—residual water content (θ_r), saturated water content (θ_s), and shape parameters (α, n)—from the POLARIS soil property database. These parameters vary spatially at 30 m resolution, allowing the physics constraints to adapt per pixel rather than imposing a single global bound or retention curve. The physics loss penalizes predictions that violate physical bounds and enforces consistency in matric potential between predictions and targets. To ensure numerical stability when converting volumetric water content θ to matric potential ψ via the van Genuchten relation, we clamp θ to the physically meaningful interval $[\theta_r + \epsilon, \theta_s - \epsilon]$, where ϵ is a small constant (e.g., 10^{-4}), and we normalize ψ by $\psi_{\max}$ to keep gradients well-scaled during

backpropagation. We employ adaptive alpha scheduling during training: the loss begins with full emphasis on data fitting (MSE) and gradually transitions toward physics enforcement. This schedule prevents early-epoch optimization stiffness (when random initial predictions may grossly violate hydraulic constraints) and progressively tightens the solution space as the model approaches a data-consistent manifold. This strategy allows the model to first learn spatial patterns from APSIM supervision, then refine predictions to align with known hydrological constraints. The resulting extension transforms DeepSoil from a surface-level estimator into a tool capable of predicting root zone soil moisture with high spatial resolution. More broadly, it establishes a general template for depth-aware mapping: mechanistic simulation provides depth-specific supervision where sensors are sparse, while physics-guided losses ensure that learned fields remain consistent with established soil hydraulic behavior.

1.4 Paper Contribution And Translation Impact

Our methodology targets the generation of high-precision soil moisture maps from observational and other ancillary data to produce high-resolution maps at temporal frequencies that can be used to inform decision-making. Our specific contributions include the following:

(1) We reconcile multimodal data from not just sensing devices, but other ancillary data.
(2) Our methodology combines the spatial coverage of remote sensing with the precision of the in-situ measurements.
(3) This study represents a novel combination of scientific models, deep networks, equations governing physics of the phenomena used to regulate how the model learns, and multimodal data working in tandem to generate soil moisture maps.
(4) We demonstrate how ancillary-related data can be combined with scientific/domain-theoretic models to inform the generation of soil moisture maps. In particular, we show how such models can be trained and how their loss functions can be calibrated and refined to guide the training of models.
(5) We extend DeepSoil beyond surface mapping by introducing a root-zone (20 cm) modeling pipeline that leverages mechanistic soil-water simulations as supervision and a physics-guided composite loss to enforce hydraulic plausibility, enabling high-resolution sub-surface soil moisture maps under sparse in-situ coverage.

Translational Impact: Similar issues arise in other scientific domains where remote sensing sources may have higher spatial coverage, but lower precision while in-situ measurements may have poor spatial coverage but higher precision. Combining such diverse observational phenomena to produce higher-precision maps at high resolutions has applications in tracking greenhouse gas emissions, ecological monitoring, health hazards due to airborne pollutants, and other phenomena where spectral measurements are available alongside in-situ observations. Finally, we note that there will always be subtle regional variations relating to ambient conditions; it is important to account for these variations to ensure model accuracy. The proposed methodology describes how to accomplish this objective.

2 DATASETS LEVERAGED IN DEEPSOIL

2.1 Input and Target Datasets

Our methodology is designed to produce daily soil moisture maps at 30m spatial resolutions. A key focus is on accurately estimating soil moisture values in the top 5 cm of soil. We have developed a model that accounts for a rich set of features from phenomena that influence soil dynamics. We also leverage a variety of scientific datasets input and target data for our model. We categorize these data sources into two main groups: 'Static Phenomena' and 'Dynamic Phenomena'. These datasets play a pivotal role in training our multi-step deep neural network. Static features represent attributes or

Table 1. Datasets leveraged for training and inferences using DeepSoil.

Dataset	Training	Inference Dataset
gNATSGO	1.6 GB	8.6 GB
Polaris	2.4 GB	45 GB
Landsat	351 MB	8 GB
NLCD	114 MB	2.3 GB
Köppen Climate	76 MB	1.7 GB
DEM	241 MB	4.2 GB
GridMET	3.3 GB	92 MB
MCD15A3H (Interpolated)	41 GB	13.6 GB
SMAP	1.2 GB	93 MB
HydroBlocks	70 GB	-
In-situ stations	1.4 GB	-
Total	**121.68 GB**	**83.56 GB**

characteristics that are relatively constant for a given region but have variability *across* spatial regions. In contrast, dynamic phenomena include features that change on a daily/hourly basis. In our model, DeepSoil, we integrate 11 distinct datasets, each described in this section. In Table 1, we provide estimates of the size of the preprocessed datasets used as inputs and targets during training; we also leverage these datasets for inferences and estimates that we produce over the states of Colorado and Oklahoma.

1) Static Phenomena

Soil Properties: We incorporate two distinct soil properties datasets: the Gridded National Soil Survey Geographic (gNATSGO) Database [74] and the Probabilistic Remapping of SSURGO (POLARIS) soil properties [14] dataset, both covering the CONUS.

From the gNATSGO database, we selectively utilize a subset of key properties at 30m spatial resolution. These include the available water storage estimate (in mm) and the thickness of soil components (in cm). The available water storage estimate is a vital factor in our soil moisture estimation, offering insights into the actual soil water content at a given location. Although it exhibits dynamic characteristics and is influenced by factors such as precipitation, plant water uptake, and evaporation, it serves as a baseline for the expected available water storage in a particular region. This is very useful, especially in regions where in-situ observation stations are lacking. Additionally, we incorporate derived features at coarser (90m) spatial resolution, encompassing available water content, field capacity, and soil porosity. The available water content represents the range of moisture available for plant uptake. In cropping settings, the available water content is between the soil's maximum water-holding capacity (field capacity) and the point at which plants can no longer extract water effectively (wilting point). Soil porosity provides information about soil pore volumes that impact water-holding capacity, aeration, and root penetration.

The POLARIS dataset provides additional details (at 30m resolution) about soil properties, including soil texture, pH, organic matter content, and mineral composition. We specifically leverage data on soil texture, such as silt, clay, and sand percentages. Additionally, we utilize bulk density (g/cm^3), residual soil water content (m^3/m^3), organic matter (%), and saturated hydraulic conductivity ($log_{10}(cm/hr)$) to account for soil mass per unit volume and its ability to transmit water. One such known relationship is between soil organic matter and soil porosity, where a high percentage of organic matter increases soil porosity value, allowing the soil to store more water and air for plant roots, thus increasing soil moisture content [32]. Also, soil moisture, soil matric potential, and hydraulic conductivity hold non-linear relationships with soil potential (energy exerted by plants to extract water from the soil) which varies

depending on soil texture and composition (soil particle size), pore size, organic matter content, etc. These interactions are also expressed using soil retention curves (WRCs) and hydraulic conductivity functions (HCFs).

Satellite Imagery: We also include cloud-free satellite imagery obtained from the Landsat 8 Collection 2 Level-2 atmospherically corrected surface reflectance product [50]. This imagery provides a high-resolution view of the land surface at 30m across the red, blue, green, and near-infrared (NIR) spectra. These spectral bands enable us to compute the Normalized Difference Vegetation Index (NDVI), a valuable indicator for assessing vegetation health. Lower NDVI values suggest moisture-stressed vegetation, while higher values indicate denser and healthier green vegetation.

Land Cover and Climatic Conditions: We incorporate two additional datasets to enhance our model's contextual grounding. The first dataset is the National Land Cover Database (NLCD) [30] that provides a comprehensive classification of the Earth's surface into 16 distinct categories, including forests, urban areas, agriculture, and water bodies. This dataset enables precise characterization of land covers within our study area.

The second dataset we leverage is the Köppen climate classifications [43]. The Köppen classifications offer insights into long-term climatic patterns across different geographic regions. In particular, these classifications segment CONUS into 30 classes organized under five primary climatic groups: tropical, dry, temperate, continental, and polar.

Elevation Dataset: Relative slopes have an impact on pooling, water movement, and soil moisture variations within a region. We include the 3D Elevation Program's (3DEP) 1 Arc-second Digital Elevation Models (DEMs) dataset [75] as our final static feature. This dataset represents the bare-earth topographic surface, offering elevation values at a 30m spatial resolution.

2) Dynamic Phenomena

Meteorological Conditions: We leverage the Gridded Surface Meteorological (GridMET) dataset [1] for access to daily surface meteorological data covering CONUS. The GridMET data is at 4 km resolution and includes several meteorological variables that exhibit dynamic changes over time such as precipitation accumulation, maximum and minimum relative humidity, temperature, etc.

Leaf Area Index (LAI): To capture the dynamic aspect of vegetation and its influence on soil moisture, we incorporate the MODIS/ Terra+Aqua Leaf Area Index/FPAR (MCD15A3H.061) dataset [58]. The leaf area index plays a pivotal role in regulating soil moisture dynamics. As vegetation grows, it can extract water from the soil through transpiration. Higher leaf area index values may indicate increased water uptake, potentially leading to reduced soil moisture levels, while lower values suggest reduced vegetation. The dataset quantifies the leaf area in a specific region; the leaf area index changes with plant growth and seasonal variations. It is available at 4-day intervals and 500m spatial resolution.

Satellite-Based Soil Moisture Maps: We also rely on the Soil Moisture Active Passive (SMAP) satellite data [23] to obtain a time-series of soil moisture measurements at a spatial resolution of 9 km within the top 5 cm of the soil. This dataset is categorized as a Level 3 product, presenting daily composites derived from Level 2 surface soil moisture data (36 km). The dataset offers a relatively coarse estimate of soil moisture, providing an overview of moisture trends within a region [25, 41], although with limited precision in capturing field-scale soil moisture.

In-situ Soil Moisture Measurements [RQ-1]: Our objective in this research extends beyond generating high-resolution (30m) soil moisture maps using deep neural networks (DNNs). A key goal is to ensure accuracy and precision that are close to in-situ station data. Notably, HydroBlocks exhibits errors when compared to in-situ observations. A key source of these errors seems to be limitations in accounting for the influence of precipitation events on soil moisture dynamics. In-situ soil moisture station data is the most accurate soil moisture estimate available; given that it is a point measurement the estimate is accurate within proximate spatial regions (typically less than 10-30m), making our task particularly challenging.

Table 2. In-situ soil moisture observational networks used for DeepSoil predictions. Individual networks are managed by different federal and state agencies.

Soil Moisture Network (no. of stations)	
TexMesonet (32)	Nebraska Mesonet (18)
Texas Soil Observation Network (69)	Oklahoma Mesonet (114)
Kentucky Mesonet (42)	USCRN (157)
Missouri Agricultural Database (3)	Soil Scape (54)
New York State Mesonet (121)	Scan (178)
Illinois Climate Network (19)	Snotel (350)
Montana Mesonet (14)	NoahHMT (21)
Delaware Environment Observing System (24)	

To produce 30m spatial soil moisture maps that closely align with in-situ data, we harness soil moisture measurements from **15 distinct soil moisture observation networks** [7, 9, 10, 12, 22, 31, 44, 48, 54, 57, 69, 71, 76, 80], as detailed in Table 2. In total, there are **1216 observation stations**, each continuously measuring soil moisture levels on an hourly/daily basis. The precise locations of these stations are depicted in Fig. 1.

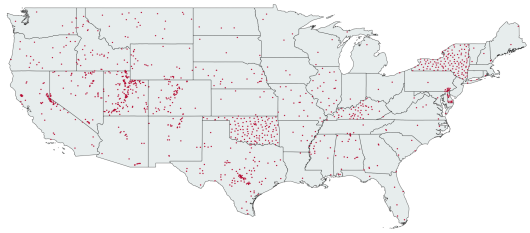


Fig. 1. Geographic distribution of in-situ soil moisture monitoring stations (operated by state/federal agencies) within the United States.

Physics-Based Soil Moisture Maps [RQ-2]: SMAP-Hydro Blocks (SMAP-HB) [13, 79] are high-resolution soil moisture maps generated through a physics-based approach using the Tau-Omega Radiative Transfer Model (RTM), reverse RTM models and Bayesian Merging Scheme. This methodology starts by identifying hydrologically similar units within a region, capturing interactions based on meteorological, topological, water flow, and hydrological properties of the soil. Next, RTM-based temperature products are computed and integrated with SMAP L3 brightness temperature observations through a Bayesian merging approach. Finally, an inverse RTM is applied to convert these temperature products into downscaled soil moisture maps. This satellite-based surface soil moisture dataset covers an entire continent, serving as our primary target dataset for training our deep neural networks. Although not representing the ground truth data, it provides accurate soil moisture (7% MAE) that closely approximates real-world conditions. The HydroBlocks product offers data at a 30m spatial resolution, with a temporal frequency of 6 hours, encompassing CONUS but only for the period spanning 2015 to 2019.

3 METHODOLOGY

Our overall approach for generating 30m daily soil moisture maps is presented in Fig. 2. Our methodology encompasses:

(1) Data wrangling and harmonization: To reconcile and harness diverse multimodal data. We distribute our raw input datasets uniformly across a cluster of machines, with each machine responsible for preprocessing and modeling a subset of the data.

(2) Design of deep neural networks: The network design includes layers suited to our task alongside a calibration of the network structure using the HyperBand algorithm.
(3) Science-guided loss functions: We have designed a novel, multi-part loss function that accounts for non-linear interactions between soil properties to regulate how our deep neural network learns.
(4) Clustering regions based on SMC dynamics: This is used to manage the competing pulls of refinements and computational requirements. Rather than train an all-encompassing model, we cluster spatial extents based on their spatial similarities along soil and hydrological characteristics. Models are trained for clusters of regions. During inferences, each spatial extent is parameterized based on their available multimodal data.

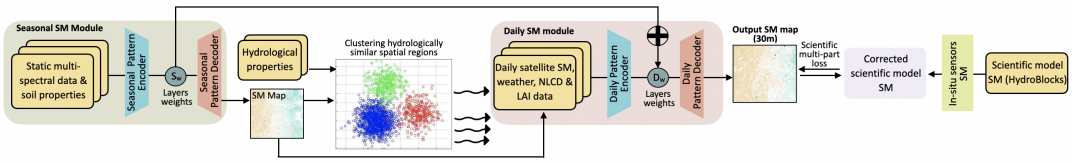


Fig. 2. System Overview: Dataset preprocessing is performed while preserving data locality. k-means++ clustering, soil moisture prediction models, and science-informed multipart loss functions are key components of the DeepSoil framework.

Our model development process proceeds in two phases. In the first phase, we design a model that forecasts monthly average soil moisture maps. The monthly-average model captures nonlinear relationships between soil characteristics within a region and the monthly average precipitation patterns within a spatial extent. The coarse temporal granularity of the monthly-average model allows us to ensure accuracy and to provide contextual grounding of the daily-moisture model.

