ISSN 2682-681X (Paper), ISSN 2705-4241 (Online) | http://unilorinjoger.com | https://doi.org/10.63745/joger.2026.06.30.001

JOURNAL OF GEOMATICS AND ENVIRONMENTAL RESEARCH
www.unilorinjoger.com

RESEARCH ARTICLE

Development of a Geospatial Machine Learning Solution for Soil Moisture Content Prediction over Nigeria

Mefe Moses*,

Department of Geomatics, Ahmadu Bello University, Zaria, Nigeria

*Corresponding author email: mosesmefe@gmail.com

Abstract

This study presents a geospatial machine learning framework for predicting soil moisture content across Nigeria using satellite-derived and environmental data. Recognizing the limitations of traditional in-situ soil moisture measurements, such as high cost, labour intensity, and limited spatial coverage, the study employs a Random Forest Regressor trained on Tropical Applications of Meteorology using Satellite (TAMSAT) soil moisture estimates, alongside climatic, soil, and environmental variables. Model performance was evaluated over an 11-year validation period (2013–2023) across Nigeria's Agro-ecological zones. The results demonstrate high predictive accuracy, particularly for surface soil moisture (Layer 1), with Mean Absolute Error (MAE) values between 1.085 and 1.194 kg/m² and Root Mean Square Error (RMSE) ranging from 1.739 to 1.931 kg/m². The model explains over 94% of the variance in surface soil moisture, as indicated by strong correlation (R > 0.986) and explanatory power (R² > 0.973). Subsurface soil moisture predictions (Layer 2) show slightly higher errors (MAE: 1.688–1.919 kg/m²; RMSE: 2.629–2.890 kg/m²), but maintain very high correlation and explanatory power (R > 0.991; R² > 0.981). Spatially, the model performs best in northern and central Nigeria, where soil moisture conditions are relatively stable, while higher uncertainties are observed in southern coastal regions and around Lake Chad due to complex hydrological interactions. Temporally, model accuracy is highest during the dry season, particularly in January. The study demonstrates the model's potential for precision agriculture, irrigation planning, and drought mitigation, while recommending further refinement through the integration of high-resolution hydrological and groundwater datasets.

ARTICLE HISTORY
Received: 01 January, 2026
Accepted: 10 March, 2026
Published: 18 April 2026

KEYWORDS
Soil moisture prediction; machine learning; remote sensing; Random Forest; hydrological modelling

Citation: Moses M. (2026). Development of a Geospatial Machine Learning Solution for Soil Moisture Content Prediction over Nigeria. *Journal of Geomatics and Environmental Research*, *9(1), Pp 1-17*

1. INTRODUCTION

Soil moisture plays a crucial role in agriculture and water resource management, influencing hydrological processes such as evapotranspiration, infiltration, and runoff (Babaeian *et al.*, 2019). Accurate soil moisture information is essential for optimizing irrigation scheduling, predicting crop yields, managing drought, and mitigating flood risks. In Nigeria, where agriculture contributes significantly to the economy, effective soil moisture monitoring is particularly important for enhancing agricultural productivity and ensuring sustainable water resource management. However, traditional methods for soil moisture measurement, such as in-situ measurements, are often laborious, time-consuming, and expensive, while providing limited spatial coverage (Babaeian *et al.*, 2019). Therefore, there is a growing need for efficient and cost-effective approaches to soil moisture prediction, especially in regions like Nigeria with limited resources and infrastructure.

Soil moisture is essential for agriculture and terrestrial ecosystem function: it regulates water and nutrient availability, influences crop productivity, and affects evapotranspiration, infiltration, and runoff (Ochsner *et al.*, 2013). By maintaining soil structure and microbial activity, adequate soil moisture is essential for sustainable agricultural practices and soil health. Agriculture contributes approximately 22% of Nigeria's GDP and supports over 70% of the population (World Bank Group, 2020). Consequently, effective soil moisture management is vital for the country's economic stability. The country's dependence on rainfed agriculture makes it highly vulnerable to climatic variability, particularly in its diverse agroecological zones, ranging from the arid northern regions to the humid southern zones (Adedolapo & Ajetomobi, 2020). However, limited access to reliable soil moisture monitoring infrastructure and the high cost of in-situ measurement technologies hinder efforts to optimize water use and improve agricultural resilience. These challenges necessitate innovative solutions for large-scale and cost-effective soil moisture assessment.

