Volume 70, Issue 3, 2026
Journal of Scientific Research
of
The Banaras Hindu University

Machine Learning-Based Downscaling of SMAP Soil Moisture and Spatiotemporal Trend Analysis (2016–2023), Middle Ganga Plain

Ratnakar Mishra*, Kuldeep Prakash

Department of Geology, Banaras Hindu University
*Corresponding Author: mishraratnakar2025@gmail.com

Abstract: **Accurate quantification of high-resolution soil moisture (SM) is essential for hydrological modelling; however, the 36-kilometre footprint of NASA's SMAP mission conceals micro-scale environmental heterogeneity. This study introduces an analytical framework that integrates machine-learning-based spatial disaggregation with non-parametric spatiotemporal trend analysis for the period 2016–2023.**

Index Terms: **Soil Moisture Active Passive (SMAP), Spatial Disaggregation, Artificial Neural Network (ANN), Spatiotemporal Trend Analysis, Mann-Kendall Trend Test.**

Introduction

Initially, an Artificial Neural Network (ANN) was employed to downscale SMAP data to a 1-kilometre grid, thereby revealing pronounced contrasts among localised extremes and effectively capturing spatial heterogeneity that is otherwise obscured within the standard 36-kilometre SMAP satellite footprint. Chronologically ordered panels demonstrate distinct intra-annual cycles of wetting and drying across the geographic domain. The high-resolution outputs capture localised, pixel-level variations driven by high-resolution covariates such as Evapotranspiration (ET) and Land Surface Temperature (LST), indicating that the neural network effectively learned and applied the non-linear environmental relationships governing localised soil water retention.

Subsequently, a pixel-by-pixel Mann-Kendall trend test was applied to the high-resolution anomalies. Unmasked trend data can create an "illusion of uniformity," suggesting basin-wide wetting. However, the application of strict statistical masking filters out meteorological noise and isolates definitive geographic "hotspots" of permanent moisture accumulation. This framework effectively separates chaotic seasonal weather cycles from enduring eco-hydrological shifts.

Soil moisture is a land surface state variable that links the land surface and the atmosphere and is a key variable for hydrological and climatological applications (Babaeian et al., 2019). The partitioning of incoming solar radiation into latent and sensible heat fluxes over land is strongly influenced by soil moisture, which, in turn, affects the global water, energy, and carbon cycles (Montzka et al., 2026). Soil moisture also directly influences the thermodynamic conditions of the boundary layer and, in turn, cloud formation, precipitation, and the initiation of extreme convective weather events (References therein). Furthermore, vegetation transpiration, as a mechanism for transferring materials and energy from the pedosphere to the atmosphere, was identified by Xie et al. (2025) as an important process closely related to changes in root-zone soil water content (SWC).

For many applications, such as the optimisation of water resource management for agriculture and the prediction of extreme climatic events like drought and flash floods, the spatial distribution of soil moisture and long-term temporal trends are important. The spatial distribution of soil moisture and the long-term temporal trend are significant for many applications, including the optimisation of water resource management for agriculture, the initialisation of soil moisture states for model runs, and the prediction of extreme climatic events such as droughts and flash floods (Kwon et al., 2026). In arid and semi-arid regions, where water is the controlling factor in vegetation growth and development, soil moisture is the primary water source for plants and the most sensitive indicator of agricultural vulnerability (Cao et al., 2026). Therefore, long-term, high-temporal-resolution soil moisture monitoring over land surfaces is a scientific need to support sustainable water resource management in a changing global climate and to address its effects on precipitation redistribution over land (Babaeian et al., 2019).

Traditionally, soil moisture estimation has relied on in situ measurements obtained by gravimetric sampling, time-domain reflectometry (TDR), and capacitance probes (Babaeian et al., 2019). The methodologies used on the ground are very precise and

*Mishraratnakar2025@gmail.com

DOI: 10.37398/JSR.2026.700301

reliable for estimating volumetric water content at a specific spatial location in the soil profile, but have significant spatial limitations (Meyer et al., 2022). Overall, in situ networks are typically sparse, span diverse terrain, and are labour-intensive and costly (Babaeian et al., 2019). The newer technologies, such as cosmic-ray neutron sensors (CRNS), also have a limited measurement area, typically several hundred meters, and are sensitive to hydrogen sources near the trees, such as root biomass (Babaeian et al., 2019) and atmospheric water vapour. The spatial variability of soil moisture, driven by precipitation, topography, soil texture, and vegetation cover, makes it impossible to assume that soil moisture is representative of the hydrological conditions of the large regional catchments from which it is derived (Korres et al., 2010).

However, the limitations of the ground-based observational network in providing continuous, space-time measurements of soil moisture at a global scale have led satellite remote sensing to become a game-changing technology (Montzka et al., 2026). Microwave remote sensing is also the most promising way to globally retrieve soil moisture, and at these wavelengths, clouds and moderate vegetation canopies can be penetrated. In particular, the L-band microwave frequency (around 1.4 GHz) is highly sensitive to the dielectric properties of liquid water in the top 5 cm of the soil profile and is expected to provide the most accurate volumetric surface-soil moisture estimates for scientific applications to date (Babaeian et al., 2019). This important physical recognition prompted the establishment of dedicated L-band satellite missions, such as the National Aeronautics and Space Administration's (NASA) Soil Moisture Active Passive (SMAP) mission launched in 2015 (Entekhabi et al., 2014).

NASA's SMAP mission's primary task was to generate the first global, high-accuracy soil moisture dataset, which would have a considerable impact on climate modelling, drought monitoring, and hydrological modelling (Entekhabi et al., 2014). Originally formulated to accomplish this objective, SMAP was to include two instruments: an L-band passive radiometer and an L-band active synthetic aperture radar (SAR). For this purpose, SMAP was designed innovatively with two instruments: a L-band passive radiometer and an L-band active synthetic aperture radar (SAR). With a coarse footprint of ~36-40 km, the passive radiometer achieved high radiometric accuracy, while the active radar had a fine footprint of ~3 km. The mission's architecture aimed to synergistically combine the two, resulting in an intermediate soil moisture product with a fine spatial resolution of 9 km (Das et al., 2018).

Unfortunately, the active radar on SMAP experienced a hardware anomaly, and within a couple of months after launch, it was converted to a passive L-band radiometer (Das et al., 2019). The high spatial resolution (36 km) of the soil moisture data retrieved from the radiometer is lower than the native spatial resolution (MIM-ODIN), posing an important challenge to be addressed during operations.To achieve a sufficient signal-to-noise ratio, passive radiometers must integrate natural thermal emissions over large geographic areas, resulting in data products that are averaged over local environmental extremes (Zhong et al., 2024).

Global- and large-scale climate models require a spatial resolution of 36 km; however, this is insufficient for regional and local climate models (Sabaghy et al., 2018). Other local-scale applications require soil moisture intelligence at local scales (less than 1 km), namely drought monitoring and irrigation management, and detailed flood forecasting at a catchment scale (up to 1 km). Other examples of applications where soil moisture intelligence is required at local scales include precision agriculture, detailed drought monitoring, and irrigation management (Sabaghy et al., 2018). It is crucial to have a robust spatial downscaling methodology (Peng et al., 2017), especially because the number of coarse SMAP radiometers is significantly smaller than the number of high-resolution applications of the data within local hydrology DTSS.