In the second phase, we design a daily-moisture estimation model that specializes in assimilating daily variations across input features such as meteorological data (GridMET), leaf area indices, land cover, and topographical information to estimate daily soil moisture maps. We also include coarse-grained, low-precision SMAP satellite soil moisture measurements. All the input datasets fed to the daily-moisture model are available daily and at diverse spatial resolutions.

Our methodology partitions the area of interest into spatial extents, S_E. Each spatial extent, S_E, comprises 9 tiles, and each tile has 64x64 pixels, where each pixel has a 30m resolution. Our training datasets are collated in terms of spatial extents. For each spatial extent, we spatiotemporally harmonize and align multimodal data across meteorological, topographical, soil properties, etc. These data are often available at different spatial resolutions. Spatial extents are also employed to adjust soil moisture estimates generated using scientific models.

3.1 Data Wrangling and Harmonization

Our research involves the integration of diverse data sources with varying value ranges, spatial resolutions, and temporal frequencies. To ensure efficient model training, we harmonize all datasets to a consistent spatial resolution (30m), temporal frequency (daily), and normalized values (0 to 1). Missing values are marked as -1, and a mask is created to identify these locations. Datasets that are not available on a daily basis undergo the nearest interpolation process to bridge temporal gaps using proximate temporal scans. For spatial consistency, we employ OpenCV's [8] inter-nearest image resizing, reducing all datasets to a uniform size of 64x64 pixels.

To handle large raw tiles, some of which encompass the entire CONUS, we adopt the quadkeys concept [63]. This approach recursively divides the geospatial coordinate space into non-overlapping bounding boxes, each assigned a

specific identifier. Smaller quadkey lengths represent larger spatial regions. We partition all datasets with 14-character-long quadkeys. For GridMET and SMAP datasets with coarser spatial resolutions, we use 12-characters long quadkeys. This partitioning strategy aligns with memory and computational constraints during model training.

Fig. 3. Example of multiple spatial extents in Northern Delaware represented as a 3x3 grid arrangement of 9 tiles. Here, each tile is 64x64 pixels in size. The center tile in each spatial extent is colored yellow and encapsulates an in-situ station. Each tile has an associated unique 14-character key.

In addition to partitioning images into 64x64 spatial regions, we introduce the concept of Spatial Extent (S_E). This is a spatial region generated by merging 9 neighboring 64x64 tiles centered around an in-situ station. The resulting image encompasses a larger spatial area, measuring 192x192 pixels. This is illustrated in Fig. 3. When we cluster the tiles and perform the warn-start model training, we have used a similar scheme of spatial extents.

3.2 Clustering Regions based on Soil Moisture Dynamics [RQ-3]

Next, we identify regions that share similar soil moisture dynamics and properties. This is achieved by integrating an extensive range of environmental and climatic datasets. These datasets encompass the Köppen climate map, landcover information, elevation data, seasonal averages of soil moisture extracted from SMAP observations, and soil property data sourced from the POLARIS and gNATSGO databases.

We cluster spatial extents characterized by similar soil moisture dynamics and properties. We leverage the k-Means++ clustering algorithm, known for its capacity to optimize cluster initialization. Identification of the number of clusters, k, is guided by the maximization of Silhouette scores which assess how well data points within a cluster are separated from each other compared to how close they are to data points in neighboring clusters. We classify 1216 spatial extents into 19 distinct clusters.

The number of identified clusters corresponds to the number of distinct model instances that we train. We train 19 distinct model instances, each specializing in the prediction of soil moisture dynamics within regions classified as hydrologically similar. Each cluster is comprised of 20 to 110 distinct locations. We distribute the training data across 19 machines, each corresponding to these distinct clusters, and simultaneously train the models. Each model utilizes locally stored data on its respective machine's disk.

3.3 DNN Architecture [RQ-2, RQ-3]

Our proposed architecture is a modified U-Net framework [33], a 4-block encoder-decoder DNN with skip connections between corresponding layers in both the encoder and decoder stacks. The network used for model training has a total of 12.9 million trainable parameters.

Each encoder block comprises a stack of 2D convolutional layers, batch normalization, and LeakyReLU activation functions within the inner layers. In contrast, the outer layer employs a ReLU activation function, ensuring that output values are constrained within the 0-1 range. These activation functions allow our network to model and capture complex non-linear relationships between input variables such as hydraulic conductivity, porosity, etc, and output soil moisture

predictions. These encoder blocks extract embeddings from the input meteorological and hydrological features. Identical decoder blocks replace the convolutional layer with Conv2DTranspose, which upsamples the latent vector (4x4x512) from the encoder back to a 64x64 image, representing the generated soil moisture. Skip connections are introduced to recover information lost during input downsampling in the Encoder blocks. Skip connections have been known to facilitate faster training and gradient flow during backpropagation, effectively mitigating the vanishing gradient problem. Dropout layers are also strategically added to prevent overfitting during model training. In the final decoder block, we employ a concatenation operation, merging low-level features extracted from the NDVI index, landcover classification maps, and Landsat-derived RGB bands with the upsampled image from the U-Net architecture. This enhances model performance by capturing intricate spatial relationships within the input data. This informs our model's understanding of road structures, agricultural patterns, and spatial region structures, ultimately improving its ability to produce high-quality outputs.

3.4 Science-Guided Loss Function [RQ-2]

A key contribution of this study is a novel loss function to regulate model training by accounting for the influence of key physical processes that account for the dynamics of soil water content within the top 5 cm of soil. In particular, we have designed a multi-part weighted loss function that accounts for various factors, including soil texture, soil dryness/wetness, and available water for uptake by the plant. The weights assigned to each component of the loss function are dynamically adjusted over training epochs to modify the impact of loss terms on accuracy across various model training scenarios. For all experiments, the predicted soil moisture values and the ground truth values are multiplied by a mask to exclude invalid or missing data points as marked in the target data.

Overall, our unified loss function is a weighted combination of the L2 error (MSE) and a science-based loss shown in Equations (1) and (2), that enforce the scientific consistency of relationships across soil characteristics, soil matric potential, and soil moisture dynamics during model training. Here, α represents a hyperparameter whose value is reduced dynamically while training to reduce the impact of the basic loss component (MSE) and increase the weight of the scientific loss term. We dynamically adjust α based on the training epoch and the total number of epochs with the objective of enhancing model performance. Our methodology divides the training process into steps, gradually decreasing α to fine-tune the learning rate over time. During initial training epochs, the model prioritizes backpropagating errors from pixel-wise squared errors of SMC from the HydroBlocks dataset. Towards the end of the training, the scientific loss term attains the highest weight as α values decrease to 0 to consider the relations among the predicted SMC and other soil properties. This approach is an effective optimization technique for achieving convergence; crucially, as our performance benchmarks demonstrate, this allows DeepSoil to achieve better generalization on unseen locations.

$$L_{\text{Total}} = \alpha \cdot L_{\text{MSE}} + (1 - \alpha) \cdot L_{\text{sci}} \tag{1}$$

3.4.1 ***Mean Squared Errors.*** We define a custom MSE loss between the predicted and target soil moisture maps. This criterion calculates the pixel-wise average of the squared difference between the corresponding valid elements of the predicted and target soil moisture maps. The squaring operation, allows the MSE to be more sensitive to outliers or large errors. The MSE plays a significant role during the initial training epochs when errors are typically higher.

$$L_{\text{MSE}} = \frac{1}{B}\sum_{k=1}^{B}\frac{1}{N}\sum_{i=1}^{N}\left(w_i \cdot \left|x_{k,i} \cdot mask_{k,i} - y_{k,i} \cdot mask_{k,i}\right|^2\right) \tag{2}$$

Here, L_{MSE} represents the average MSE over the training dataset. B represents the number of images in the batch. N is the total number of pixels in each image. $x_{k,i}$ is the value of the pixel at index i in the predicted image of the k-th image in the batch. $y_{k,i}$ is the value of the pixel at index i in the target image of the k-th image in the batch. w_i is the weight associated with the pixel at index I. For the 16 neighboring pixels around the given index where in-situ is located, w_i has a higher value, giving them more importance than remaining pixels, and $mask_{k,i}$ provides valid pixel locations.

3.4.2 ***Soil Water Retention Curves (WRC).*** As reported by the authors of HydroBlocks, the SMC dataset exhibits notable errors that can range from approximately 7% to 15%, when compared to in-situ station measurements of soil moisture. Given our reliance on HydroBlocks as a scientifically consistent target dataset, addressing the inherent disparities between these target values and ground-truth measurements is necessary.

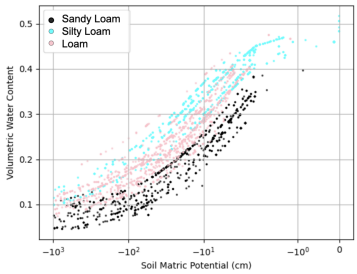


Fig. 4. Non-linear relation captured using van-Genuchten equation between water content and soil matric potential (in logarithmic scale) for different soil types using 1200+ in-situ sensors observation over a year.

The soil Water Retention Curve (WRC) represents a well-known relationship between soil characteristics - soil matric potential (ψ) and water content (θ) [60]. Soil matric potential, also known as capillary pressure, denotes the pressure or energy required to extract water from the soil against gravity, and the force with which water is withheld within soil pores and particles. (ψ) serves as a direct measure of the available water for plants to uptake from soil. As soil moisture decreases, capillary forces increase, causing water to be retained more tightly within the soil pores. A decrease in soil moisture thus results in a more negative matric potential (ψ), indicating a greater energy requirement for water extraction from the soil and vice versa. The relationship is accurately captured in a widely-used numerical equation called the van-Genuchten equation [77], which is based on observed field/lab values of soil properties. Our model training relies on this scientific relation between soil properties, soil matric potential, and SMC to adjust the model's trainable parameters. Key parameters characterizing this relationship include soil type, residual water content, saturated water content, the α parameter influencing curve shape (shown in Fig. 4), and the n parameter representing soil pore-size distribution. This relationship can be mathematically formulated as below -

$$\psi = \frac{1}{\alpha} \cdot \left[\left(\frac{\theta_s - \theta_r}{\theta - \theta_r}\right)^{n/n-1} - 1\right]^{1/n} \tag{3}$$

here, ψ is soil matric potential, α is inverse of air-entry value, θ_s is saturated water content, θ_r is residual water content, θ is volumetric water content, n, m are some constants defining curve.

Our scientific loss term then becomes -

$$L_{sci} = \psi_{predicted} - \psi_{target} \tag{4}$$

We capture this non-linear relationship in our science-guided loss function to calculate ψ using Equation (3) by leveraging some of the variables from POLARIS datasets available at 30m resolution and ground-truth SMC using in-situ sensors. The soil matric potential serves as a regularization term for our DeepSoil model by calculating how far the desired ψ is from the ψ calculated using model-generated SMC. In regions where SMC is unavailable through in-situ sensors, our target dataset, HydroBlocks SMC is used to calculate ψ using the van-Genuchten equation. DeepSoil learns the relation between soil matric potential and soil moisture content by learning patterns over different soil conditions but at different locations.

By integrating station-based error adjustments through loss function into our approach, we achieve a higher level of precision and reliability in the target soil moisture maps that we used to train our model against. This scientific approach enables us to capture localized variations and address discrepancies between ground truth measurements and physics-modeled values, ultimately increasing the accuracy of our predictions.

3.5 Root Zone Soil Moisture Simulation Using APSIM [RQ-4, RQ-2]

Depth-specific root-zone soil moisture fields at fine spatial granularity are not routinely available from existing gridded products. This creates a supervision gap for RQ-4: we want 30m maps at 20 cm, but we do not have dense, depth-resolved measurements that can serve as training targets at that scale. To address this limitation, APSIM (The Agricultural Production Systems SIMulator) is executed at a large set of representative locations using its SWIM3 soil-water module to generate point-based, physics-driven estimates that supervise and constrain learning of a two-dimensional, map-producing predictive model. In other words, APSIM provides physically consistent "labels" at depth, and the learning model later converts these sparse point labels into spatially continuous rasters by learning how covariates (soil/terrain/climate/weather) map to the APSIM depth state. The SWIM3 simulations provide point-level soil-moisture trajectories as supervisory targets, while the predictive model learns a mapping from gridded covariates to spatial fields; training uses sparse (masked) supervision so that the loss is evaluated only where supervisory targets are available. This separation of responsibilities is deliberate: SWIM3 enforces physics in the vertical (profile) dimension; the CNN enforces spatial structure and produces dense maps.

In SWIM3, unsaturated vertical flow is represented by the mixed-form Richards equation (derived from mass conservation combined with the Darcy–Buckingham flux law), which can be written as:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[K(\theta)\left(\frac{\partial h}{\partial z} - 1\right)\right], \tag{5}$$

where θ is volumetric water content, h is pressure head, $K(\theta)$ is unsaturated hydraulic conductivity, z is depth (positive downward), and t is time. These simulations yield physically consistent, point-level time series of soil moisture states that are used as supervisory targets for learning spatially continuous soil-moisture maps. Richards' equation is essential here because it couples storage (θ) and flow (via $K(\theta)$ and pressure gradients), allowing the 20 cm state to respond coherently to precipitation pulses and drying cycles rather than being a heuristic "shift" of surface moisture.

All APSIM simulations are conducted in a **fallow configuration** (crop growth disabled and no irrigation applied) so that soil-water dynamics are driven by **meteorological forcing** and **soil hydraulic properties**. We adopt fallow simulations to remove confounding management effects (crop water uptake, rooting depth dynamics, irrigation scheduling)

so that the generated supervisory targets reflect physics that is consistent across the landscape and directly attributable to inputs we can represent everywhere (weather + soil + terrain). Meteorological inputs are provided by **GridMET**, supplying the boundary conditions required to evolve the soil profile through time (e.g., precipitation and near-surface atmospheric variables used to represent evaporative demand). This choice also aligns the simulator's forcing with the covariates later used by CNN, ensuring that the supervisory signal is conditioned on the same information the predictor will receive at the inference time.

SWIM3 requires hydraulically consistent parameter sets. Accordingly, soil hydraulic properties are derived from texture using a structured **pedotransfer** procedure, where a pedotransfer function (PTF) translates readily available soil descriptors (here, texture fractions) into the hydraulic parameters required by a mechanistic flow model. This PTF step is not merely convenience: it is what makes statewide scaling feasible, because texture is available as a gridded product while full hydraulic curves rarely are. First, for each APSIM job location we obtain sand, silt, and clay fractions from **POLARIS** and normalize the fractions to enforce closure in percent units. Let $T = \text{sand} + \text{silt} + \text{clay}$. The normalized texture components are

$$\text{sand}_N = 100\,\frac{\text{sand}}{T}, \qquad \text{silt}_N = 100\,\frac{\text{silt}}{T}, \qquad \text{clay}_N = 100 - \text{sand}_N - \text{silt}_N, \tag{6}$$

so that $\text{sand}_N + \text{silt}_N + \text{clay}_N = 100$. Normalization is important for numerical stability and physical plausibility because small inconsistencies in reported texture fractions can propagate into large differences in inferred hydraulic parameters.