Satellite remote sensing offers a promising solution for large-scale soil moisture monitoring, providing spatially continuous and temporally frequent observations (Srivastava *et al.*, 2016). Several satellite missions, such as NASA's Soil Moisture Active Passive (SMAP) and the European Space Agency's Sentinel-1, offer valuable data products relevant to soil moisture estimation (Bauer-Marschallinger *et al.*, 2021; Entekhabi *et al.*, 2010; Peng *et al.*, 2021). These datasets, when combined with advanced machine learning techniques, enable the development of predictive models that can account for regional variability and improve soil moisture estimation accuracy. Machine learning algorithms, such as Random Forests (RF), Support Vector Machines (SVM), and Neural Networks (NN), have demonstrated their potential in environmental modelling, offering flexible and scalable solutions to complex prediction tasks (Pal & Sharma, 2021; Reichstein *et al.*, 2019).

Machine learning models can learn complex relationships between satellite-derived features and soil moisture, enabling accurate and efficient predictions (Hossain & Kabir, 2023). Studies have demonstrated the potential of machine learning for various environmental applications, including land surface modelling (Pal & Sharma, 2021), drought prediction (Pallapoth, 2025), and ionospheric modelling (Moses *et al.*, 2020; Reddybattula *et al.*, 2022; Smirnov *et al.*, 2023). For instance, Osman et al. (2025) provide a comprehensive review of machine learning approaches for drought monitoring and forecasting, demonstrating how hybrid and deep learning models can effectively capture complex nonlinear relationships in hydro-climatic data.

Despite notable advances, a critical methodological gap persists between satellite data validation, climatological analysis, and operational soil moisture prediction in Nigeria. While previous studies have established the reliability of satellite-derived soil moisture products using International Soil Moisture Network (ISMN) observations (Moses, 2025a) and documented long-term spatial and temporal variability across the country (Moses, 2025b) Nigeria still lacks a robust, data-driven predictive framework capable of translating satellite and environmental information into accurate, national-scale soil moisture estimates. This limitation is further compounded by persistent challenges in applying machine learning approaches to Nigerian agricultural systems, including data scarcity, limited access to high-quality satellite products, and the need for models that can effectively accommodate Nigeria's diverse environmental conditions. Strong heterogeneity in soil texture, land use, and vegetation cover increases modelling complexity, leading to spatial and seasonal inconsistencies in prediction accuracy and reliability. Consequently, the practical utility of existing soil moisture datasets for precision agriculture, drought mitigation, and water resource management remains constrained, particularly in regions where in-situ observations are sparse or entirely absent. Addressing these interconnected gaps requires a comprehensive and scalable framework that integrates satellite observations, local environmental parameters, and advanced machine learning techniques to improve soil moisture prediction across Nigeria (Peng *et al.*, 2021). This study aims to develop machine learning models for soil moisture prediction using satellite-derived data to address the limitations of traditional methods and enhance agricultural productivity and water resource management in Nigeria by providing timely and accurate soil moisture information for informed decision-making. The findings contribute to the growing body of literature on machine learning in environmental modelling and provide valuable insights for soil moisture monitoring and management in Nigeria.

2. MATERIALS AND METHODS

2.1 The Study Area

Nigeria, situated in West Africa and bordered by the Gulf of Guinea between Benin and Cameroon, covers an area of 923,768 km² and spans latitudes 4° to 14° N and longitudes 2° to 15° E (Olasore *et al.*, 2021). This extensive landmass exhibits remarkable spatial variability in climate and topography that governs the distribution of its natural vegetative zones. As a tropical country, Nigeria experiences a climate that is seasonally damp and very humid in the south, while large parts of the north receive significantly lower rainfall and have shorter rainy seasons. Such climatic gradients, coupled with variations in temperature, humidity, and soil characteristics, play a decisive role in determining the types of indigenous vegetation that can thrive in different regions (FAO, 2007). The country's diverse agro-ecological zones (AEZs) result from the interplay of natural climatic forces and human activities, which directly impact soil moisture. A widely accepted classification of Nigeria's AEZs includes eight principal zones (see Figure 1). The Mangrove Forest and Coastal Vegetation zone, located along the Niger Delta, is influenced by brackish waters and is characterized by poorly aerated, waterlogged, and saline soils. This zone is primarily associated with coastal swamps that are only selectively cultivated, often for swamp rice, where stabilization is achieved. Immediately inland lies the Freshwater Swamp Forest zone, which occupies low-lying areas that are regularly flooded by rainwater. This zone supports a mosaic of palm and fibre plants, such as various Raphia species and oil palm, and is critical for both fishing and fibre production due to the abundant silt and sediment deposition (Numbere & Numbere, 2018).