Downscaling of coarse resolution satellite SM data to fine scale resolution is another traditional approach of downscaling that is possible using the Spatial downscaling (also referred to as Spatial disaggregation) theoretical framework, and it can also incorporate the auxiliary variables with fine scale resolution of the environment that can explain the spatial variability in the landscape. A number of theoretical approaches have been put forward over the last decade, all of which are extremely dependent on optical and infrared TOA observations, most notably land surface temperature (LST), and vegetation indices such as the Normalised Difference Vegetation Index (NDVI) (Kumar et al., 2025).

One of the most well-known theoretical methods employed in the remote sensing community is the Universal Triangle (or Trapezoid) method (Kumar et al., 2025). This involves them plotting a scatterplot of LST against a vegetation index (VI), from which they can determine the soil's evaporative efficiency and infer the distribution of soil moisture at a given location (Kumar et al., 2025). The traditional versions of the Triangle Method, however, do not always account for the complexity of vegetation modulation and have suggested adding the Modified Soil Adjusted Vegetation Index (MSAVI) to make the method more sensitive in heterogeneous landscapes (Kumar et al., 2025). The reduction in the effectiveness of optical and thermal sensors is significant due to cloud contamination of optical sensors, and the assumption of a linear relationship among soil moisture, temperature, and vegetation is often observed to be highly non-linear in the field (Chakrabarti et al., 2015).

Other approaches have used geomorphometric variables such as topographic slope, elevation, and the Topographic Wetness Index (TWI) as the primary controls for spatial downscaling (Senanayake et al., 2024). Guevara and Vargas (2019) demonstrated that using terrain parameters in local modelling may enhance the ability to resolve differences, simulate soil moisture without ecological or climatic data, and capture the mechanics of lateral subsurface water flow. However, terrestrial models cannot capture natural vegetation and farm-crop water extraction, and thus may exhibit temporal inaccuracies during peak crop water extraction (Senanayake et al., 2024).

Theoretical advantages of strictly theoretical models and the physical constraints of active-passive fusion architectures have, to date, inspired the use of Machine Learning (ML) methods in Geospatial Downscaling (GD), with these methods becoming an extremely popular solution (Senanayake et al., 2024). The spatiotemporal dynamics of soil moisture–environmental driving factors are highly complex, non-linear, and dynamic, and can be captured by purely data-driven machine learning algorithms (Montzka et al., 2026). Unlike empirical or physical models, machine learning does not require strict assumptions about the physical environment, enabling extremely flexible adaptation across different hydro-climatic regions and robustness to noisy data inputs (Senanayake et al., 2024).

The fundamental method of the ML soil moisture downscaling is to build a regression model that can be applied to produce soil moisture from the coarse resolution (e.g., SMAP) input with high-resolution auxiliary covariates that are also aggregated to the same resolution as the coarse resolution (e.g., SMAP) input (Abbaszadeh et al., 2019). In general, the predictive soil covariates selected are related to variations in soil moisture caused by meteorological forcing (such as evapotranspiration), vegetation activity (such as NDVI or EVI), thermal conditions (daytime and nighttime LST), and static soil properties (such as sand and clay fractions) (Abbaszadeh et al., 2019). After successfully mapping the complex relationships among the aggregated covariates and coarse soil moisture, the trained model is then fed with high-resolution covariates to accurately map soil moisture at a fine scale of 1 km (Ma et al., 2023).

Today's literature is replete with a broad array of highly developed ML and Deep Learning (DL) algorithms that have been rigorously tested. Random Forest (RF) and eXtreme Gradient Boosting (XGBoost) are widely used ensemble decision-tree models that deliver outstanding performance on high-dimensional tabular data and have minimal risk of overfitting (Senanayake et al., 2024). On the Tibetan Plateau, Ma et al. (2023) demonstrated the effectiveness of the proposed approach within a Random Forest framework, which overcomes the limitations of existing native satellite maps, especially during snow cover, to achieve quantitative retrievals at scales much larger than those of current native satellite maps. Likewise, Abbaszadeh et al. (2019) showed that an ensemble learning method can be used to effectively downscale SMAP radiometer data over the Continental United States (CONUS), and its derived unbiased Root Mean Square Error (ubRMSE) metrics would be better compared to traditional uniform disaggregation methods (Abbaszadeh et al., 2019).

Advanced DBNs (Deep Belief Networks) have also been shown to be more effective at learning high-level, abstract relationships in the environment than ANNs (Artificial Neural Networks). However, Deep Belief Networks (DBNs) have been shown to be a powerful approach for extracting nonlinear features from multiple covariate sources, thereby overcoming the limitations of shallow machine learning methods, as demonstrated by Cai et al. (2022). With these efficient computation architectures and powerful scaling methods, scientists can effectively tackle this important spatial challenge and realise SMAP's promise of providing actionable agricultural intelligence at 1 km resolution (Zhong et al., 2024).

The synthesis of a high-resolution soil moisture product at 1 km resolution is critical for enabling highly localized long-term environmental and climatological assessment. This is not an end goal, but rather, a prerequisite soil moisture data cube needed to conduct the spatiotemporal trend analysis (Zeng et al., 2025). After producing continuous, spatially disaggregated soil moisture maps, the next important step in the development of hydrological research is to perform comprehensive spatiotemporal trend analysis to uncover the effects of global climate change (Kim et al., 2023).

The magnitude of climate variations' influence on soil moisture is highly dependent on strong human interventions, which are sometimes coupled with climate variations (Zeng et al., 2025). Due to variations in precipitation anomalies and rising global temperatures, many subtropical, arid, and semi-arid regions have undergone substantial changes in water balance, resulting in abrupt regional shifts and more frequent droughts (Cao et al., 2026). Intense soil moisture depletion has occurred in some regions, while unexpected increases in soil moisture have been observed in parts of Northwest China (Wang et al., 2023).

In the past few decades, researchers have adopted robust and nonparametric statistical techniques to quantify these complex changes in space and time, particularly the Mann-Kendall (MK) trend test and Sen's slope estimator (Zeng et al., 2025). It is a very robust test for high-resolution data, especially for remote-sensing-based data, due to its sensitivity to missing values, seasonal discontinuities, and extreme outliers in multi-year hydrological time-series data. The results of the Mann-Kendall test can be processed algorithmically on a per-pixel basis to generate detailed maps of soil moisture trends, which can be used to understand the localisation of soil moisture drying or the signal of rapid soil moisture changes in an ecosystem (Kim et al., 2023).