Second, we apply a Rosetta pedotransfer model to map normalized texture to van Genuchten–Mualem hydraulic parameters:

$$(\theta_r, \theta_s, \alpha, n, K_s) = f_{\text{PTF}}(\text{sand}_N, \text{silt}_N, \text{clay}_N). \tag{7}$$

These parameters define both storage (via the retention curve) and flow (via conductivity), so they form the physical "signature" of a soil column that governs how precipitation infiltrates and redistributes with depth.

This mapping is evaluated through a precomputed Rosetta lookup table defined over integer texture fractions. Parameters for an arbitrary texture are obtained by interpolation over neighboring texture vertices in this table. Let $\mathbf{x} = (\text{sand}_N, \text{silt}_N, \text{clay}_N)$ and let $\mathbf{x}_i$ denote the surrounding vertices with associated table parameters $\mathbf{p}_i$. Barycentric weights $w_i(\mathbf{x})$ are computed such that

$$\sum_i w_i = 1, \qquad w_i \geq 0, \qquad \mathbf{x} = \sum_i w_i\,\mathbf{x}_i, \tag{8}$$

and parameters are interpolated as

$$\mathbf{p}(\mathbf{x}) = \sum_i w_i(\mathbf{x})\,\mathbf{p}_i. \tag{9}$$

For parameters reported by Rosetta in base-10 logarithmic form (e.g., $\log_{10}\alpha, \log_{10} n, \log_{10} K_s$), interpolation is performed in the reported log space to reflect the scale on which these quantities are parameterized and to preserve positivity upon back-transformation:

$$\alpha = 10^{\log_{10}\alpha}, \quad n = 10^{\log_{10} n}, \quad K_s = 10^{\log_{10} K_s}. \tag{10}$$

Interpolating in log space is critical because α, n, and K_s can vary over orders of magnitude; linear-space interpolation would bias parameters toward unrealistically large values and can destabilize the flow solution.

Next, given $(\theta_r, \theta_s, \alpha, n)$, the van Genuchten retention function with $m = 1 - \frac{1}{n}$ is defined as

$$\theta(h) = \theta_r + \frac{\theta_s - \theta_r}{[1 + (\alpha|h|)^n]^m}, \qquad m = 1 - \frac{1}{n}, \tag{11}$$

and standard water contents used in SWIM-style parameterizations are computed, including **LL15** (a lower-limit water content commonly associated with permanent wilting conditions) and **DUL** (a drained upper limit commonly associated with field-capacity conditions), obtained by evaluating $\theta(h)$ at standard reference suctions (denoted as h_{15} and h_{33}, respectively). Hydraulic conductivity is represented using the Mualem–van Genuchten formulation,

$$K(S_e) = K_s\, S_e^{\ell}\left[1 - \left(1 - S_e^{1/m}\right)^{m}\right]^2, \qquad S_e = \frac{\theta - \theta_r}{\theta_s - \theta_r}, \tag{12}$$

and is evaluated at a representative state (e.g., θ = DUL) to obtain a conductivity value used to support numerically stable simulation. Together, the retention and conductivity functions ensure that SWIM3 responds differently for coarse vs. fine textures (e.g., rapid drainage in sandy soils vs. stronger retention and slower redistribution in clay-rich soils), which is exactly the depth-dependent behavior we need the downstream mapping model to learn.

Finally, explicit physical and numerical validity constraints are applied to the resulting hydraulic set (e.g., enforcing the water-content hierarchy $0 \le \text{AirDry} < \text{LL15} < \text{DUL} < \text{SAT} \le 1$, and separately bounding saturation, bulk density, and conductivities; here AirDry denotes a residual (near air-dry) volumetric water content and SAT denotes saturated volumetric water content) to ensure plausibility and stability in SWIM3. These checks act as a guardrail against physically impossible soils (e.g., $\theta_s \le \theta_r$) and against parameter combinations that can cause numerical stiffness or nonconvergence in Richards' equation. In practice, they are also what makes it safe to deploy the same procedure at scale across heterogeneous landscapes.

SWIM3 outputs soil moisture states by soil layer. To define a depth-consistent supervisory target at 20 cm, we compute $\theta(20\text{ cm})$ by **linear interpolation between the two adjacent layer midpoints that bracket 20 cm**. In our configuration, the reported states used for this interpolation are taken from the fourth (**SW[4]**) and fifth (**SW[5]**) layers, whose mid-depths are approximately $z_1 = 17.5$ cm and $z_2 = 25$ cm, respectively. Specifically,

$$\theta(z^*) = \theta(z_1) + (\theta(z_2) - \theta(z_1))\,\frac{z^* - z_1}{z_2 - z_1}, \quad z^* = 20\text{ cm}, \tag{13}$$

where $z_1 < z^* < z_2$ are the bracketing mid-depths and $\theta(z_1), \theta(z_2)$ are the corresponding SWIM3 layer moisture states. This produces a single daily, point-level estimate at 20 cm depth that is used to supervise learning of spatially continuous soil-moisture maps. This interpolation step ensures that the target depth is consistent across locations even when SWIM3 layer boundaries do not align exactly at 20 cm, and it avoids introducing discontinuities that would occur if we selected the nearest layer state without depth adjustment.

Connection to RQ-2 and the physics-guided learning objective: A key advantage of this APSIM-based supervision strategy is that it is naturally compatible with the same soil physics used elsewhere in DeepSoil. The hydraulic parameters inferred via Rosetta (e.g., θ_r, θ_s, α, n) are the same quantities that appear in the van Genuchten relation used later for physics-guided losses. This creates a consistent physical "language" across simulation and learning: SWIM3 generates depth dynamics using these parameters, and the learning model can be regularized using the same parameterized constraints to discourage physically implausible moisture states.

3.6 Generating Daily, Tile-based 20 cm Root Zone Soil Moisture Targets [RQ-4]

Building on the APSIM simulation framework described previously, we describe our approach for generating spatially continuous soil moisture predictions at 20 cm depth. The APSIM outputs provide point-based, physically consistent soil moisture estimates that serve as training targets. Our task is to learn a mapping from gridded covariates to spatial fields that generalizes across the landscape. At a high level, this stage converts (sparse) APSIM point trajectories into a

supervised learning problem over image-like tiles. The key design challenge is that the predictor must output dense 30 m rasters while the APSIM "labels" exist only at a subset of pixels. We represent each training example as a tile-date pair augmented with a binary supervision mask, and we optimize the model using losses computed only where APSIM targets exist.

3.6.1 Model Architecture. We employ a lightweight U-Net architecture optimized for tile-based soil moisture prediction. The model takes 64×64 pixel input tiles (with 20 feature channels) and produces corresponding 64×64 pixel output tiles containing predicted soil moisture at 20 cm depth. Each pixel corresponds to 30-meter spatial resolution. To generate a complete soil moisture map for Colorado, we apply the model to tiles covering the state and mosaic the predictions into a seamless raster.

This tile-based inference strategy makes statewide mapping tractable: each forward pass produces a local field estimate, and the final statewide raster is obtained by stitching the outputs in geographic space. Because the model operates on fixed-size windows, it can be executed in parallel over tiles and dates, and it does not require loading the full state into memory.

The architecture consists of three encoder blocks, a bottleneck layer, and three decoder blocks with skip connections. The encoder path progressively downsamples the spatial resolution through max-pooling: from 64×64 to 8×8. Simultaneously, the number of feature channels increases at each level (from 20 to 256), allowing the network to learn increasingly abstract representations as the spatial detail is reduced. The first two encoder blocks use double convolution layers (Conv → BatchNorm → ReLU → Conv → BatchNorm → ReLU), while the third uses a single convolution for computational efficiency. The bottleneck operates at 8×8 resolution with 256 channels. The decoder path mirrors this structure, using transposed convolutions for upsampling and concatenating skip connections to recover spatial detail lost during encoding. Skip connections are particularly important here because the target variable (20 cm soil moisture) varies smoothly at broad scales but still exhibits sharp spatial transitions at land use boundaries and terrain breaks. The skip paths let the decoder reintroduce high-frequency spatial cues (e.g., land cover edges, topographic gradients) that would otherwise be lost during downsampling. An important design choice is the reduced bottleneck capacity. Standard U-Net architectures often use 16× the base filters at the bottleneck, resulting in approximately $40M$ parameters. Our architecture uses only 4× the base filters (256 channels with base filter of 64), reducing the parameter count to $\approx 3M$. This reduction is appropriate for our 64×64 input size and prevents overfitting while maintaining sufficient capacity to learn the deterministic soil moisture patterns from APSIM. The output layer applies a sigmoid activation to constrain the predictions to the valid soil moisture range $[0, 1]$. This parameter-budget choice is also aligned with the supervision regime: because labels are sparse within each tile, excessively large models can overfit to local idiosyncrasies at labeled pixels and generalize poorly to unlabeled pixels within the same tile. The smaller bottleneck serves as an implicit regularizer, encouraging the network to learn the dominant, physically plausible controls on root-zone moisture rather than memorizing site-specific artifacts.

3.6.2 Input Feature Representation. The model receives 20 input channels organized into static and dynamic components. The static channels (8 total) capture time-invariant soil and terrain properties. These static covariates provide the "hydrologic template" that shapes how water is stored and redistributed at depth (e.g., texture controls retention and conductivity; terrain modulates redistribution and local wetness gradients). **Soil texture (3 channels):** Sand, silt, and clay fractions from POLARIS, normalized to $[0, 1]$ from their original percentage values. **Terrain attributes (3 channels):** Elevation, slope, and aspect derived from the digital elevation model. **Land cover and climate (2 channels):** NLCD land cover classification and Köppen climate zone, both normalized to $[0, 1]$.

The dynamic channels (12 total) capture time-varying meteorological conditions from GridMET. These variables act as the daily forcing that drives infiltration and drying, allowing the model to represent event responses (e.g., rainfall pulses) and seasonal drying/wetting cycles in the root zone. **Daily weather (4 channels):** Precipitation, mean temperature, vapor pressure deficit, and potential evapotranspiration for the prediction date. **Antecedent conditions (6 channels):** 7-day and 30-day cumulative precipitation, 7-day and 30-day cumulative potential evapotranspiration, and 7-day and 30-day net water balance (precipitation minus evapotranspiration). **Seasonal encoding (2 channels):** Sine and cosine transformations of day-of-year to capture seasonal patterns without discontinuity at year boundaries. Including antecedent aggregates is essential for root-zone prediction because moisture at 20 cm depends on the *integrated* history of wetting and drying, not just same-day forcing. The 7-day windows capture short-term memory after precipitation events, while the 30-day windows capture slower seasonal transitions and persistent deficits.

3.6.3 Training Data Organization. Training samples are organized as patch-date pairs. Each sample corresponds to a specific 64×64 spatial patch on a specific date. Labels within each patch are sparse: only pixels coinciding with APSIM simulation locations have ground truth values. The patch index contains pre-computed valid locations where APSIM simulations provide sufficient label coverage. We create a binary mask indicating valid label locations and compute losses only over masked pixels. This sparse supervision approach allows the convolutional architecture to learn spatial patterns and propagate information to unlabeled regions. Operationally, each training instance consists of $(X_{t,p}, Y_{t,p}, M_{t,p})$, where $X_{t,p}$ is the 20-channel input tensor for patch p at date t, $Y_{t,p}$ is the target tensor containing APSIM-derived $\theta(20\text{ cm})$ values at labeled pixels (and placeholders elsewhere), and $M_{t,p}$ is the binary mask indicating which pixels are labeled. During optimization, gradients are computed only where $M_{t,p} = 1$, ensuring that the model is trained against physically derived ground truth while still producing dense outputs across the entire tile. This setup also encourages "spatial generalization within a tile": even though the loss is applied sparsely, the convolutional filters are shared across the full image, so the model learns to extrapolate from labeled pixels to nearby unlabeled pixels using consistent relationships between covariates and soil moisture.

For mixed-supervision UNet training, we construct the APSIM-derived and in-situ datasets using a consistent season-aware, date-based splitting strategy to ensure temporally disjoint evaluation while preserving seasonal coverage. Specifically, for each supervision source (APSIM simulations and in-situ observations), all labeled samples are first grouped by calendar year and then partitioned into the four meteorological seasons (DJF, MAM, JJA, SON). Within each season, the available dates (not pixels) are randomly assigned to train/validation/test according to fixed fractions (default 0.70/0.10/0.20), yielding three non-overlapping date lists per dataset. All labels and patches associated with a given date are restricted to that date's split, preventing leakage through shared days while ensuring that every split contains examples from each season. Mixed training is then implemented after these independent splits are formed: optimization steps draw minibatches from APSIM-train with specified probability (e.g. 0.2) and from in-situ otherwise, whereas model selection and reporting are performed exclusively on held-out in-situ validation/test dates, aligning evaluation with the most reliable observational reference while still benefiting from APSIM's dense spatiotemporal supervision during training.

3.7 Training With 20 cm Depth Physics-guided Composite loss [RQ-2, RQ-4]

We believe that training the 20 cm soil moisture model requires a loss function that balances data-driven learning with physical constraints. We design a composite loss that combines mean squared error against APSIM targets with physics-based regularization derived from the van Genuchten soil water retention model. This design addresses RQ-4 and RQ-2

simultaneously: APSIM provides depth-consistent supervision at sparse pixels, while physics-based regularization discourages the network from producing root-zone states that violate known hydraulic bounds or retention behavior. In practice, the MSE term drives agreement with the simulator where labels exist, and the physics term improves stability and generalization by constraining solutions in regimes that are underrepresented or noisy in the labeled set (e.g., very dry or near-saturated conditions). Because supervision is sparse (masked) within each tile, the loss is formulated to (1) compute data fidelity only at labeled pixels and (2) optionally impose physically meaningful constraints more broadly (at labeled pixels, or across all pixels if desired). This is important because the model must still output a dense 30 m field at inference time even though only a subset of pixels contribute direct supervision during training.

3.7.1 Composite loss formulation. The total loss combines a data-driven component and a physics-guided component through adaptive weighting:

$$L_{\text{total}} = \alpha \cdot L_{\text{MSE}} + (1 - \alpha) \cdot L_{\text{physics}} \tag{14}$$

Here, $\alpha \in [0, 1]$ controls the balance between fitting the APSIM targets and enforcing physical constraints. We adopt a convex combination so the objective remains well-scaled and interpretable: when $\alpha = 1$ the optimization reduces to purely data-driven fitting, and when $\alpha = 0$ the optimization emphasizes physical consistency (while still being computed in a differentiable manner through the same forward pass). The value of α changes during training according to a schedule described below.