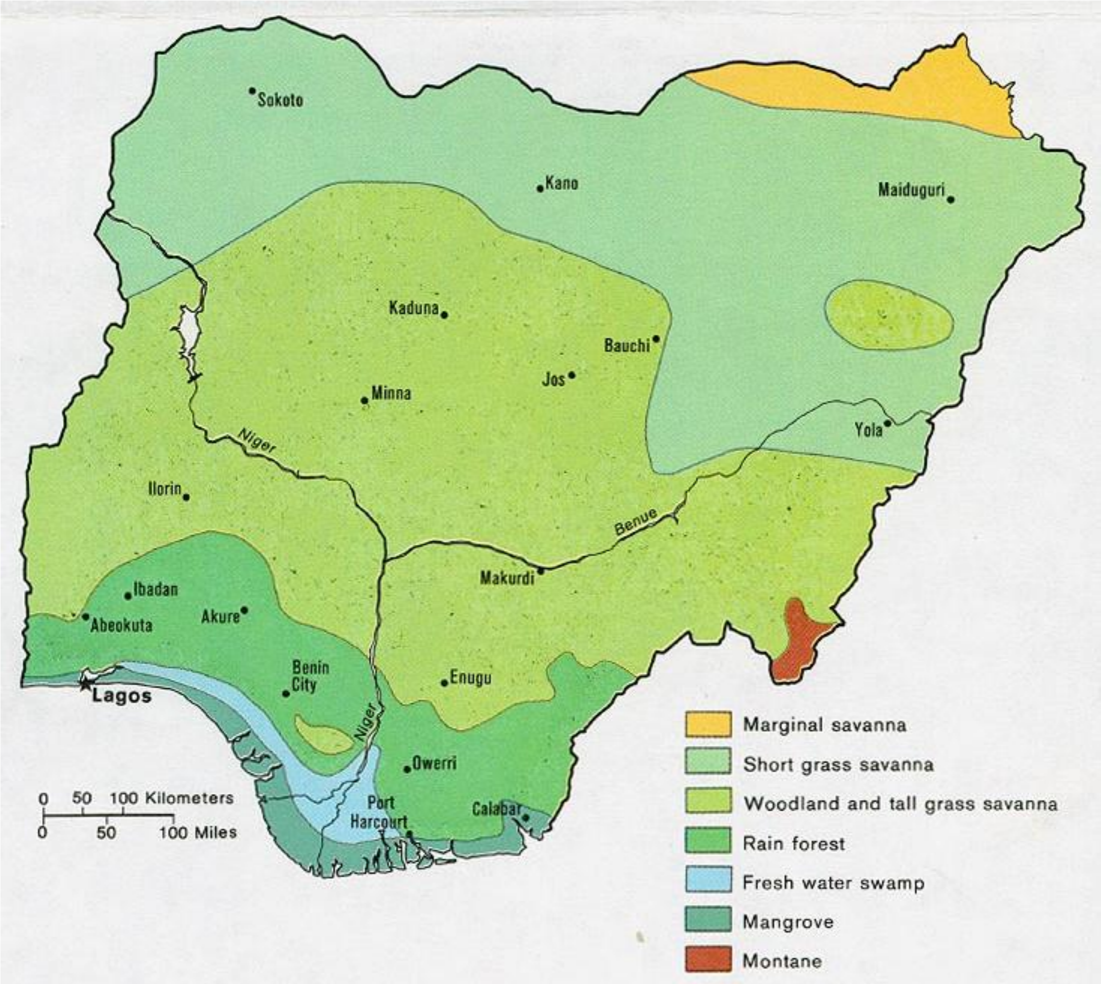

Figure 1. Agro-ecological zones in Nigeria (source: https://www.agriculturenigeria.com/agro-ecological-zone/)

Further north, the Tropical High Forest Zone experiences a prolonged rainy season with annual rainfall frequently exceeding 2000 mm. This climate supports dense vegetation and a variety of plantation and staple crops, including cocoa, rubber, coffee, and oil palm. However, intense human activity over the years has led to significant forest degradation and conversion to agricultural land in parts of this zone. Transitioning from these humid forests, the Derived Guinea Savannah serves as an intermediary between the dense tropical forests and the more open savanna zones, with an average annual rainfall of approximately 1314 mm and mean temperatures around 26.5ºC. Here, extensive bush burning, overgrazing, and cultivation have replaced much of the original forest with grasses and scattered trees, favouring the production of crops like maize, cassava, yams, and rice (Ekwealor *et al.*, 2020; Yusuf *et al.*, 2017).