Moreover, analysing trends and applying advanced geospatial detection models can help identify the specific causes of drought and soil moisture variations. Zeng et al. (2025) used an Optimal Parameter Geographic Detector to detect non-linear localised soil drought in Hebei province,where both land cover (NDVI) and topographic altitude have been shown to affect soil drought, suggesting that the human-induced land cover changes may offset the pure meteorological drought for agricultural plains (Zeng et al., 2025). Likewise, Yin et al. (2023) applied OPGD to show that land surface temperature and elevation are alternative primary drivers of soil moisture as a function of season and soil moisture baseline (Yin et al., 2023).

One of the most important applications of high-resolution spatiotemporal trend analysis (Kim et al., 2023) is to uncouple and examine the complex climate stress-induced dynamics among soil moisture, groundwater, and vegetation. Kim et al. (2023) found a very strange phenomenon in the Limpopo River Basin where ecological and hydrological drought can end up being fundamentally uncoupled (Kim et al., 2023). Using Mann-Kendall statistics, they calculated changes in downscaled anomalies over 18 years and found no significant long-term negative trends in vegetation indices (NDVI), although a statistically significant loss of underlying groundwater storage was observed (Kim et al., 2023). Such decoupling suggests that short-term soil moisture changes in shallow soil layers could compensate for short-term changes in vegetation growth, providing a buffer to the long-term hydrologic depletion of the deeper aquifer storage by agricultural managers (Kim et al., 2023)

This increase in greening, however, did not occur across the entire Irtysh basin; Cao et al. (2026) did, however, find a positive vegetation greening trend throughout the basin, despite localised warming and humidification (Cao et al., 2026). They demonstrated that the growing season was lengthened and the soil moisture in summer was moistened by higher rainfall because of the increase in temperature, which offset the increase in transpiration, reducing drought stress and allowing photosynthesis to keep up with the increase in transpiration that would occur without the moisture from the summer rain (Cao et al., 2026). However, Xie et al. (2025) found that excessive greenery led to negative hydrological feedback in the subtropical, humid regions of Fujian Province. By performing simulations with a global land data assimilation system (GLDAS), Xie et al. (2025) supported the well-argued assertion that during the summer months, much of this loss is due to intense vegetation transpiration, thereby producing a natural subsurface water deficit generated by the plants themselves.

The occurrence of these cascading, complex eco-hydrological feedbacks clearly demonstrates the absolute necessity of integrating high-resolution machine-learning downscaling with careful spatiotemporal statistical tracking. The spatial averaging of native satellite footprints (Zhong et al., 2024) filters out complex interactions among the various crop types, local topographic drainage, and atmospheric heat fluxes.

While advances in machine-learning-based downscaling and in regional trend analysis were quite remarkable, there remains a strong need to integrate these two approaches into a single analytical framework from a unified end-to-end perspective.

Many previous studies have successfully downscaled coarse soil moisture data for very short test durations, or the models used are unable to filter out a patch of statistically insignificant environmental noise from the time series. However, coarse reanalysis data (e.g., ERA5) or native 36 km satellite observations can be used for long-term trend analysis and will not obscure the important, highly localised mechanisms of the agricultural drought and areas of water scarcity.

The first objective is to build, train, and optimise a robust Artificial Neural Network (ANN) to downscale NASA coarse (native ~10 km) soil moisture radiometer data to a fine (1 km) spatial grid. Static USGS topographic elevation parameters and intensive use of high-resolution multispectral optical and thermal environmental covariates, such as MODIS-derived Normalised Difference Vegetation Index (NDVI), Evapotranspiration (ET), Enhanced Vegetation Index (EVI) and Land Surface Temperature (LST). The network architecture will be optimised to minimise overfitting and redundancy by using a strict permutation feature-importance evaluation to mathematically isolate the most dominant drivers of soil moisture variability, thereby providing an optimal, concise and accurate model.

Second, the fully developed spatiotemporal trend analysis over an extended 8-year study period (2016–2024) using the new, continuous soil moisture dataset at 1 km resolution. In this study, using the pixel-by-pixel nonparametric Mann-Kendall trend test, geographically detailed maps will be generated to assess statistically significant changes in soil wetting and drying in the region of interest.

The primary objective of this study is to demonstrate that an integrative approach, combining advanced Artificial Intelligence for spatial enhancement with holistic non-parametric statistical analysis for continuous-time tracking, can provide highly localised information on changing hydrological regimes. This framework delivers 1-kilometre actionable intelligence, serving as a valuable resource for environmental monitoring, drought-resistant agricultural planning, and strategic climate change adaptation.

2. Methodology

Cloud-based Geospatial Computing and Advanced Machine Learning Methodologies were a major focus of the computational framework for this research. It was done using the Google Earth Engine (GEE) Python API, starting with the initial data acquisition and geospatial processing. The environment was set up and authenticated to access the GEE high-volume API endpoint, enabling processing of large datasets of satellite imagery. The Region of Interest (ROI), also called the spatial domain, was defined interactively using the geemap library, which was used to extract a geometry representing the spatial extent of the data to be filtered and clipped in all subsequent processing steps.To perform a comprehensive spatiotemporal analysis, it was decided to include an 8-year study period from January 2016 to January 2024. The native temporal resolutions of the different satellite products varied, necessitating a coordinated temporal frame. To do so, a custom temporal sequence generator was programmed to compute the number of months between the start and end dates. This sequence was then mapped using an advanced lambda function, which generated a precise, iterative sequence of monthly start dates (time_list) as the temporal backbone for harmonising monthly composites across all distinct satellite sensors. The predictive modelling was based on combining multiple remote sensing variables through multi-sensor data acquisition and aggregation. Each image collection was processed by a custom aggregation function (monthly) that removed all time steps from the collection that were outside of one month, and then computed the mean value of each pixel over the one-month period, producing monthly composites that were normalised. The dependent variable for the downscaling model was soil moisture data at a coarse spatial resolution (~10 km) from the NASA Soil Moisture Active Passive (SMAP) mission (specifically, the SMAP Soil Moisture Active Passive (SMAP) SPL3SMP_E/005 product). For programmatic convenience, the soil_moisture_am band was chosen and renamed. Soil water retention is strongly influenced by the vegetation dynamics and can be represented by a vegetation covariate. Hence, Normalised Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) data were obtained from the MODIS MOD13Q1 version 061 data set and then averaged monthly. Thermal and Hydrological Covariates: The MODIS MOD11A2 dataset provided diurnal Land Surface Temperature (LST) measurements (both daytime and nighttime). Furthermore, total Evapotranspiration (ET) data were merged with those of the MODIS MOD16A2GF product. Topography and land cover are categorised as static topographic variables, rather than dynamic variables, because their extent of change over the study period is relatively small. The USGS GTOPO30 digital elevation model was used to incorporate elevation data. The categorical land cover data were obtained from the MODIS MCD12Q1 product; however, rather than the temporal mean, the statistical mode was used over the time series to classify the main LC_Type1 for each pixel. All continuous variables, along with the static topographical and categorical bands, were consolidated into a single multi-band Earth Engine ImageCollection and clipped to the ROI.

The xee and xarray libraries were used to establish a critical link between the Google Earth Engine cloud environment and local machine learning. Precise geographical coordinates of the study area were extracted and converted into a local Shapely geometry to obtain precise parameters for the grid mapping.