3.7.2 Data-driven loss component. The MSE loss measures pixel-wise agreement between predictions and APSIM targets over valid (masked) pixels:

$$L_{\text{MSE}} = \frac{1}{N_{\text{valid}}} \sum_{i=1}^{N} m_i \cdot \left(\theta_i^{\text{pred}} - \theta_i^{\text{target}}\right)^2 \tag{15}$$

where θ_i^{pred} is the predicted soil moisture at pixel i, θ_i^{target} is the APSIM-simulated target, m_i is the binary mask (1 where labels exist, 0 otherwise), and $N_{\text{valid}} = \sum_i m_i$ is the count of valid pixels. It allows learning from spatially sparse supervision while the convolutional architecture propagates information to unlabeled regions. Masking is essential: without it, unlabeled pixels would introduce arbitrary gradients (from sentinel values or missing data), and the model could minimize loss by exploiting artifacts rather than learning physically meaningful relationships. To avoid instability when N_{valid} is small for a given tile-date sample, we enforce a minimum label count when constructing the patch index (as noted earlier) and compute the loss with safe normalization (e.g., adding a small ϵ to N_{valid}) so gradients remain bounded.

3.7.3 Physics-guided loss component. The physics loss enforces consistency with known soil hydraulic relationships. It comprises two terms: a bound penalty and a matric potential consistency term. Intuitively, these two terms play complementary roles: the bound penalty enforces feasibility (predictions must lie in physically permissible moisture ranges), while the matric potential term enforces shape/behavior (predicted moisture should correspond to a matric potential consistent with soil hydraulic parameters). This combination reduces the chance that the network "fits" APSIM targets at labeled pixels but produces unrealistic moisture states elsewhere in the tile.

Bound Penalty. Soil moisture must lie between the residual water content θ_r and saturated water content θ_s for a given soil type. We penalize predictions that violate these physical bounds:

$$L_{\text{bound}} = \frac{1}{N}\sum_{i=1}^{N}\left[\max(0, \theta_r^i - \theta_i^{\text{pred}}) + \max(0, \theta_i^{\text{pred}} - \theta_s^i)\right] \tag{16}$$

This term is zero when predictions respect the physical limits and increases linearly with the magnitude of violations. We compute (θ_r^i, θ_s^i) per pixel from soil hydraulic parameters inferred from POLARIS texture (via the same pedotransfer pipeline used to parameterize SWIM3). This makes the constraint spatially heterogeneous: sandy soils and clay-rich soils have different feasible moisture ranges, and the penalty respects that heterogeneity rather than imposing a single global bound. We apply L_{bound} across all pixels in the tile (unmasked) because feasibility is required everywhere, not only at labeled locations. This helps prevent spurious predictions in unlabeled regions, especially during early training when the network may produce out-of-range values.

Matric Potential Consistency. The van Genuchten equation (Equation 3) establishes a nonlinear relationship between soil moisture and matric potential. We compute matric potential from both predicted and target soil moisture values and penalize their disagreement:

$$L_\psi = \frac{1}{N_{\text{valid}}}\sum_{i=1}^{N} m_i \cdot \left|\frac{\psi_i^{\text{pred}}}{\psi_{\max}} - \frac{\psi_i^{\text{target}}}{\psi_{\max}}\right| \tag{17}$$

where ψ^{pred} and ψ^{target} are computed by inverting the van Genuchten equation:

$$\psi(\theta) = \frac{1}{\alpha_{\text{vg}}}\left[\left(\frac{\theta_s - \theta_r}{\theta - \theta_r}\right)^{n/(n-1)} - 1\right]^{1/n} \tag{18}$$

Here, α_{vg} and n are van Genuchten shape parameters obtained from the Rosetta pedotransfer function using POLARIS soil texture data. The normalization by $\psi_{\max}$ cm ensures numerical stability across the wide range of matric potential values. We compute ψ^{target} from θ^{target} (APSIM-derived) and ψ^{pred} from θ^{pred} (network output) using the same pixel-wise hydraulic parameters. This ensures the penalty reflects differences attributable to moisture state rather than differences in parameterization. Because the van Genuchten inversion can become numerically unstable as $\theta \to \theta_r$ (denominator approaches zero), we clamp θ to $[\theta_r + \epsilon, \theta_s - \epsilon]$ when computing $\psi(\theta)$, with a small ϵ (e.g., 10^{-4}) to maintain differentiability and avoid exploding gradients. We apply L_ψ only at labeled pixels (masked) so that matric potential consistency is anchored to the APSIM-supervised moisture states; this avoids over-constraining unlabeled regions where the network should be free to infer spatial structure from covariates.

The complete physics loss combines these terms:

$$L_{\text{physics}} = \lambda_b \cdot L_{\text{bound}} + L_\psi \tag{19}$$

where λ_b weights the bound penalty term and controls how aggressively the model is prevented from generating infeasible outputs. In practice, we set λ_b so that L_{bound} is comparable in magnitude to L_ψ during mid-training epochs; this prevents one physics component from dominating optimization simply due to scale differences.

3.7.4 Adaptive alpha scheduling. One of the key insights in our work is that the relative importance of data-driven and physics-based losses should evolve during training. We implement adaptive alpha scheduling where α decreases from an initial value (typically 1.0) toward 0 over the course of training. This schedule serves two purposes. Early in training, when model weights are randomly initialized, enforcing physics constraints can impede learning. Starting with $\alpha = 1$ allows the model to first learn spatial patterns from the data. As training progresses and predictions become more

accurate, decreasing α gradually introduces physics enforcement to refine predictions and improve generalization. We implement linear scheduling option $\alpha(t) = \alpha_0 \cdot \left(1 - \frac{t}{T_{\text{total}}}\right)$, where t is the current epoch and T_{total} is the total number of epochs.

This schedule mirrors the training dynamics of many constrained learning systems: first learn a reasonable data-consistent solution manifold, then "tighten" the manifold by introducing physically motivated penalties once the network outputs are within a plausible range. Empirically, this reduces optimization stiffness and avoids early-epoch failures where the model is pushed toward satisfying constraints before it has learned the basic mapping from covariates to moisture.

4 EMPIRICAL BENCHMARKS & EVALUATION

To assess the performance of our models, we performed empirical benchmarks over a 19-node cluster. Each node in our distributed cluster is equipped with an Intel Xeon E5-2620v3 processor, 64 GB of RAM, and a Quadro P2200 GPU with 5GB of memory and 1280 cores. Our approach involves the uniform distribution of datasets, organized by spatial clusters identified through quadkey assignments. Data originating from different sources but sharing the same quadkey are placed on the same machine to ensure data locality and minimize network I/O due to data transfers during model training.

We use the PyTorch framework and the Adam Optimizer to train all our models. We train our models for 1000-1500 epochs, leveraging a dataset comprising 13,000 randomly selected training samples. The input dataset used for training consists of soil moisture maps from the year 2019, focusing on areas close to in-situ stations, as visualized in Fig. 1. We evaluate the performance of our models using data from unfamiliar geographical locations.

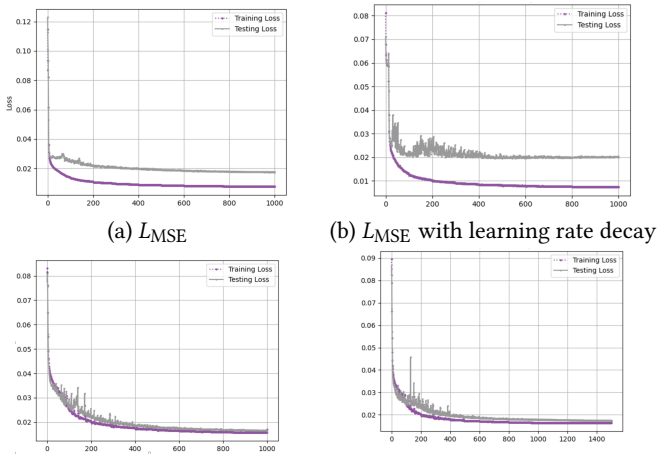
4.1 Loss function components [RQ-2]

We begin by assessing the significance of incorporating a science-guided loss function into our model training process. Our combined loss function comprises two primary components: the traditional MSE loss (L_{MSE}) and the science-guided loss component (L_{sci}). Each of the component's weights is dynamically adjusted while training. We start our analysis by individually evaluating the contribution of each of these loss terms.

In Fig. 5, we present the training and testing loss profiles under four different scenarios - (1) Training just with the MSE loss function (also BASELINE model i.e. without using science-guided learning), (2) Training with MSE loss while gradually reducing learning rate by a factor of 0.8 every 100 epochs using step scheduler, (3) Training with the MSE loss and scientific loss using ψ prediction based on HydroBlocks' model outputs, and (4) Training with the MSE loss and scientific loss using ψ prediction based on in-situ sensor observations.

The optimal model training results are achieved when the loss function combines (L_{MSE} and L_{sci}) using in-situ observations while reducing α parameter using step decay. This combination yields an outstanding overall testing loss of 0.017 (1.7%) and a testing PSNR of 34.48 dB. While a model trained only with the MSE loss exhibits relatively low testing MAE (1.9%) and PSNR accuracy of (33.92 dB), the images produced are slightly blurry and lack finer details.

An important observation is that while the loss value of a model trained with the MSE loss is considerably low, the image quality is quite low when assessing using measures such as PSNR and SSIM (structural similarity index measure; higher the better), which quantifies the fidelity of images. This is evident in Fig. 6, where 6(b) and (c) corresponds to an image generated by a model trained only with MSE in the loss function. The image appears blurry and lacking in finer, sharper spatial details such as road structures and landcover boundaries. Overall, the training and testing loss show that **much better generalization achieved using the scientific loss component, and the model does not overfit**.



(a) L_{MSE} (b) L_{MSE} with learning rate decay

(c) L_{MSE} and L_{sci} (HydroBlocks) (d) L_{MSE} and L_{sci} (In-situ sensors)

Fig. 5. Training and testing loss for models with different combinations of loss components. The training loss combines various loss terms that guide the model to convergence, while the testing loss is measured by the MAE between the predicted and target soil moisture maps on unseen locations.

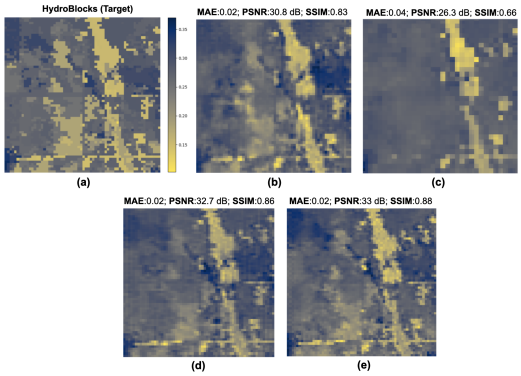


Fig. 6. Sample test images generated by models trained with different configurations: (a) Target SMC map; (b) L_{MSE}, (c) L_{MSE} with learning rate decay, (d) L_{MSE} and L_{sci} using HydroBlocks-based soil matric potential. and (d) L_{MSE} and L_{sci} using in-situ sensors observations.

This shows that DeepSoil is able to cope with limited training samples (due to the low number of in-situ sensors) while achieving a solid performance over entirely unseen locations. As can be seen, the SMC maps are much sharper than pure MSE-based loss functions.

We also observed an emerging behavior within our DNN models, as depicted in Fig. 7. These models were trained with a science-guided loss function, leveraging inputs from landcover maps, NDVI indices, and RGB bands extracted from Landsat imagery. Notably, we observed that approximately 1.1% of pixels in the HydroBlocks ground-truth dataset were missing, primarily associated with water bodies. While these missing pixels were intentionally excluded during the loss calculation process to decrease training errors, our DNN models exhibited a remarkable, emergent ability to

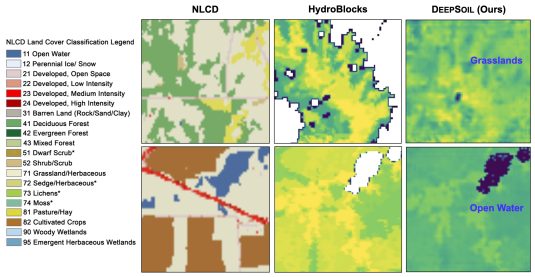


Fig. 7. Emergent behavior in DEEPSOIL: Regeneration of missing pixels based on landcover types.

regenerate these water bodies effectively. During the testing phase, the model indicated the ability to regenerate the missing pixels within the image and accurately distinguish soil moisture values based on the land cover type.

4.2 Clustering [RQ-3]

We did an accuracy analysis of our model, comparing its performance with and without the incorporation of our k-means++ clustering scheme. Specifically, we evaluate two different model training approaches: one involving a single global model trained on all locations in the training set but with fewer time steps and the other utilizing models trained on clustered spatial extents. The clustering process is based on a combination of soil, land, and hydrological properties, resulting in 19 distinct clusters.

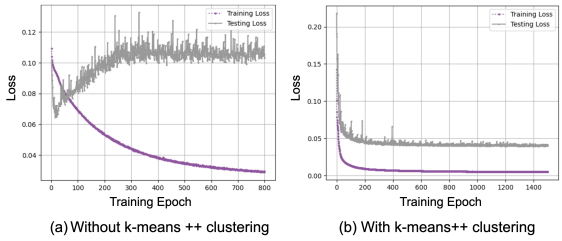


Fig. 8. Training and testing loss in models with and without clustering.

As illustrated in Fig. 8, the model trained without clustering indicates significantly higher errors. It converges slowly during training but also demonstrates poor performance on unseen data, with MAE errors increasing to 10-12%. Conversely, the models trained using multiple clusters exhibit faster convergence and consistently achieve lower test loss, with an average MAE of 5%. These models, when trained on clustered spatial extents, become more specialized as they focus on a smaller number of distinct spatial regions. This experiment highlights the advantages of our clustering-based approach in enhancing model performance and achieving better generalization overall.

4.3 SMC Maps at 30m Resolution [RQ-1, RQ-2]

Finally, we assess the overall effectiveness of our approach by contrasting soil moisture prediction errors with different variations of loss functions in DEEPSOIL, HydroBlocks, and the satellite-based SMAP 36 km data product. These assessments are performed by contrasting estimates with in-situ, ground truth measurements. This evaluation is performed over 100+ testing locations spanning the months of March-November of 2019, which were *unseen* during the

Table 3. Comparing SMAP, HydroBlocks, and variations of our model DeepSoil w.r.t to ground station data.