The methodology used was a two-stage resolution. For training the neural network, the multi-band SWIR dataset was first downloaded from GEE at a coarse spatial resolution equal to the native scale of the SMAP target variable. This was accomplished by using a grid_scale of 0.1 degrees (approximately 10km) in the EPSG:4326 coordinate reference system. For the actual downscaling and inference phase, a high-resolution grid scale of 0.01 degrees (approximately 1 km) was achieved by adjusting the grid_scale parameter in the model. Both data sets were reordered in time and cast into the multi-dimensional xarray data type.

The 10 km spatial grid has been resampled into a 2d Pandas DataFrame for traditional machine learning operations. Data integrity was maintained by removing rows from the training set with missing numeric values (NaNs). The data was quite varied and needed to be manipulated for the analysis (this is called 'feature engineering'). A neural network cannot be applied directly to the categorical land-cover variable (LC_Type1). Thus, it was processed into a sparse numerical representation using one-hot encoding (pd.get_dummies), which creates separate binary columns (prefixed with 'lc').

The continuous predictor variables (NDVI, EVI, ET, daytime LST, nighttime LST, and elevation) were isolated and standardised to optimise the neural network's gradient descent convergence. The data was standardised using a StandardScaler from the scikit-learn library, which centres each feature at 0 and rescales it to have unit variance. A similar independent standardisation was performed on the target variable, coarse soil moisture. These discretisation codes were then horizontally merged with the scaled continuous attributes, producing the final high-dimensional input matrix with NumPy. This led to the creation of the final high-dimensional input matrix by horizontally concatenating the discretisation codes with the scaled continuous attributes, implemented in NumPy. The core of the downscaling methodology was an Artificial Neural Network (ANN) built using the Sequential API in TensorFlow/Keras. The preprocessed 10 km dataset was split (random_state=42) into 80% for training and 20% for validation and testing, with the split independent of the training data. The original MLP architecture used a decreasing number of nodes to induce the network to learn a compressed, generalised representation of environmental covariates. The network accepted the input matrix through three successive Dense hidden layers, each with 64, 32, and 16 neurons. The non-linear relationships in the environment were modelled using the Rectified Linear Unit (ReLU) activation function in each hidden layer. The output layer comprised one neuron with a linear activation function to predict the standardised continuous soil moisture value.

Model compilation is done with the Adaptive Moment Estimation (Adam) optimiser. The primary loss function for parameter updates was Mean Squared Error (MSE), with Mean Absolute Error (MAE) monitored as a secondary metric. This network is trained for more than 100 epochs with a batch size of 8 and a 20% internal validation set to monitor for overfitting.

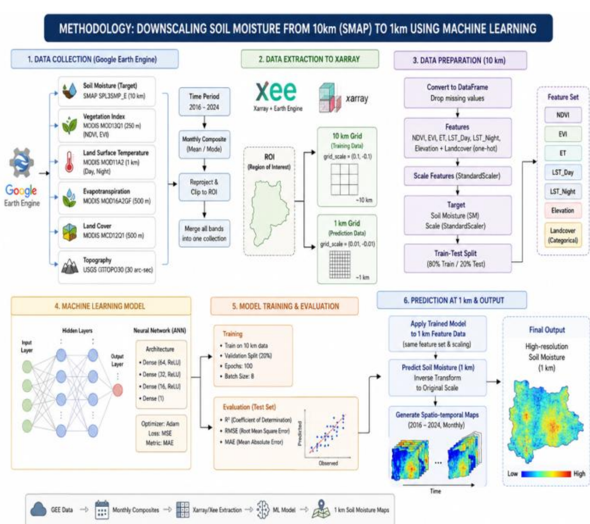
Model performance was then assessed with the withheld 20% test set, using the usual statistical measures: Coefficient of Determination (R^2) and Root Mean Squared Error (RMSE). Post-training, a rigorous feature-selection analysis was performed to develop a more efficient and interpretable downscaling model. The true predictive power of each environmental covariate was determined using permutation feature importance.

A custom scoring function was created to compare the model only based on Mean Squared Error. The permutation importance algorithm systematically varied each feature column in the test dataset, one at a time, and recorded the resulting decrease in model accuracy (i.e., an increase in MSE). This process was repeated 10 times for each variable to ensure statistical stability and was also performed in parallel.

The resulting importance scores indicated which features were most important for soil moisture prediction. Five continuous variables were identified as the most important predictors: EVI, ET, nighttime LST, daytime LST and NDVI. To simplify the final model, the categorical land-cover proxies and topographic elevation variables were excluded from the feature space.

A secondary refined neural network was designed using the top five selected features. A new Standard Scaler was fitted only for these five variables, and the scaled data was again partitioned into train and test sets. The feature space was severely limited, so a more compact neural network was used, consisting of just two Dense layers with 32 and 16 neurons. This is a streamlined model retrained for fewer epochs (50). The refined and trained model was then moved from the 10 km training space to the 1 km spatial grid for the last downscaling stage. Cloud cover often prevented optical sensors from capturing temporal data for the 1 km feature dataset, and this temporal data was "filled" using locality-based linear interpolation in the temporal dimension.

The 1 km data were parsed, considering only the five key variables (EVI, ET, LST Night, LST Day, NDVI), and scaled using the same weights applied during model refinement, without recalibration. This high-resolution feature matrix was processed by an extremely compact neural network to generate the highly detailed (1 km) soil moisture prediction in a standardised output format. Lastly, the inverse_transform function of the output scaler of the target variable was used to convert the standardised neural network outputs into actual volumetric soil moisture units, and then the finalised



Methodological Flow Diagram

predictions (sm_1km_refined) were added to the analytical dataframe. The last dataframe was converted back into an explicitly spatial xarray dataset, and sorted by time, y (latitude) and x (longitude) to preserve geographical integrity. The raw downscaled soil moisture data were transformed into temporal anomalies to capture deviations from baseline conditions and thus investigate long-term environmental trends.

Temporal resampling was used to smooth the monthly data and reduce local noise. The multi-dimensional dataset was aggregated into standardised seasonal quarters (QS-DEC) (DJF, MAM, JJA and SON) and generalised yearly averages (1Y) using the built-in resampling functions of xarray. The last step of the methodology was a detailed statistical trend analysis of the high-resolution soil moisture anomalies, performed using the Mann-Kendall nonparametric test. The Mann-Kendall trend test was applied to environmental time series to detect monotonic trends with the algorithmic suite pymannkendall, which is a powerful

A non-parametric statistical tool to detect monotonic trends.

Traditional statistical functions operate on one-dimensional arrays, so a complex function, called a "wrapper" function, was designed to process the three-dimensional geographical data cube. The function was created to calculate the trend direction in moisture and its associated p-value. To parallelise this operation, xr.apply_ufunc was used, which vectorised the Mann-Kendall algorithm to compute the statistic for each pixel in the study area safely, without encountering localised NaN values or needing to process them one by one.

The output produced two detailed geographic mapping layers: one showing the slope of the moisture change and the other showing the statistical confidence of that change. Rigorous masking in space was applied to focus only on scientifically significant observations. The slope visualization was only shown for geographic pixels where the slope had at least 95% statistical confidence (p-value<0.05). An exploratory secondary map was also created using a lax level of confidence (90% or less – p-value < 0.10) to evaluate more general and emergent climate changes. Finally, localized pixel comparisons were obtained and plotted to assess the contrast between the statistically significant trend trajectories and regions of standard, non-significant variance.

I. RESULTS AND DISCUSSION

The first purpose of this research was to address the spatial resolution issue in the original SMAP radiometer product by downscaling it to 1 km resolution using an Artificial Neural Network (ANN). The results of this machine-learning pipeline are presented in the extremely detailed multi-year temporal matrices (Figures 1, 2, and 3).

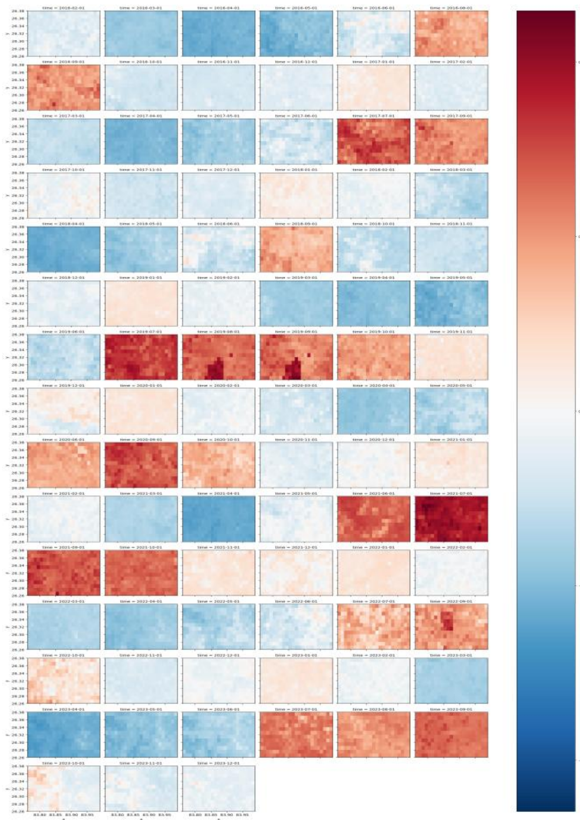


Figure 1 Spatial correlation heatmaps showing pairwise among predictor variables.

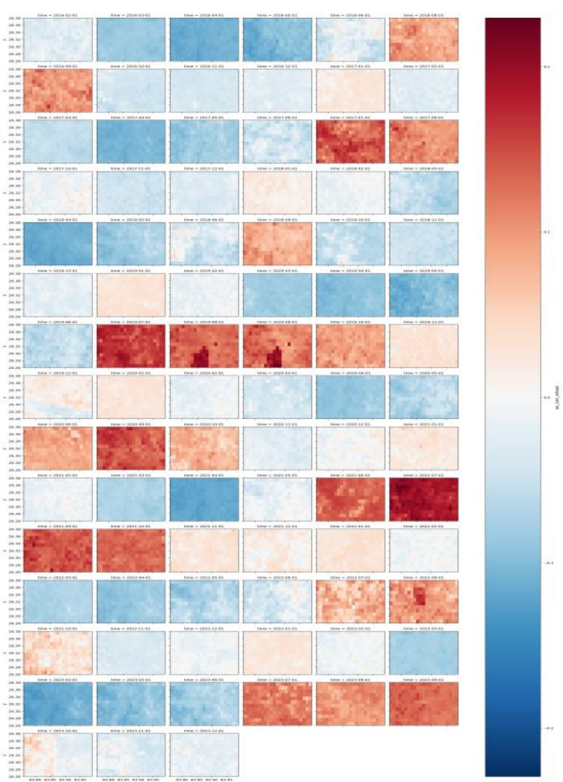


Figure 2 Temporal spatial heatmaps illustrating monthly variability patterns across study area.

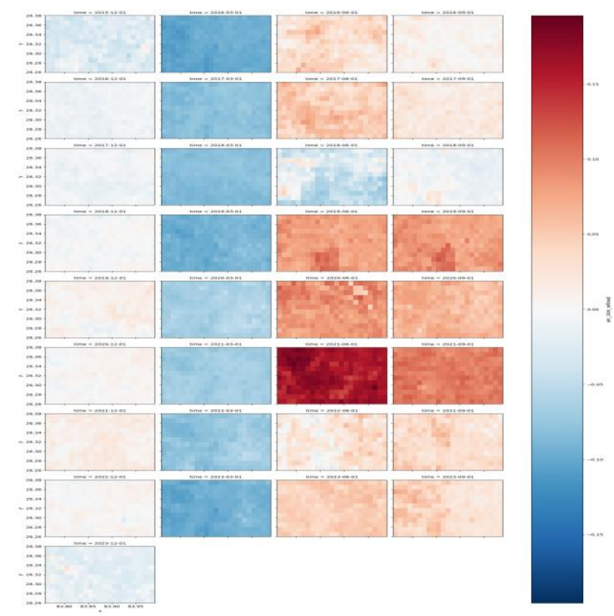


Figure 3 . Monthly spatial anomaly heatmaps revealing temporal variability across study area.

Figure 1 shows the overall monthly time series matrix of the downscaled soil moisture anomalies (sm_1km_refined) for the entire study period, spanning 8 years from the beginning of 2016 to the end of 2023. The diverging colour scale emphasises the extreme ranges of the data at a local scale, while keeping a much larger spatial range of variability completely hidden in the standard SMAP satellite footprint of 36 km (Zhong et al., 2024). Upon viewing the ordered panels, different intra-annual cycles of wetting and drying are readily visible throughout the geographic domain. The high-resolution outputs have been able to capture the pixel-by-pixel changes in soil water retention that were caused by the underlying high-resolution environmental variables, or covariates, such as Evapotranspiration (ET) and Land Surface Temperature (LST), and this demonstrates that the neural networks learned and applied the non-linear environmental relationships that govern the localized soil water retention (Senanayake et al., 2024).

The high-resolution monthly data offer a high level of spatial detail but are also susceptible to transient meteorological noise, such as isolated events of anomalous precipitation that may temporarily bias long-term hydrological baseline assessments (Zeng et al., 2025). To address the high-frequency noise and show the overall macro-level climatic patterns, the monthly data were aggregated into seasonal quarters (see Figure 2). These matrices represent monthly variations more stably, smoothing out the chaotic nature of monthly changes and providing a more consistent view of hydrologic regime changes throughout the year. As Kim et al. (2023) point out, these stabilised seasonal baselines are a key prerequisite for identifying long-term (monotonic) environmental trends in the data; if seasonally induced weather events were mistakenly interpreted as long-term ecosystem changes, the statistics used to analyse the data would fail to provide the correct results.