Model	MAE	Standard Deviations
SMAP	0.26	+/- 0.06
HydroBlocks	0.077	+/- 0.067
BASELINE (L_{MSE})	0.061	+/- 0.043
L_{MSE} with learning rate decay	0.16	+/- 0.062
L_{MSE} and L_{sci} (HydroBlocks)	0.0638	+/- 0.046
L_{MSE} and L_{sci} (In-situ sensors)	0.046	+/- 0.0331

model training phase. We report MAE and standard deviations in Table 3 for locations where ground-truth stations are available.

Notably, our DeepSoil has **a 4.6% soil moisture error when compared to in-situ station observations**. We achieve a (substantial) **82.3% reduction in SMC percentage error over SMAP** 9 km and **40.25% reduction in error over HydroBlocks**. As we experiment with different loss terms, the errors reduce significantly while incorporating scientific knowledge using van-Genuchten equation 3. Our model outperforms the BASELINE model that is based on regular MSE-based loss with a 24.59% reduction in errors. Our DeepSoil model reduces the weight of MSE errors and increases the overall impact of the scientific equation over training epochs in the loss function. For fair comparisons, we also performed experiment by training model using pure MSE-based learning but with reducing learning rates over epochs (using step learning rate decay). From Table 3, this variation leads to the worst performance with an increase in errors by 162.2% and 247.8% increase compared to BASELINE and our science-guided model respectively.

Lastly, we compare the model performance and show how incorporating in-situ sensors in scientific-loss term helps in reducing errors in HydroBlocks. Overall, we achieve 27.8% decrease in errors when in-situ sensor observations are leveraged to drive the model training. These results demonstrate the accuracy and precision of our model in predicting soil moisture. Our approach has a demonstrably high correlation with in-situ stations.

For a comprehensive assessment of DeepSoil's performance, we conducted an extensive evaluation encompassing the entire state of California and Oklahoma (depicted in Fig. 9 and 10). We compared our predictions with satellite-based SMC from SMAP at its original sensing resolution of 36 km. We generated the SMC map for the months of April (for CA, 2024) and July (for OK, 2023) because these months influence planting and irrigation plans. April marks the start of the growing season in temperate regions such as California. Similarly, mid-summer (July), marks a critical period for crops such as corn as they are in pivotal stage of development and are highly sensitive to available water content. Lack of water intake can lead to significant loss of crop yield. This evaluation involved utilizing datasets of over 200,000+ 64x64 pixel tiles, amounting to approximately 83 gigabytes of data. To contrast the scales involved: our model generated images for the area of 163,000 mi^2 for California with 35097x38100 pixels (compared to the 23x26 pixels for SMAP) and 24849x11930 pixels for Oklahoma (versus SMAP's 9x21 pixels).

SMAP data products include several flags regarding the poor quality of the data, especially for surfaces with permanent ice and snow, urban areas, wetlands, and proximity of large water bodies or coastlines [42, 62, 73]. Our experiment includes coastal regions and urban areas where SMAP is known to have high errors [15]. Our model performs well with the limited number of in-situ stations in this target region and demonstrates an **accuracy of higher than 99.7% across the areas unseen by the model**. In particular, this result highlights the effectiveness of our approach in the regions where SMAP underperforms significantly.

On the other hand, in Oklahoma, in-situ stations are near-uniformly dispersed throughout the entire state, providing rich data coverage for soil moisture mapping. Our model consistently demonstrates its capacity to generate high-quality

soil moisture predictions that closely align with SMAP observations that are very coarse in spatial resolution. Crucially, our approach captures subtle spatial variations in soil moisture, particularly in distinguishing between riverbeds and the surrounding land areas.

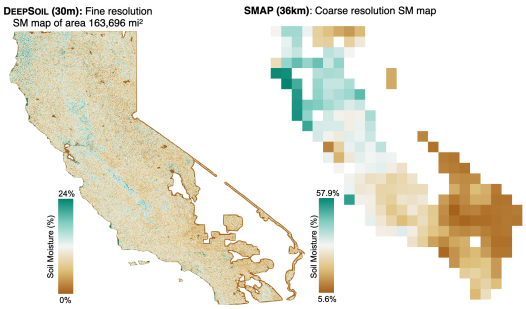


Fig. 9. Visualizations contrasting soil moisture maps for April 6th, 2024 over California generated by DeepSoil (30m) vs SMAP (36km sensed resolution).

We compared our predictions with the soil volumetric water content (VWC) data produced by Oklahoma State University at an 800m by 800m spatial resolution[61], as illustrated in Fig. 10(c). This map's generation process involves utilizing in-situ SMC measurements from the Oklahoma Mesonet for a regression kriging algorithm. This algorithm integrates soil texture information, specifically the percentage of sand, along with the precipitation index. The map depicts elevated levels of VWC in the western region of the state, with several localized hotspots (high VWC). However, this diverges from both SMAP satellite data and our prediction.

4.4 Predicting 20 cm root zone soil moisture content [RQ-4]

RQ-4 evaluation uses two complementary tracks. First, we measure how well the learned mapper reproduces mechanistic root-zone dynamics by training and evaluating against APSIM/SWIM3-derived targets; this serves as a mechanistic-consistency baseline (upper bound) and validates that the architecture, covariates, and loss can recover depth behavior at 30 m when dense supervision is available. Second, we evaluate our mixed-supervision, observation-calibrated variant (see section 4.4.2) in which in-situ measurements provide the primary calibration signal; this is the primary external-validity evaluation, but it is expected to yield lower headline metrics due to sparse and noisy station labels and point-to-pixel representativeness differences. Because the two tracks use different target sources (and therefore different label distributions and sparsity), their metrics should not be interpreted as directly comparable.

APSIM simulation locations were selected using a multi-stage, coverage-oriented sampling pipeline designed to yield statewide sets of points that are representative across land cover, climate regimes, soil texture, and meteorological

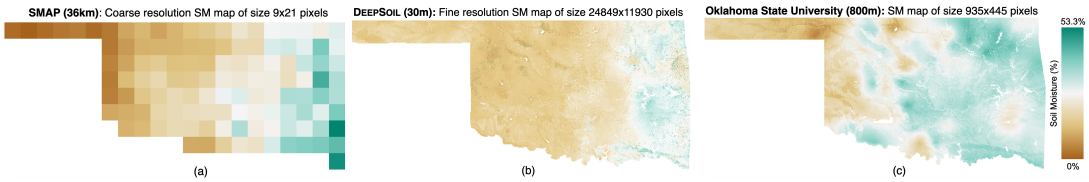


Fig. 10. Visualizations contrasting soil moisture maps over Oklahoma for 17th July 2023 generated by - (a) SMAP at 36 km resolution, (b) DeepSoil at 30m spatial resolution, and (c) A kriging-based data product for the state generated by OSU.

forcing grids. We first construct a Colorado-wide candidate pool by sampling valid-land pixels on the 30m master grid and tagging each candidate with its NLCD land-cover class. Candidates are then intersected with the Köppen climate raster to form discrete NLCD×Köppen strata, enabling controlled allocation across both common and rare land-cover–climate combinations. From this stratified pool, we select a fixed annual budget of approximately 500,000 unique APSIM locations using proportional allocation across strata so that no single regime dominates the simulation supervision. Within large strata, representative points are chosen using MiniBatchKMeans clustering over a compact feature vector that captures soil texture (sand/silt/clay), topographic context (elevation), and GridMET spatial indexing (GridMET row/column/cell_id), which encourages coverage across both soil–terrain variability and the discrete meteorological forcing grid. Smaller strata, where clustering is unnecessary or unstable, use randomized selection while preserving the global proportional allocation. We repeat this procedure independently for each study year, producing ∼500k locations for 2023 and a separate ∼500k locations for 2024, and we simulate each annual set under the same APSIM configuration for its corresponding year.

4.4.1 Evaluation of APSIM-only supervision training. Having defined (1) APSIM–SWIM3 as a physics-based generator of depth-specific supervisory targets and (2) a tile-based U-Net trained under sparse supervision, we now evaluate whether the resulting model can recover temporally and spatially coherent 20 cm soil moisture fields on unseen conditions. We evaluate the performance of our 20 cm root zone soil moisture prediction model on a held-out test set. The test set comprises all data from October through December 2024, which was never used during model training. This temporal split ensures evaluation on genuinely unseen conditions and prevents data leakage from future to past. The training data spanned January 2023 through September 2024, covering a full seasonal cycle. The test set contains approximately 2.9 million labeled pixels throughout the 91-day evaluation period. We compute aggregate metrics by comparing model predictions against APSIM-simulated ground truth values at all labeled locations. For each test date we run inference over the set of valid patch-date pairs, apply the binary mask m_i to align predictions with available APSIM targets, and compute per-pixel residuals in volumetric water content units (m^3/m^3). We then aggregate errors across space and time by pooling all labeled pixels in the evaluation window, which yields a single summary MAE/RMSE and a global R^2 over the full test period. This pooling-based aggregation is appropriate in our setting because labeled pixel density varies across patches, and it naturally weights regions by the amount of available supervisory information.

The model achieves a Mean Absolute Error (MAE) of 0.012 in volumetric water content units, computed across all test pixels. This means that for a typical soil moisture value of 21%, our prediction deviates by only 1.2 percentage points on average. The Root Mean Square Error (RMSE) is 0.016. The coefficient of determination $R^2 = 0.882$ indicates that the model explains approximately 88% of the variance in root zone soil moisture across diverse soil types, terrain conditions, and weather patterns in Colorado. The gap between RMSE and MAE suggests that large residuals are relatively infrequent, indicating that the model's errors are not dominated by a small number of extreme failures (e.g., isolated high-variance events). Fig. 11 presents an example soil moisture map (at 20 cm depth) generated by our model for October 15th, 2024. The 30-meter resolution captures fine-scale spatial heterogeneity driven by terrain, soil texture, and land cover variations. The map reveals coherent spatial patterns that align with topographic features and land use boundaries across Colorado. This qualitative behavior is consistent with the model's design: static covariates (texture/terrain/land cover) provide the spatial scaffold, while dynamic meteorological forcings modulate daily departures from that scaffold, enabling depth-consistent root-zone responses rather than purely surface-correlated patterns.

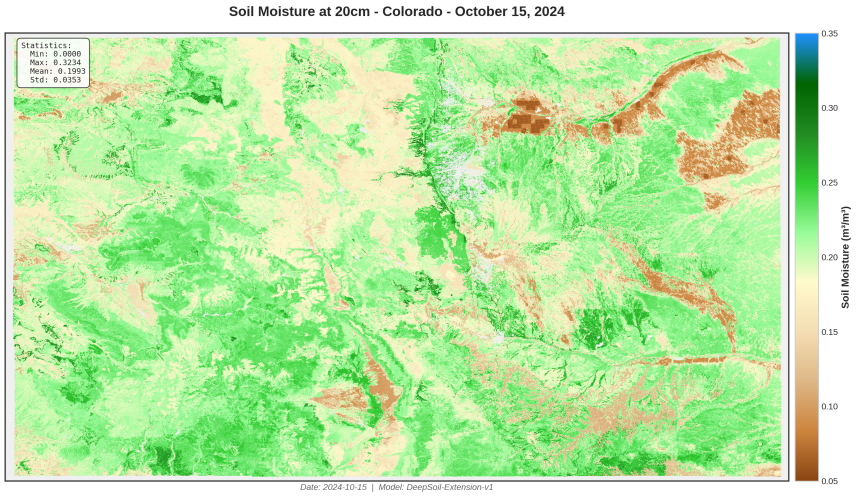


Fig. 11. Predicted soil moisture map at 20 cm depth for October 15th, 2024 over Colorado generated by DeepSoil Extension at 30-meter resolution.

The proposed extension demonstrates the key strength. It uses APSIM simulations as ground truth for model training rather than relying solely on sparse in-situ measurements. APSIM can be executed at any location given soil texture data and meteorological forcing. This means our methodology is not constrained to regions with dense sensor networks. The framework can be extended to any state or region where POLARIS soil data and GridMET weather data are available. In particular, APSIM provides a physically grounded mechanism for generating depth-specific supervision at scale, while the trained U-Net amortizes that computational cost by learning a fast mapping from widely available gridded covariates to 30 m root-zone fields. This pairing is central to RQ-4: it reconciles depth-aware physical consistency with the operational need to produce daily, high-resolution maps over large spatial extents.

Table 4. Soil moisture label distributions (APSIM vs in-situ, full scan).

Dataset	Rows	Mean	Std	Min	Max
APSIM	358,708,860	0.234371	0.044081	0.035750	0.469606
In-situ	35,457	0.185810	0.119462	0.000000	0.996000

Dataset	P1	P5	P50	P95	P99
APSIM	0.122681	0.166204	0.232412	0.310158	0.340769
In-situ	0.000000	0.023612	0.176000	0.384750	0.500000

Metric	Value
Histogram L1 distance	1.091946

4.4.2 Mixed-Supervision Training for Statewide 30 m Resolution Root Zone (20 cm) Soil Moisture Maps (2023–2024). In this subsection, we evaluate the end-to-end behavior of our 20 cm root zone modeling under two competing goals: (1) statewide spatial completeness at 30 m spatial resolution and daily cadence, and (2) fidelity to the most reliable observational reference available. Because mechanistic simulations and field sensors provide complementary but

Table 5. Mixed training performance on in-situ data.

Run	MAE_{train}	RMSE_{train}	r_{train}	MAE_{test}	RMSE_{test}	r_{test}
20% in-situ and 80% APSIM	0.0572	0.0748	0.3231	0.0490	0.0721	0.0935
50% in-situ and 50% APSIM	0.0453	0.0570	0.6938	0.0433	0.0568	0.6743
100% in-situ	0.0542	0.0715	0.4260	0.0444	0.0688	0.2662
80% in-situ and 20% APSIM	0.0429	0.0567	0.6986	0.0377	0.0552	0.6095

Table 6. Test performance with mixed-source ratios by season

Run	Season	MAE	RMSE	r
80% in-situ and 20% APSIM	DJF	0.0161	0.0414	0.2439
80% in-situ and 20% APSIM	MAM	0.0808	0.0899	0.6669
80% in-situ and 20% APSIM	JJA	0.0348	0.0451	0.7981
80% in-situ and 20% APSIM	SON	0.0226	0.0264	0.1944
50% in-situ and 50% APSIM	DJF	0.0329	0.0433	0.3431
50% in-situ and 50% APSIM	MAM	0.0838	0.0931	0.6887
50% in-situ and 50% APSIM	JJA	0.0310	0.0391	0.7809
50% in-situ and 50% APSIM	SON	0.0280	0.0346	0.2591
20% in-situ and 80% APSIM	DJF	0.0166	0.0367	0.1924
20% in-situ and 80% APSIM	MAM	0.1170	0.1239	0.6833
20% in-situ and 80% APSIM	JJA	0.0492	0.0662	0.6740
20% in-situ and 80% APSIM	SON	0.0189	0.0236	0.2800
100% in-situ	DJF	0.0084	0.0149	0.2910
100% in-situ	MAM	0.1198	0.1285	0.6343
100% in-situ	JJA	0.0437	0.0535	0.6924
100% in-situ	SON	0.0118	0.0140	0.3230

imperfect signals, we benchmark training strategies that reconcile coverage-driven supervision with observation-driven calibration, and we summarize the resulting statewide soil moisture root zone maps for 2023 and 2024.