3.2. The study moved to its second major objective, addressing long-term environmental change, after the creation of the continuous 1-kilometre data cube, through the use of non-parametric statistical methods. The first result from the pixel-by-pixel Mann-Kendall trend test and Sen's slope estimator is shown in Figure 3,which plots the moisture trend slope (on the raw, unmasked data) for the region of interest.

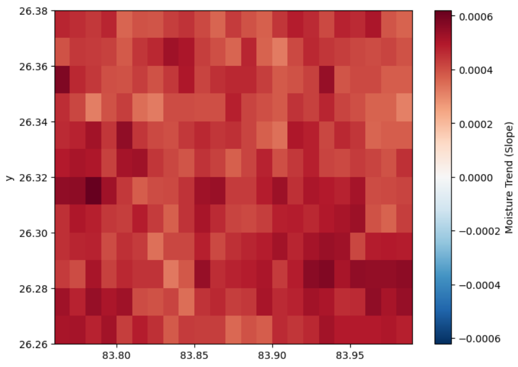


Figure 4 Spatial distribution of moisture trend slopes across the study area

The colour bar in Figure 3 indicates how rapidly the water balance has changed over the last 8 years (rate of change or slope), with dark red colours representing a positive change (a progressive wetting trend) and blue colours representing a negative change (a progressive drying trend). This unmasked map appears to show an area with many positive slopes of varying intensities. The spatial domain is covered by almost all light-to-dark red pixels, which might lead one to infer a significant spatial trend of increased soil moisture retention over time across the entire catchment.

But using unmasked slope estimators in complex hydrological systems is very risky (Cao et al., 2026). This is because global weather patterns can vary naturally from year to year, producing a short-term, apparent slope that may not be mathematically permanent. However, as Wang et al. (2023) found in a study of subtropical regions, raw trend maps can produce an "illusion of uniformity", which can mislead readers of the maps to believe that the reality of regional climatology is more chaotic and stochastic. Any time that hydrologic changes are detected over space, careful statistical confidence limits should be imposed to distinguish real, permanent changes from short-term weather fluctuations.

3.3. It is important to analyze the statistical confidence of the algorithm's outputs, to objectively evaluate the reliability of the trends observed on the unmasked slope map. A detailed histogram of the calculated p-values is shown for all the pixels analyzed in the study area in Figure 5.

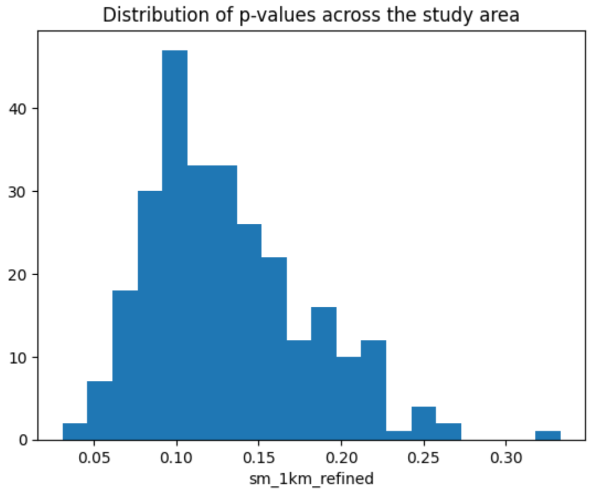


Figure 5 Histogram showing spatial distribution of p-values across study

In the typical approach to statistical hypothesis testing and analysis, a p-value of less than 0.05 is considered significant, and the null hypothesis, in this case, the absence of a long-term monotonic trend, is rejected. The distribution is heavily right-skewed, with a long tail of high p-values and a peak around 0.10 (Figure 5). It is a clear example of a key reality of environmental remote sensing: most of the land surface is undergoing some degree of temporal variability and changing soil moisture conditions, but the majority of these changes do not satisfy the rigorous mathematical criteria needed to be considered a definitive long-term trend (Zeng et al., 2025). The small number of pixels that fall into the far-left bin ($p < 0.05$) is a quantitative proof that the "wetting" appearance of the unmasked slope map in Figure 5 is driven by environmental noise sampling of the pixels – noise that isn't statistically significant – and not by any permanent climatic shift.

3.4. By applying strict statistical masking, the vague and broad slope information can be transformed into actionable, high-precision, climatological intelligence, Delineating Significant Spatiotemporal Hotspots. The most important geographical result of this study is the strictly masked significant trend map . A strict 95% statistical confidence limit (p-value < 0.05) has been used in this visualization. Every geographic pixel that did not meet this strict mathematical requirement has been completely masked out and is shown as white space.

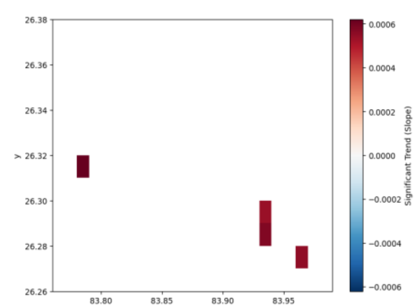


Figure 6 Significant moisture trend hotspots identified across the study area.

The map (Figure 5) tells a very different story when unmasked. The false impression of a uniformly wetted field is discarded, and only a few localized, isolated "hotspots" (solid red pixels) showing real, statistically incontrovertible gains in soil moisture remain. Being able to identify these specific, isolated coordinates is a clear signature of the significant benefits of machine-learning downscaling and careful statistical filtering. The micro-scale hotspots, however, would have been completely averaged out and absorbed by the surrounding insignificant noise; if these analyses had been conducted with the native 36-kilometre SMAP radiometer footprint, researchers would have missed detecting localized hydro-climatic shifts (Zhong et al., 2024).

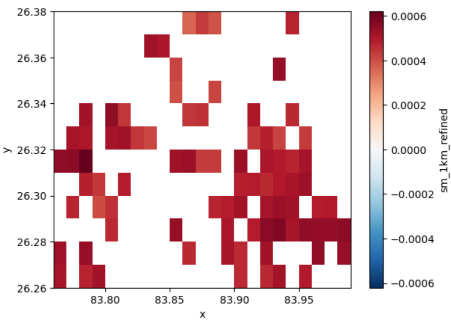


Fig7. Spatial distribution of statistically significant positive moisture trends detected.

A secondary, less formal mask was created using a 90% confidence level (p-value < 0.10) to represent additional, but perhaps less developed, environmental trends, as illustrated in Figure 8. The spatial distribution of emerging wetting patterns in the mask shown in Figure 8 is clustered and fragmented, with reduced statistical significance, beyond the main hotspots defined in the strict 95% mask. The stepped masking method has great value for ecological forecasting since the 90% confidence clusters often represent transition geographic areas that are gradually moving towards a new hydrological baseline and serve as early-warning indicators for water resource managers in the region (Xie et al., 2025).