Calibration supervision from in-situ sensors. Given the objective of ensuring model accuracy, we consider in-situ observations to be the most reliable reference for the target variable. This assessment is consistent when compared with outputs from physics-based simulations. As depicted in Table 4, the physics-based model does not fully reproduce the complete range of soil moisture content (SMC) captured by in-situ sensors, particularly under extremely dry and wet conditions, although it effectively fills spatiotemporal gaps in the observations. This limitation motivates a carefully designed training strategy that leverages both in-situ sensor data and simulation outputs. In particular, we treat in-situ measurements as the primary calibration signal for model selection and reporting: hyperparameters, mixing schedules, and early-stopping decisions are chosen to optimize agreement with held-out in-situ dates. This choice reflects the fact that station measurements capture local extremes and event responses (e.g., rapid wet-up after precipitation and dry-down under strong evaporative demand) that may be damped by model structural assumptions, forcing uncertainty, or pedotransfer parameterization in the mechanistic pipeline. Mechanistic simulations provide dense spatiotemporal supervision that encodes physically consistent vertical redistribution and drainage dynamics, but may exhibit amplitude compression relative to observations. Our benchmarks, therefore, focus on strategies that preserve the spatial continuity learned from simulation while correcting systematic bias using sparse but high-fidelity station targets.

Batch-level mixing. The APSIM simulation labels encourage physically coherent spatiotemporal structure and provide statewide coverage; however, they can be biased and may underrepresent extremes. Conversely, the in-situ labels are accurate but extremely sparse. Also, in-situ sensor networks are often deployed to serve specific objectives (e.g., agriculture or mountain snowpack monitoring) and therefore do not provide representative coverage across large spatial extents. Batch-level mixing provides a direct balance between training the model with high fidelity to sparse but accurate targets and capturing geospatial variability from broadly representative simulation outputs.

From an optimization perspective, batch-level mixing avoids two failure modes that we observed in ablations: (1) APSIM-only training can converge to a solution that is spatially smooth and physically plausible but systematically biased relative to station observations; and (2) in-situ-only training can overfit to a small number of station-dominated environments and exhibit unstable updates due to extreme label sparsity (high variance gradients). Interleaving batches provides frequent "structure" updates (APSIM) and frequent "calibration" updates (in-situ), yielding a controllable bias-variance trade-off governed by p. Batch-level mixing provides a direct and controllable trade-off: larger p emphasizes simulation-driven coverage and structure, while smaller p emphasizes observational calibration. In practice, we select p based on in-situ validation performance.

Mixed-source training at the batch-level. Mixed-source training is implemented by interleaving APSIM-sourced and in-situ-sourced minibatches within a *single* training run. Let $p \in [0, 1]$ denote the probability of sampling an APSIM batch at any optimization step. At each step, we draw:

- an APSIM batch with probability p, sampled from the training split generated by APSIM simulations; or
- an in-situ batch with probability $1 - p$, sampled from the training split captured by in-situ sensors.

APSIM batches provide spatially dense supervision (mask coverage is high across the tile) and are computed from the same covariate stack used at inference time (soil texture/terrain/landcover + GridMET-derived dynamics). In-situ batches are sparse and are masked to station-aligned pixels (or small neighborhoods around stations, when applicable), so gradients are driven by high-precision targets but at far fewer labeled locations. This approach differs from simply concatenating datasets across the training process, where one dataset is used in the early stages of training and another is used solely to refine model performance. Rather than replacing the broader training samples from APSIM, we adjust the sampling rate as the probability, p is reduced. Meanwhile, samples collected from in-situ sensors are incorporated throughout training to continuously correct bias. Each dataset is split independently *by date* (rather than by pixel) to assess temporal generalization, and validation and test performance are reported using the in-situ split by default to better reflect real-world conditions.

Tables 4–6 summarize the central challenge of statewide 30 m root-zone mapping under mixed supervision: the two label sources provide *complementary* information, but they differ substantially in both coverage and value range. Table 5 reports in-situ performance for four training regimes, spanning APSIM-heavy mixtures (20% in-situ / 80% APSIM) to in-situ-only supervision. A key observation is that the best overall in-situ generalization is achieved by *mixed* training, with the 80% in-situ / 20% APSIM configuration attaining the lowest test error ($\mathrm{MAE}_{test} = 0.0377$, $\mathrm{RMSE}_{test} = 0.0552$) while maintaining strong correlation ($r_{test} = 0.6095$). Importantly, this improvement is not simply due to in-situ supervision: the in-situ-only setting yields a higher MAE_{test} (0.0444) and substantially lower r_{test} (0.2662), consistent with the instability expected under extreme label sparsity. Conversely, the APSIM-heavy mixture exhibits low-to-moderate errors but weak correlation (e.g., $r_{test} = 0.0935$ at 80% APSIM), highlighting that physics-guided supervision alone can produce spatially plausible maps yet remain miscalibrated relative to station dynamics. Table 5

empirically supports the premise of batch-level mixing: APSIM batches supply dense spatiotemporal structure while in-situ batches continually re-anchor the mapping to observed magnitudes and event responses.

Seasonal robustness under temporally held-out evaluation. Table 6 decomposes test performance by climatological season, which is crucial because soil moisture exhibits distinct regimes across Colorado (snow-dominated winter recharge, spring wet-up, summer dry-down under high evaporative demand, and autumn transition). Two patterns are noteworthy. First, correlations are consistently highest in MAM and JJA across mixtures (often $r \approx 0.67$–0.80), indicating that the model captures the dominant hydroclimatic drivers during active wetting/drying seasons. Second, DJF and SON correlations are generally lower despite small MAE in some cases, a behavior that is expected when the absolute variance of the target is reduced (or when snow/freeze-thaw processes and sub-grid heterogeneity introduce observation noise not fully represented in the forcing). The mixed configurations preserve strong warm-season skill while avoiding the pronounced degradation observed when relying exclusively on one supervision source, suggesting that mixed training yields a more seasonally stable mapping function under unseen date sequences.

Mixed supervision results interpretation Table 4 quantifies the distribution shift between APSIM-derived labels and in-situ observations. APSIM labels are narrowly distributed (std = 0.0441; max = 0.4696) relative to in-situ (std = 0.1195; max = 0.9960), with in-situ exhibiting near-zero dry states and high wet extremes that the mechanistic pipeline rarely attains. This mismatch is non-trivial: optimizing against a narrow-label surrogate typically induces amplitude compression and under-prediction of extremes when evaluated on a broad-label target. The histogram L1 distance of 1.0919 further indicates substantial divergence in label mass. In this context, achieving the best in-situ test performance with mixed supervision (Table 4), while retaining statewide spatial completeness learned from APSIM, is particularly compelling: the model must simultaneously (1) inherit spatial priors and continuity from dense simulation labels and (2) learn to expand its response range to match observed variability from sparse stations. The gains of the mixed configurations therefore reflect effective reconciliation of two imperfect signals rather than incremental tuning on a single homogeneous dataset.

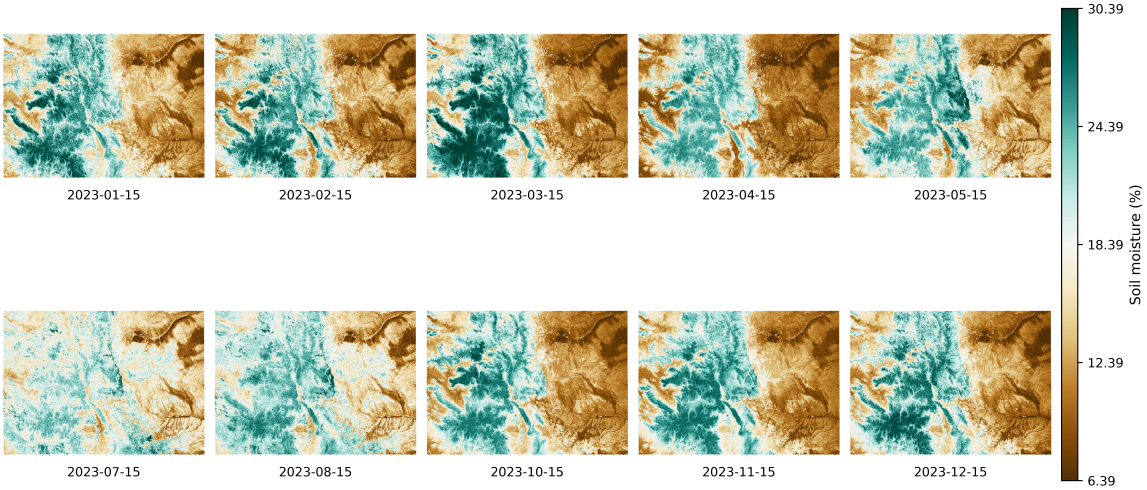


Fig. 12. Predicted soil moisture map at 20 cm depth for 2023 over Colorado generated by DeepSoil++ at 30-meter resolution.

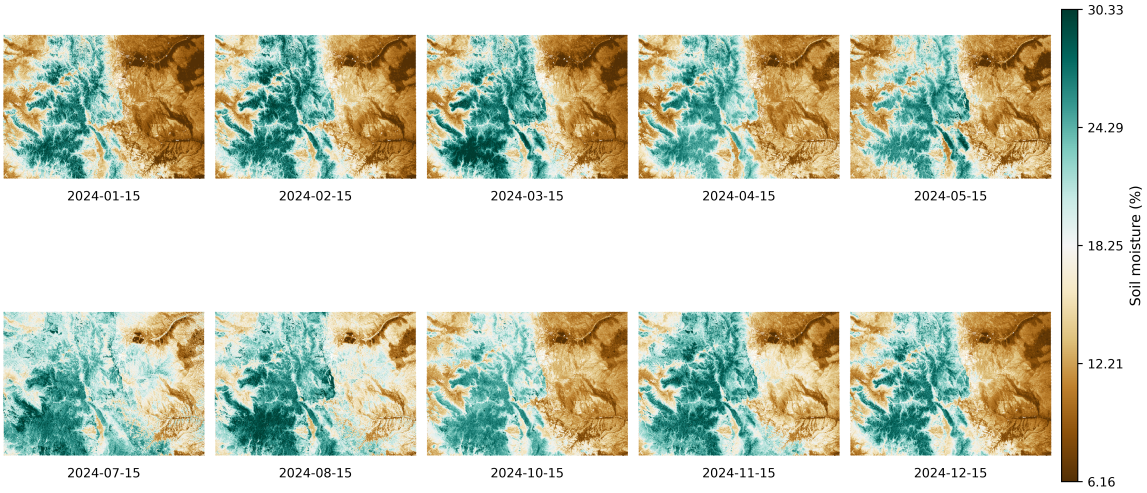


Fig. 13. Predicted soil moisture map at 20 cm depth for 2024 over Colorado generated by DeepSoil++ at 30-meter resolution.

Overall, the tables indicate that the best-performing regime is neither purely simulation-driven nor purely observation-driven, but a calibrated mixture that leverages APSIM for coverage and in-situ sensors for realism. This finding motivates the batch-level mixed training design described next.

4.4.3 Results summary (mixed-source vs. single-source training). Across experiments, mixed training consistently outperforms observational fidelity relative to APSIM-only training, while retaining the broad spatial variability patterns and seasonal dynamics learned from simulation (Table 6). When evaluated on held-out in-situ dates, mixed-source training reduced systematic bias and effectively captures the relations across observations compared to simulation-only baselines. Also, training with mixed-source data resulted in more stable learning and effective convergence compared to training with in-situ data alone. Incorporating APSIM batches in the mixed-source training regularizes the learning process by providing dense supervision across a broader geospatial extent and a wider range of dates.

Figures 12 and 13 visualize statewide 30 m predictions of volumetric soil moisture at 20 cm depth (root-zone) for Colorado, reported in percent (%). Each subfigure is a mid-month snapshot (date shown under the map). For readability, the figures use a 1st–99th percentile stretch, meaning the color scale is anchored to the central distribution of predicted values so that statewide spatial structure is visible without being dominated by rare extremes; thus, the maps are best interpreted as showing spatial gradients and seasonal evolution rather than emphasizing a few outlier pixels.

Across both years, the maps exhibit a coherent and physically plausible west–east moisture contrast, with higher root-zone moisture in the higher-elevation Rocky Mountain headwaters and lower moisture over the eastern plains and other drier regions, while month-to-month changes reflect the expected seasonal wetting/drying cycle (springtime recharge and broader wetness in many mountain and adjacent regions versus more widespread summer drying, particularly over the plains).

We report metrics on temporally held-out in-situ dates to emphasize real-world forecasting conditions (unseen weather sequences and antecedent moisture states). We additionally inspect map-level behavior such as spatial coherence across elevation and landcover transitions, absence of unphysical saturation artifacts, and continuity across tile boundaries; this is because statewide maps stress both the learned spatial priors and the robustness of the masking

strategy. Topographical characteristics in the 2023 and 2024 mosaics are well preserved. Terrain- and climate-aligned regions consistently follow spatial gradients while remaining responsive to seasonal wetting and drying. These results demonstrate that mixed-source training effectively transfers the mechanistic model's spatial completeness into a data-calibrated mapping product.

5 RELATED WORK

Estimation of soil moisture maps at high precision for different spatiotemporal scopes is critical given its crucial role in diverse applications ranging from precision agriculture to hydrological modeling. Research efforts can be broadly categorized as being based on machine learning techniques, conventional physics-based models, and hybrid methods that seek to integrate both paradigms.

Numerical models: Physics-based approaches use fundamental equations to predict soil moisture dynamics. For example, ParFlow, Hydrus-1D/2D, pedotransfer functions [13, 16, 45, 53, 55, 65, 66, 72] are numerical hydrology models that employ Richards' equations to simulate 2D/3D subsurface flows. Richards' equation [67] is a partial differential equation that estimates water flow through unsaturated soils as a function of space and time. The equation considers various factors, including soil properties, hydraulic conductivity, and initial moisture content. These 2D/3D models account for surface and subsurface water flow, as well as land surface processes (CLM) such as evapotranspiration and snow. While it is very effective for simulations, its resolution is limited by the resolution (spatial) of input datasets. Furthermore, parameterizing the governing equations can be a challenge alongside the substantial computational requirements. These models also have known deficiencies in estimating soil moisture values over water-stressed regions.

Machine Learning Approaches: Most recent efforts have focused on utilizing time-series deep neural networks [34, 36, 46], such as the dynamic NN in Adeyemi et al. [2], offering a one-day-ahead irrigation-oriented soil moisture predictions using LSTM networks while leveraging dynamic and static climate features. Peng et al.'s [64] introduces the External Parameter Orthogonalization-Partial Least Squares model, improving soil moisture predictions by mitigating the effect of parameters such as soil salinity.

In a comparative study, Senyurek et al. [70] assess the performance of Artificial Neural Networks (ANN), Random Forests (RF), and Support Vector Machines (SVM) for soil moisture prediction. Other efforts involve downsampling satellite-based measurements from Landsat and SMAP-based soil moisture using RFs [81, 82]. Ge et al. [27] propose a method that combines UAV data with Extreme Learning Machines (ELM) for automated parameter selection. While the above techniques demonstrate solid performance, often achieving soil moisture errors within the range of 5-10%, their applicability at larger scales poses a significant challenge. Finally, point-based prediction and evaluations at smaller spatial scales restrict the generalizability and scalability of these models.

Physics-informed machine learning models: To address these limitations efforts have leveraged knowledge-guided machine learning in other domains. For example, Karpatne et al. [17, 35] proposed an approach that integrates physics-based principles and domain knowledge into machine learning models while predicting lake water temperature with minimal in-situ measurements. This involves modifications to loss functions, network architectures, and ingesting numerically simulated input datasets. Such knowledge-guided efforts have shown promise in various applications, including weather and climate modeling [84] and carbon cycling [47]. Some physics-informed machine learning models [5, 6, 18, 28] formulate water flow by utilizing numerical models outputs to solve Richard's equation in the loss function. In super resolving tasks, machine learning models like GANs, Pix2Pix networks, and ESRGAN are often coupled with PDE-constrained loss functions by penalizing errors based on energy spectrum differences [49, 78]. Several efforts have explored physics/knowledge-guided machine learning (KGML) methods for modeling spatiotemporally evolving

phenomena. These methods often operate in concert with mechanistic or process-based domain-theoretic models, and frequently incorporate multipart loss functions that encode scientific knowledge directly into model training. Examples include physical constraints from soil hydrology, such as the van Genuchten water-retention equations and hydraulic-conductivity models based on Richards' Equation [38–40]; vegetation indices [11]; evapotranspiration dynamics [3]; preservation of graph-theoretic properties such as betweenness centrality [51]; cloud-occlusion masking in satellite imagery [37]; perceptual constraints for visualization [56]; soil salinity estimation [20]; hyperspectral imagery modeling [26, 52]; correlations among soil spectroscopic properties [4], and time-series projections [19, 21].

In our research, we adopt a science-guided approach that combines the U-Net machine learning model with a science-informed loss function. This combination allows us to predict high-precision soil moisture maps at 30m resolution at scale.

6 CONCLUSIONS & FUTURE WORK

In this study, we described our methodology to combine remote sensing, in-situ sensing, and scientific models to produce high-precision soil moisture maps. Our results demonstrate the effectiveness and potential of our model for accurate and precise soil moisture mapping across diverse geographic regions.

RQ1: Our approach dynamically assigns higher weights to loss values in the surroundings of in-situ stations. This tailored approach ensures that the model training is guided by ground truth data. Despite the sparsity of in-situ data, our model successfully integrates essential information, enabling the generation of large-scale soil moisture maps. Our empirical evaluation demonstrates that our approach achieves soil moisture predictions with a very low 4% MAE when compared to in-situ station measurements. For deeper root-zone depths where in-situ coverage is insufficient, we complement this reconciliation strategy with APSIM-supervised learning (RQ-4), thus extending DeepSoil beyond surface moisture while preserving physically consistent behavior.

RQ2: Our weighted multipart loss function assimilates scientific insights and conventional loss functions to regulate learning and enforce physically plausible soil moisture behavior. In particular, we incorporate the van Genuchten water-retention relationship to couple predicted soil moisture with soil hydraulic parameters, and we dynamically schedule the loss weights so the model first learns data-driven spatial structure and then increasingly emphasizes physics consistency. This design improves numerical stability, reduces overfitting under sparse supervision, and yields sharper, more coherent maps that preserve landcover and terrain-driven boundaries while remaining consistent with known soil hydraulic constraints.

RQ3: Our approach of clustering spatial extents based on soil properties, monthly soil moisture averages, landcover types, and other land-surface properties proved to be effective in reducing training errors when compared to a single global model trained for CONUS. Additionally, the clustered approach significantly reduces the required memory footprint and training times by converging faster. This allows our models to scale while tuning individual models for specific spatial regions. By constraining each model to a hydrologically coherent regime, clustering also reduces spurious averaging across incompatible climates/soils and improves generalization on unseen tiles within a cluster. In effect, clustering provides a practical bias–variance tradeoff: we retain regional specialization without incurring the cost of training and serving a fully bespoke model per location.

RQ4: Our extension of the model to a 20 cm root-zone depth demonstrates that physically grounded, depth-specific supervision can be generated at scale using APSIM–SWIM3 and translated into high-resolution (30 m) root-zone soil moisture fields through tile-based learning with sparse (masked) targets. By aligning APSIM point outputs to the 30 m grid, training a lightweight U-Net under masked losses, and regularizing predictions with a physics-guided composite

objective (bounds and matric-potential consistency), we achieve accurate root-zone estimates on temporally unseen conditions while maintaining physical plausibility across heterogeneous soils. This enables sub-surface mapping in regions without dense in-situ networks.

As part of future work, we will extend DeepSoil to multi-depth soil moisture mapping through the root zone, targeting depths from 20 cm down to 60 cm, by coupling depth-specific supervision with cross-depth consistency constraints. We note that 50–60 cm spans the lower root zone, where plants draw on stored moisture during dry spells, making it a strong indicator of drought resilience and irrigation need. In particular, we will (1) expand APSIM configurations beyond fallow to incorporate crop uptake and management scenarios where appropriate, (2) introduce a vertical-coherence loss that penalizes physically implausible gradients and enforces consistency between shallow and deeper layers under wetting and drying events, and (3) quantify uncertainty by comparing ensembles across cluster-specific models and physics-weight schedules. Finally, we will validate transferability across additional states and hydroclimatic regimes and assess how 60 cm predictions improve downstream tasks such as irrigation decision support, yield-risk monitoring, and drought early warning.

7 ACKNOWLEDGMENTS

This research was supported by the National Institute of Food Agriculture (COL014021223, 2025-77039-45531), the National Science Foundation (1931363, 2312319), an NSF/NIFA Artificial Intelligence Institutes AI-LEAF Award [2023-03616], and a Clare Booth Luce Professorship. This research is made possible by the New York State (NYS) Mesonet. Original funding for the NYS Mesonet was provided by Federal Emergency Management Agency grant FEMA-4085-DR-NY, with the continued support of the NYS Division of Homeland Security & Emergency Services; the State of New York; the Research Foundation for the State University of New York (SUNY); the University at Albany; the Atmospheric Sciences Research Center (ASRC) at the University at Albany.

REFERENCES

[1] John T Abatzoglou. Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology* 33, 1 (2013), 121–131.

[2] Olutobi Adeyemi, Ivan Grove, Sven Peets, Yuvraj Domun, and Tomas Norton. Dynamic neural network modelling of soil moisture content for predictive irrigation scheduling. *Sensors* 18, 10 (2018), 3408.

[3] Samuel Armstrong, Paahuni Khandelwal, Dhruv Padalia, Gabriel Senay, Darin Schulte, Allan Andales, F Jay Breidt, Shrideep Pallickara, and Sangmi Lee Pallickara. Attention-based convolutional capsules for evapotranspiration estimation at scale. *Environmental Modelling & Software* 152 (2022), 105366.

[4] Andrei Bachinin, Rupasree Dey, Paahuni Khandelwal, Sam Leuthold, M Francesca Cotrufo, Shrideep Pallickara, and Sangmi Lee Pallickara. Science-Informed Multitask Transformer for Soil Property Prediction from FTIR Spectroscopy. In *2025 IEEE International Conference on eScience (eScience)*. IEEE, 48–57.

[5] Toshiyuki Bandai and Teamrat A Ghezzehei. Physics-informed neural networks with monotonicity constraints for Richardson-Richards equation: Estimation of constitutive relationships and soil water flux density from volumetric water content measurements. *Water Resources Research* 57, 2 (2021), e2020WR027642.

[6] Toshiyuki Bandai and Teamrat A Ghezzehei. Forward and inverse modeling of water flow in unsaturated soils with discontinuous hydraulic conductivities using physics-informed neural networks with domain decomposition. *Hydrology and Earth System Sciences* 26, 16 (2022), 4469–4495.

[7] Jesse E Bell, Michael A Palecki, C Bruce Baker, William G Collins, Jay H Lawrimore, Ronald D Leeper, Mark E Hall, John Kochendorfer, Tilden P Meyers, Tim Wilson, et al. US Climate Reference Network soil moisture and temperature observations. *Journal of Hydrometeorology* 14, 3 (2013), 977–988.

[8] Gary Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* 25, 11 (2000), 120–125.

[9] Fred V Brock, Kenneth C Crawford, Ronald L Elliott, Gerrit W Cuperus, Steven J Stadler, Howard L Johnson, and Michael D Eilts. The Oklahoma Mesonet: a technical overview. *Journal of Atmospheric and Oceanic Technology* 12, 1 (1995), 5–19.

[10] Jerald A Brotzge, J Wang, CD Thorncroft, E Joseph, N Bain, N Bassill, N Farruggio, JM Freedman, K Hemker Jr, D Johnston, et al. A technical overview of the New York State Mesonet standard network. *Journal of Atmospheric and Oceanic Technology* 37, 10 (2020), 1827–1845.

[11] Kevin Bruhwiler, Paahuni Khandelwal, Daniel Rammer, Samuel Armstrong, Sangmi Lee Pallickara, and Shrideep Pallickara. Lightweight, embeddings based storage and model construction over satellite data collections. In *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 246–255.

[12] Todd G Caldwell, Tara Bongiovanni, Michael H Cosh, Thomas J Jackson, Andreas Colliander, Charles J Abolt, Richard Casteel, Toti Larson, Bridget R Scanlon, and Michael H Young. The Texas soil observation network: A comprehensive soil moisture dataset for remote sensing and land surface model validation. *Vadose Zone Journal* 18, 1 (2019), 1–20.

[13] Nathaniel W Chaney, Peter Metcalfe, and Eric F Wood. HydroBlocks: A field-scale resolving land surface model for application over continental extents. *Hydrological Processes* 30, 20 (2016), 3543–3559.

[14] Nathaniel W Chaney, Eric F Wood, Alexander B McBratney, Jonathan W Hempel, Travis W Nauman, Colby W Brungard, and Nathan P Odgers. POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. *Geoderma* 274 (2016), 54–67.

[15] Andreas Colliander, Thomas J Jackson, Rajat Bindlish, S Chan, N Das, SB Kim, MH Cosh, RS Dunbar, L Dang, L Pashaian, et al. Validation of SMAP surface soil moisture products with core validation sites. *Remote sensing of environment* 191 (2017), 215–231.

[16] Monidipa Das and Soumya K Ghosh. FB-STEP: a fuzzy Bayesian network based data-driven framework for spatio-temporal prediction of climatological time series data. *Expert Systems with Applications* 117 (2019), 211–227.

[17] Arka Daw, Anuj Karpatne, William D Watkins, Jordan S Read, and Vipin Kumar. Physics-guided neural networks (pgnn): An application in lake temperature modeling. In *Knowledge Guided Machine Learning*. Chapman and Hall/CRC, Boca Raton, FL, 353–372.

[18] Ivan Depina, Saket Jain, Sigurdur Mar Valsson, and Hrvoje Gotovac. Application of physics-informed neural networks to inverse problems in unsaturated groundwater flow. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards* 16, 1 (2022), 21–36.

[19] Rupasree Dey, Andrei Bachinin, Tanjim Bin Faruk, Abdul Matin, Yao Zheng, Mu Hong, Shrideep Pallickara Pallickara, and Sangmi Lee Pallickara. XFORMER: A Multi-Stage Uncertainty-Guided Deep Learning Framework for Time Series Extreme Event Forecasting. In *Proceedings of the IEEE Conference on Artificial Intelligence (CAI)*. IEEE, Granada, Spain.

[20] Rupasree Dey, Abdul Matin, Everett Lewark, Tanjim Bin Faruk, Andrei Bachinin, Sam Leuthold, M Francesca Cotrufo, Shrideep Pallickara, and Sangmi Lee Pallickara. DeepSalt: Bridging Laboratory and Satellite Spectra through Domain Adaptation and Knowledge Distillation for Large-Scale Soil Salinity Estimation. (2025).

[21] Rupasree Dey, Abdul Matin, Nathan Orwick, Yao Zhang, Shrideep Pallickara, and Sangmi Lee Pallickara. When to Trust, How to Distill: Multi-Foundation Model Guidance for Lightweight, Robust Scientific Time Series Forecasting. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD)*. ACM, Jeju, South Korea. To appear.

[22] Howard J Diamond, Thomas R Karl, Michael A Palecki, C Bruce Baker, Jesse E Bell, Ronald D Leeper, David R Easterling, Jay H Lawrimore, Tilden P Meyers, Michael R Helfert, et al. US Climate Reference Network after one decade of operations: Status and assessment. *Bulletin of the American Meteorological Society* 94, 4 (2013), 485–498.

[23] Dara Entekhabi, Eni G Njoku, Peggy E O'neill, Kent H Kellogg, Wade T Crow, Wendy N Edelstein, Jared K Entin, Shawn D Goodman, Thomas J Jackson, Joel Johnson, et al. The soil moisture active passive (SMAP) mission. *Proc. IEEE* 98, 5 (2010), 704–716.

[24] Jianling Fan, Brian McConkey, Hong Wang, and Henry Janzen. Root distribution by depth for temperate agricultural crops. *Field Crops Research* 189 (2016), 68–74.

[25] Kuai Fang, Ming Pan, and Chaopeng Shen. The value of SMAP for long-term soil moisture estimation with the help of deep learning. *IEEE Transactions on Geoscience and Remote Sensing* 57, 4 (2018), 2221–2233.

[26] Tanjim Bin Faruk, Abdul Matin, Shrideep Pallickara, and Sangmi Lee Pallickara. Accounting for spatial variability with the histogram of oriented gradients based masking improves performance of masked autoencoder over hyperspectral satellite imagery (student abstract). In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 29365–29367.