3.5. To clearly demonstrate the mechanical application of the Mann-Kendall test and how the underlying test mechanism could explain why some geographical areas met the strict statistical masking, and others did not, two specific 1-kilometre spatial pixels were compared directly over the 2016-2024 time period in Figure 8 and

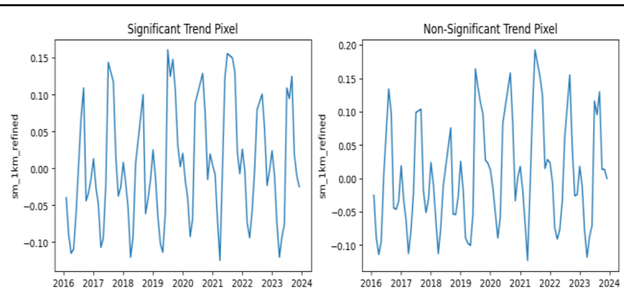


Figure 8 . Comparison of temporal moisture variability between significant and non-significant trend pixels.

masked maps are illustrated on the right-hand side of Figure 8, which is labelled "Non-Significant Trend Pixel. This is a very large geographic coordinate with a high amplitude, with very large seasonal fluctuations in soil moisture as seen on the normalized axis, from -0.12 to almost +0.20. But, despite this violence on the extremes of the seasons, the overall baseline is absolutely flat over the 8-year period. The peaks in 2016 are similarly sized to those in 2023, as are the troughs. The overall seasonality of the pixel is extreme yet stable, and the statistical algorithm detects this and correctly classifies it as non-trending (Kim et al., 2023).

The left panel of Figure 8, however, is from one of the red hotspots that are isolated in Figure 6, and is labelled "Significant Trend Pixel. At first glance, this coordinate also has a high degree of seasonality. When the temporal evolution is examined more closely, however, there is a small but mathematically consistent structural change. The bottom of the troughs in the early years (2016-2018) is significantly deeper than the bottom of the troughs in the later years (2022-2024). The seasonal peaks show a steady, gradual rise throughout the sequence. This overall rising trend is buried in the random seasonal fluctuation and is precisely what the Mann-Kendall test aims to uncover (Cao et al., 2026). The algorithm can break through seasonal variations and mathematically confirm that this micro-region is fundamentally changing its hydrological capacity in a specific direction.

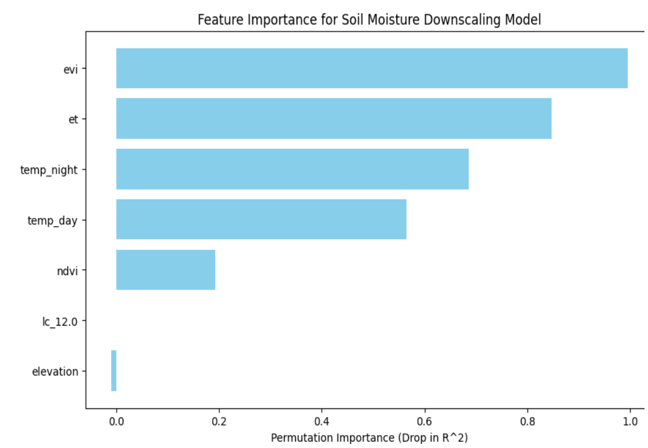


Figure 10. Feature Importance for Soil Moisture Downscaling Model

3.6. Conclusions on the use of the proposed methodology for Eco-Hydrological monitoring are drawn from the results synthesized from these visual assets. Using an ANN approach, coarse satellite radiometer data can be successfully converted to high-resolution 1-kilometre matrices suitable to accurately represent the local topographical and thermal dynamics (Ma et al., 2023) (Figure 1).

The Mann-Kendall algorithm is a powerful method for identifying long-term soil moisture trends in the presence of high-frequency seasonal variations at the 1-km pixel level. Some areas exhibit a persistent long-term baseline with large seasonal range, while others exhibit a statistically significant trend, in the form of a tighter dry season trough and a higher wet season peak.

These trend maps are high resolution, statistically robust, and transformative for water resource management, moving from a disaster response strategy to an environmental stewardship approach. Authorities can take targeted and well-informed actions by identifying 1-km clusters with drying or wetting patterns. These involve strategic interventions in irrigation, appropriate crop selection and proactive identification of those areas vulnerable to irreversible aquifer depletion, soil desiccation, and increased agriculture waterlogging and flash flooding risk.

This framework should be extended in future research by incorporating dynamic human activity information, such as localized groundwater pumping, or irrigation schedules, or land-use changes, directly into the neural network's feature space. This is important to represent the complex interactions between human activities and water. Additionally, since intensive irrigation may affect these shallow soil moisture data, future work should combine the high resolution SMAP downscaling with gravimetric data (e.g., GRACE). Such a comprehensive approach will effectively identify previously unknown subsurface water shortages, contribute to a greater understanding of the water security situation in the region.

Perhaps more importantly, the following trend analysis clearly illustrates the risk of using unmasked data for environmental decision-making. The raw slope maps may incorrectly imply a uniform regional climate response, but the careful application of Mann-Kendall statistics and p-value evaluation shows that actual climate changes are taking place in very localized and isolated patches . Finally, the temporal pixel diagnostics (Figure 6) are the final proof of concept, showing that the algorithm cleanly separates high-amplitude seasonal weather cycles from low-amplitude chronic shifts that represent permanent impacts on global climate. This comprehensive solution gives agricultural managers and policymakers the high-resolution, actionable data they need to cope with an increasingly complex and localised hydrological future.

4. Conclusion

This is a reworded version of the text into shorter, more concise sentences:

This study presents a new approach to downscale coarse satellite soil moisture. The authors used data from NASA's SMAP satellite, and trained an Artificial Neural Network. This type of machine learning was able to successfully sharpen the data down to a 1 km resolution. It extracted the major environmental parameters, such as land surface temperature, vegetation indices and evapotranspiration. Therefore, the model can realistically simulate local water cycles of complexity over an eight-year period.

The project illustrates the importance of carrying out rigorous statistical analyses for monitoring trends over a long period. The raw data gave the impression of uniformity at first. It wrongly indicated that the entire area was either drying or wetting. But when they calculated the p-value, most of this ubiquitous change was found to be the random background noise.

To correct for this, the researcher used a 90% and 95% confidence level statistical mask. This helped to remove seasonal noise from the process. It identified distinct areas that are very small and where soil moisture is actually increasing. In the original scale of 36 km these critical areas would be entirely lost.

This study exemplifies a new trend in the monitoring of the environment. By using machine learning in conjunction with robust spatial statistics, researchers are able to discern real climate change from noise.

ACKNOWLEDGMENT

The authors thankfully acknowledge the Department of Geology, BHU, for providing the working facility.