[27] Xiangyu Ge, Jingzhe Wang, Jianli Ding, Xiaoyi Cao, Zipeng Zhang, Jie Liu, and Xiaohang Li. Combining UAV-based hyperspectral imagery and machine learning algorithms for soil moisture content monitoring. *PeerJ* 7 (2019), e6926.

[28] Mohammad Reza Hajizadeh Javaran, Mohammad Mahdi Rajabi, Nima Kamali, Marwan Fahs, and Benjamin Belfort. Encoder–Decoder Convolutional Neural Networks for Flow Modeling in Unsaturated Porous Media: Forward and Inverse Approaches. *Water* 15, 16 (2023), 2890.

[29] Dean P Holzworth, Neil I Huth, Peter G deVoil, Eric J Zurcher, Neville I Herrmann, Greg McLean, Karine Chenu, Erik J van Oosterom, Val Snow, Chris Murphy, et al. APSIM–evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software* 62 (2014), 327–350.

[30] Collin H Homer, Joyce A Fry, Christopher A Barnes, et al. The national land cover database. *US geological survey fact sheet* 3020, 4 (2012), 1–4.

[31] Illinois Climate Network. Water and Atmospheric Resources Monitoring Program. Illinois State Water Survey.

[32] Ashok Kumar Indoria, Kishori Lal Sharma, and Kotha Sammi Reddy. Hydraulic properties of soil under warming climate. In *Climate change and soil interactions*. Elsevier, Amsterdam, 473–508.

[33] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Honolulu, HI, 1125–1134.

[34] Noureddine Jarray, Ali Ben Abbes, Manel Rhif, Hanen Dhaou, Mohamed Ouessar, and Imed Riadh Farah. SMETool: A web-based tool for soil moisture estimation based on Eo-Learn framework and Machine Learning methods. *Environmental Modelling & Software* 157 (2022), 105505.

[35] Anuj Karpatne, Gowtham Atluri, James H Faghmous, Michael Steinbach, Arindam Banerjee, Auroop Ganguly, Shashi Shekhar, Nagiza Samatova, and Vipin Kumar. Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on knowledge and data engineering (TKDE)* 29, 10 (2017), 2318–2331.

[36] Anuj Karpatne, Imme Ebert-Uphoff, Sai Ravela, Hassan Ali Babaie, and Vipin Kumar. Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering (TKDE)* 31, 8 (2018), 1544–1554.

[37] Paahuni Khandelwal, Samuel Armstrong, Abdul Matin, Shrideep Pallickara, and Sangmi Lee Pallickara. Cloudnet: A deep learning approach for mitigating occlusions in landsat-8 imagery using data coalescence. In *2022 IEEE 18th International Conference on e-Science (e-Science)*. IEEE, 117–127.

[38] Paahuni Khandelwal, Andrei Bachinin, Tanjim Bin Faruk, Everett Lewark, Sangmi Lee Pallickara, and Shrideep Pallickara. DeepSoil: Physics-Guided Learning for High-Resolution Soil Moisture Mapping from Surface to Root-Zone via Multimodal Sensing and Process-Based Modeling. *ACM Transactions on Spatial Algorithms and Systems* (2026).

[39] Paahuni Khandelwal, Jeffrey D Niemann, David J Mulla, Shrideep Pallickara, and Sangmi Lee Pallickara. Subterra: Estimating soil moisture at root zone depths using science-guided learning. In *2025 IEEE Conference on Artificial Intelligence (CAI)*. IEEE, 328–335.

[40] Paahuni Khandelwal, Sangmi Lee Pallickara, and Shrideep Pallickara. Deepsoil: A science-guided framework for generating high precision soil moisture maps by reconciling measurement profiles across in-situ and remote sensing data. In *Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems*. 233–246.

[41] Hyunglok Kim, Sangchul Lee, Michael H Cosh, Venkataraman Lakshmi, Yonghwan Kwon, and Gregory W McCarty. Assessment and combination of SMAP and Sentinel-1A/B-derived soil moisture estimates with land surface model outputs in the Mid-Atlantic Coastal Plain, USA. *IEEE Transactions on Geoscience and Remote Sensing* 59, 2 (2020), 991–1011.

[42] Seung-bum Kim, Jakob van Zyl, Scott Dunbar, Eni Njoku, Joel Johnson, Mahta Moghaddam, Jiancheng Shi, and Leung Tsang. *Algorithm Theoretical Basis Document SMAP L2 & L3 Radar Soil Moisture (Active) Data Products*. Technical Report. Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA.

[43] Markus Kottek, Jürgen Grieser, Christoph Beck, Bruno Rudolf, and Franz Rubel. World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift* 15, 3 (2006), 259–263.

[44] David R Legates, Daniel J Leathers, Tracy L DeLiberty, Geoff E Quelch, Kevin Brinson, Jason Butke, Rezaul Mahmood, and Stuart A Foster. DEOS: The Delaware Environmental Observing System. In *21st International Conference on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*. American Meteorological Society, American Meteorological Society, San Diego, CA, 8 pages.

[45] Pei Leng, Zhao-Liang Li, Qian-Yu Liao, Yun-Jing Geng, Qiu-Yu Yan, Xia Zhang, and Guo-Fei Shang. Enhanced Surface Soil Moisture Retrieval at High Spatial Resolution From the Integration of Satellite Observations and Soil Pedotransfer Functions. *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), 1–11.

[46] Qingliang Li, Yuheng Zhu, Wei Shangguan, Xuezhi Wang, Lu Li, and Fanhua Yu. An attention-aware LSTM model for soil moisture and soil temperature prediction. *Geoderma* 409 (2022), 115651.

[47] Licheng Liu, Wang Zhou, Kaiyu Guan, Bin Peng, Shaoming Xu, Jinyun Tang, Qing Zhu, Jessica Till, Xiaowei Jia, Chongya Jiang, et al. Knowledge-guided machine learning can improve carbon cycle quantification in agroecosystems. *Nature communications* 15, 1 (2024), 357.

[48] Rezaul Mahmood, Megan Schargorodski, Stuart Foster, and Andrew Quilligan. A technical overview of the Kentucky Mesonet. *Journal of Atmospheric and Oceanic Technology* 36, 9 (2019), 1753–1771.

[49] Ashray Manepalli, Adrian Albert, Alan Rhoades, Daniel Feldman, and Andrew D Jones. Emulating numeric hydroclimate models with physics-informed cGANs. In *AGU Fall Meeting Abstracts*. American Geophysical Union, San Francisco, CA, 1 pages.

[50] Jeffrey G Masek, Eric F Vermote, Nazmi E Saleous, Robert Wolfe, Forrest G Hall, Karl Fred Huemmrich, Feng Gao, Jonathan Kutler, and Teng-Kui Lim. A Landsat surface reflectance dataset for North America, 1990-2000. *IEEE Geoscience and Remote sensing letters* 3, 1 (2006), 68–72.

[51] Abdul Matin, Samuel Armstrong, Saptashwa Mitra, Shrideep Pallickara, and Sangmi Lee Pallickara. Rapid betweenness centrality estimates for transportation networks using capsule networks. In *2022 Fourth International Conference on Transdisciplinary AI (TransAI)*. IEEE, 89–96.

[52] Abdul Matin, Rupasree Dey, Tanjim Bin Faruk, Shrideep Pallickara, and Sangmi Lee Pallickara. Knowledge-Guided Masked Autoencoder with Linear Spectral Mixing and Spectral-Angle–Aware Reconstruction. (2026).

[53] Reed M Maxwell and Norman L Miller. Development of a coupled land surface and groundwater model. *Journal of Hydrometeorology* 6, 3 (2005), 233–247.

[54] Renee A McPherson, Christopher A Fiebrich, Kenneth C Crawford, James R Kilby, David L Grimsley, Janet E Martinez, Jeffrey B Basara, Bradley G Illston, Dale A Morris, Kevin A Kloesel, et al. Statewide monitoring of the mesoscale environment: A technical update on the Oklahoma Mesonet. *Journal of Atmospheric and Oceanic Technology* 24, 3 (2007), 301–321.

[55] Nuno Cirne Mira, João Catalão, Giovanni Nico, and Pedro Mateus. Soil moisture estimation using atmospherically corrected C-band InSAR data. *IEEE Transactions on Geoscience and Remote Sensing* 60 (2021), 1–9.

[56] Saptashwa Mitra, Daniel Rammer, Shrideep Pallickara, and Sangmi Lee Pallickara. Glance: A generative approach to interactive visualization of voluminous satellite imagery. In *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, 359–367.

[57] M Moghaddam, A Silva, D Clewley, R Akbar, SA Hussaini, J Whitcomb, R Devarakonda, R Shrestha, RB Cook, G Prakash, et al. Soil moisture profiles and temperature data from SoilSCAPE sites, USA. ORNL DAAC. https://doi.org/10.3334/ORNLDAAC/1339

[58] Ranga Myneni, Yuri Knyazikhin, and Taejin Park. MOD15A2H MODIS/Terra Leaf Area Index/FPAR 8-Day L4 Global 500m SIN Grid V006. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD15A2H.006

[59] National Soil Moisture Network. *National Soil Moisture Network*. National Soil Moisture Network. http://nationalsoilmoisture.com/About.html

[60] Viliam Novák and Hana Hlaváčiková. Soil-water retention curve. In *Applied Soil Hydrology*. Springer, Cham, 77–96.

[61] Tyson E Ochsner, Evan Linde, Matthew Haffner, and Jingnuo Dong. Mesoscale soil moisture patterns revealed using a sparse in situ network and regression kriging. *Water Resources Research* 55, 6 (2019), 4785–4800.

[62] Peggy O'Neill, Steven Chan, Eni Njoku, Tom Jackson, and Rajat Bindlish. *Soil Moisture Active Passive (SMAP) Algorithm Theoretical Basis Document Level 2 & 3 Soil Moisture (Passive) Data Products.* Technical Report. Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA.

[63] OpenStreetMap Wiki. *QuadTiles.* OpenStreetMap Foundation. https://wiki.openstreetmap.org/wiki/QuadTiles

[64] Xiang Peng, Dan Hu, Wenzhi Zeng, Jingwei Wu, and Jiesheng Huang. Estimating soil moisture from hyperspectra in saline soil based on EPO-PLS regression. *Transactions of the Chinese Society of Agricultural Engineering* 32, 11 (2016), 167–173.

[65] Giuseppe Provenzano. Using HYDRUS-2D simulation model to evaluate wetted soil volume in subsurface drip irrigation systems. *Journal of Irrigation and Drainage Engineering* 133, 4 (2007), 342–349.

[66] Luca Pulvirenti, Giuseppe Squicciarino, Luca Cenci, Giorgio Boni, Nazzareno Pierdicca, Marco Chini, Cosimo Versace, and Paolo Campanella. A surface soil moisture mapping service at national (Italian) scale based on Sentinel-1 data. *Environmental Modelling & Software* 102 (2018), 13–28.

[67] Lorenzo Adolph Richards. Capillary conduction of liquids through porous mediums. *physics* 1, 5 (1931), 318–333.

[68] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015.* Springer, Springer, Cham, 234–241.

[69] Garry L Schaefer and Ron F Paetzold. SNOTEL (SNOwpack TELemetry) and SCAN (soil climate analysis network). *Automated Weather Stations for Applications in Agriculture and Water Resources Management: Current Use and Future Perspectives* 1074 (2001), 187–194.

[70] Volkan Senyurek, Fangni Lei, Dylan Boyd, Mehmet Kurum, Ali Cafer Gurbuz, and Robert Moorhead. Machine learning-based CYGNSS soil moisture estimates over ISMN sites in CONUS. *Remote Sensing* 12, 7 (2020), 1168.

[71] Martha Shulski, Stonie Cooper, Glen Roebke, and Al Dutcher. The Nebraska Mesonet: Technical overview of an automated state weather network. *Journal of Atmospheric and Oceanic Technology* 35, 11 (2018), 2189–2200.

[72] J Šimůnek, M Šejna, and M Th Van Genuchten. *HYDRUS-1D: Simulating the One-Dimensional Movement of Water, Heat, and Multiple Solutes in Variably-Saturated Media.* US Salinity Laboratory, USDA-ARS, Riverside, CA.

[73] SMAP Science Team. *SMAP L2 & L3 Radar/Radiometer Soil Moisture (Active/Passive) Data Products.* Technical Report. Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA.

[74] Soil Survey Staff. Gridded National Soil Survey Geographic (gNATSGO) Database for the Conterminous United States. USDA-NRCS. https://nrcs.app.box.com/v/soils

[75] US Geologic Survey. 1 Arc-second Digital Elevation Models (DEMs)–USGS National Map 3DEP Downloadable Data Collection.

[76] Texas Water Development Board. *TexMesonet.* Texas Water Development Board. https://www.texmesonet.org/

[77] M Th Van Genuchten. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil science society of America journal* 44, 5 (1980), 892–898.

[78] Thomas Vandal, Evan Kodra, Jennifer Dy, Sangram Ganguly, Ramakrishna Nemani, and Auroop R Ganguly. Quantifying uncertainty in discrete-continuous and skewed data with Bayesian deep learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* ACM, London, UK, 2377–2386.

[79] Noemi Vergopolan, Nathaniel W Chaney, Ming Pan, Justin Sheffield, Hylke E Beck, Craig R Ferguson, Laura Torres-Rojas, Sara Sadri, and Eric F Wood. SMAP-HydroBlocks, a 30-m satellite-based soil moisture dataset for the conterminous US. *Scientific data* 8, 1 (2021), 264.

[80] Robert J Zamora, F Martin Ralph, Edward Clark, and Timothy Schneider. The NOAA Hydrometeorology Testbed soil moisture observing networks: Design, instrumentation, and preliminary results. *Journal of Atmospheric and Oceanic Technology* 28, 9 (2011), 1129–1140.

[81] Yufang Zhang, Shunlin Liang, Zhiliang Zhu, Han Ma, and Tao He. Soil moisture content retrieval from Landsat 8 data using ensemble learning. *ISPRS Journal of Photogrammetry and Remote Sensing* 185 (2022), 32–47.

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

[83] Jingyao Zheng, Tianjie Zhao, Haishen Lü, Jiancheng Shi, Michael H Cosh, Dabin Ji, Lingmei Jiang, Qian Cui, Hui Lu, Kun Yang, et al. Assessment of 24 soil moisture datasets using a new in situ network in the Shandian River Basin of China. *Remote Sensing of Environment* 271 (2022), 112891.

[84] Yinhao Zhu, Nicholas Zabaras, Phaedon-Stelios Koutsourelakis, and Paris Perdikaris. Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. *J. Comput. Phys.* 394 (2019), 56–81.