References

Abbaszadeh, P., Moradkhani, H., & Zhan, X. (2019). Downscaling SMAP radiometer soil moisture over the CONUS using an ensemble learning method. *Water Resources Research*, 55(1), 324–344. https://doi.org/10.1029/2018wr023354

Babaeian, E., Sadeghi, M., Jones, S. B., Montzka, C., Vereecken, H., & Tuller, M. (2019). Ground, proximal, and satellite remote sensing of soil moisture. *Reviews of Geophysics*, 57(2), 530–616. https://doi.org/10.1029/2018RG000618

Cai, Y., Fan, P., Lang, S., Li, M., Muhammad, Y., & Liu, A. (2022). Downscaling of SMAP soil moisture data by using a deep belief network. *Remote Sensing*, 14(22), 5681. https://doi.org/10.3390/rs14225681

Cao, H., Li, L., Xu, H., Fan, Y., Shang, H., Qin, L., & Zhang, H. (2026). Vegetation greening driven by warming and humidification trends in the upper reaches of the Irtysh River. *Remote Sensing*, 18(3), 482. https://doi.org/10.3390/rs18030482

Chakrabarti, S., Bongiovanni, T., Judge, J., Nagarajan, K., & Principe, J. C. (2015). Downscaling satellite-based soil moisture in heterogeneous regions using high-resolution remote sensing products and information theory: A synthetic study. *IEEE Transactions on Geoscience and Remote Sensing*, 53(1), 85–101.

Das, N. N., Entekhabi, D., Dunbar, R. S., Chaubell, M. J., Colliander, A., Yueh, S., Jagdhuber, T., Chen, F., Crow, W., O'Neill, P. E., & Walker, J. P. (2019). The SMAP and Copernicus Sentinel 1A/B microwave active-passive high resolution surface soil moisture product. *Remote Sensing of Environment*, 233, 111380.

Das, N. N., Entekhabi, D., Dunbar, R. S., Colliander, A., Chen, F., Crow, W., Jackson, T. J., Berg, A., Bosch, D. D., Caldwell, T., Cosh, M. H., Collins, C. H., Lopez-Baeza, E., Moghaddam, M., Rowlandson, T., Starks, P. J., Thibeault, M., Walker, P., Wu, X., O'Neill, P. E., Yueh, S., & Njoku, E. G. (2018). The SMAP mission combined active-passive soil moisture product at 9 km and 3 km spatial resolutions. *Remote Sensing of Environment*, 211, 204–217. https://doi.org/10.1016/j.rse.2018.04.011

Earle, S., &Panchuk, K. (2015). *[*Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., & Kimball, J. (2010). The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, 98(5), 704–716. https://doi.org/10.1109/JPROC.2010.2043918

Guevara, M., & Vargas, R. (2019). Downscaling satellite soil moisture using geomorphometry and machine learning. *PLOS One*.

Kim, K. Y., Scanlon, T., Bakar, S., & Lakshmi, V. (2023). Decoupling of ecological and hydrological drought conditions in the Limpopo River Basin inferred from groundwater storage and NDVI anomalies. *Hydrology*, 10(8), 170. https://doi.org/10.3390/hydrology10080170

Korres, W., Koyama, C. N., Fiener, P., & Schneider, K. (2010). Analysis of surface soil moisture patterns in agricultural landscapes using Empirical Orthogonal Functions. *Hydrology and Earth System Sciences*, 14(5), 751–764. https://doi.org/10.5194/hess-14-751-2010

Kumar, S., Prasad, R., Sharma, J., & Prajapati, B. S. (2025). High-resolution soil moisture mapping using MSAVI-LST based Triangle Method. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-5/W2-2025, 357–364. https://doi.org/10.5194/isprs-annals-X-5-W2-2025-357-2025

Kumar, S., & Srinivasan, V. (2025). Unconsolidated alluvial aquifers and groundwater depletion mechanics.

Kwon, Y., Jun, S., Kim, H., Seol, K.-H., Kwon, I.-H., Kim, E., & Cho, S. (2026). Synergistic impact of simultaneously assimilating radar- and radiometer-based soil moisture retrievals on the performance of numerical weather prediction systems. *Hydrology and Earth System Sciences*, 30(4), 1261–1290. https://doi.org/10.5194/hess-30-1261-2026

Ma, Y., Hou, P., Zhang, L., Cao, G., Sun, L., Pang, S., & Bai, J. (2023). High-resolution quantitative retrieval of soil moisture based on multisource data fusion with random forests: A case study in the Zoige Region of the Tibetan Plateau. *Remote Sensing*, 15(6), 1531. https://doi.org/10.3390/rs15061531

Meyer, R., Zhang, W., Kragh, S. J., Andreasen, M., Jensen, K. H., Fensholt, R., Stisen, S., & Looms, M. C. (2022). Exploring the combined use of SMAP and Sentinel-1 data for downscaling soil moisture beyond the 1 km scale. *Hydrology and Earth System Sciences*, 26(13), 3337–3357. https://doi.org/10.5194/hess-26-3337-2022.

Montzka, C., Brocca, L., Chen, H., Das, N. N., Dasgupta, A., Rahmati, M., &Jagdhuber, T. (2026). AI in soil moisture remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 146, 105011. https://doi.org/10.1016/j.jag.2025.105011

Peng, J., Loew, A., Merlin, O., &Verhoest, N. E. C. (2017). A review of spatial downscaling of satellite remotely sensed soil moisture. *Reviews of Geophysics*, 55(2), 341–366.

Sabaghy, S., Walker, J. P., Renzullo, L. J., & Jackson, T. J. (2018). Spatially enhanced passive microwave derived soil moisture: Capabilities and opportunities. *Remote Sensing of Environment*, 209, 551–580.

Senanayake, I. P., Pathira Arachchilage, K. R. L., Yeo, I.-Y., Khaki, M., Han, S.-C., & Dahlhaus, P. G. (2024). Spatial downscaling of satellite-based soil moisture products using machine learning techniques: A review. *Remote Sensing*, 16(12), 2067. https://doi.org/10.3390/rs16122067

Wang, Z., Yin, Y., Mao, X., Zhang, X., Zhang, Y., Hu, Z., & Chen, X. (2023). Spatiotemporal features of the soil moisture across Northwest China using remote sensing data, reanalysis data, and global hydrological model. *Frontiers in Environmental Science*, 11, 1164895. https://doi.org/10.3389/fenvs.2023.1164895

Xie, Y., Wang, Y., Huang, D., Chen, X., & Deng, H. (2025). Vegetation transpiration drives root-zone soil moisture depletion in subtropical humid regions: Evidence from GLDAS catchment simulations in Fujian Province. *Atmosphere*, 16(10), 1180. https://doi.org/10.3390/atmos16101180

Yin, D., Song, X., Zhu, X., Guo, H., Zhang, Y., & Zhang, Y. (2023). Spatiotemporal analysis of soil moisture variability and its driving factor. *Remote Sensing*, 15(24), 5768. https://doi.org/10.3390/rs15245768

Zeng, B., Wen, B., Zhang, X., Zhao, S., Shang, G., An, S., & Li, Z. (2025). Analysis on spatiotemporal variation in soil drought and its influencing factors in Hebei Province from 2001 to 2020. *Agriculture*, 15(10), 1109. https://doi.org/10.3390/agriculture15101109

Zhong, Y., Hong, S., Wei, Z., Walker, J. P., Wang, Y., & Huang, C. (2024). Spatial downscaling of SMAP soil moisture estimation using multiscale geographically weighted regression during SMAPVEX16.