The Sudan Savannah (shortgrass savannah) in the northwest experiences' lower rainfall than the Guinea Savannah, and its vegetation is dominated by short grasses. The Sahel Savannah (Marginal Savannah) in the far northeast receives less than 700 mm of rainfall annually and is characterized by extremely sparse vegetation (Yusuf *et al.*, 2017). These arid conditions necessitate irrigation for sustained cultivation and are primarily used by nomadic herders for grazing livestock. Higher elevations, including the Jos Plateau, Mandara, Adamawa, and Obudu mountains, comprise the Montane Vegetation zone, where milder temperatures (average annual temperature of approximately 21.5ºC) and moderate rainfall (about 1450 mm) create a mosaic of grasslands and forested slopes. Although terrain constraints limit large-scale commercial farming, this zone supports certain crops and extensive pastoral activities, particularly by nomadic Fulani herders. This agro-ecological diversity is critical for soil moisture modelling because it encapsulates a wide spectrum of environmental conditions, ensuring that the developed model is robust and generalizable across different climatic zones.

2.2 Data Collection

This study utilized soil moisture data from the Tropical Applications of Meteorology using Satellite and Ground-based Observations (TAMSAT) dataset (https://data.tamsat.org.uk/). Daily soil moisture estimates at rooting depth were obtained for the period spanning January 1983 to December 2023, with a spatial resolution of 0.25° × 0.25°. These satellite-derived datasets were selected due to their extensive temporal coverage and compatibility with high-resolution geospatial analysis, making them suitable for training machine learning models for soil moisture prediction over Nigeria. Other ancillary data collected included climatic variables such as evaporation, potential evaporation, precipitation, runoff components, soil temperature, and total column water vapor. Additionally, soil variables were considered, including sand, silt, clay, and gravel content, bulk density, volumetric water content at -10 kPa, topsoil texture, and available water capacity. These factors were chosen based on their influence on soil texture, drainage, water retention, and compaction, providing comprehensive environmental characterization for modelling soil moisture dynamics. Spatio-temporal attributes, including latitude, longitude, year, and day of the year, were incorporated to capture geospatial and temporal variability in soil moisture patterns.

2.3 Data Preprocessing

The dataset processing workflow involved extracting soil moisture data from NetCDF files stored in a structured repository. Each file in the annual subdirectory containing the raw data represented the daily estimations of soil moisture content for that particular day. The geographic extent for Nigeria was defined as longitude 2.375°E to 14.875°E and latitude 3.875°N to 14.125°N. The soil moisture content values within this extent were extracted after identifying the latitude and longitude indices that corresponded to the Nigerian boundary. Since the dataset covers several decades (1983 to 2023), changes in calendar years, including leap years, were taken into account. The extracted data for each year was concatenated along the temporal dimension, and all yearly data were subsequently stacked into a five-dimensional array. The final processed dataset was stored as a pickle file for efficient retrieval and further geospatial machine learning analysis. The pickle format was chosen due to its ability to preserve multi-dimensional data structures that support seamless integration with machine learning frameworks. Climatic, soil, and environmental variables in NetCDF format were processed using Python libraries such as NumPy, Pandas, and netCDF4.

2.4 Feature Selection and Engineering

The feature selection methodology is based on the inherent capability of Random Forest models to estimate the relative importance of each predictor variable in forecasting the target outcome. To ensure data quality and consistency, rows of the target values with missing values were dropped. A random sample of one million rows was extracted from the multi-decade dataset. This sample size was determined to balance computational efficiency with spatial and temporal representativeness. By employing a stratified random sampling approach across the eight agro-ecological zones and the 41-year study period, we ensured that the training set captured the full range of climatic and soil moisture variability present in the original dataset. An 80/20 ratio is used to divide the target dataset for each of the two layers into training and testing groups. After training a Random Forest Regressor on the training data, the model's built-in feature importance attribute is used to calculate the contribution of each feature to the prediction of that target variable. Finally, the resulting importance scores are recorded and visualized using a line plot where each target variable's importances are plotted against the list of features. In order to guide the refinement of the model and guarantee that the most important predictors are given priority in subsequent analyses, the importance scores plot makes it easier to clearly compare which properties are most influential across the two soil moisture layers being predicted.

2.5 Model Development and Hyperparameter Tuning

A Random Forest Regressor (RFR) was chosen for the model's implementation using Scikit-learn because of its robustness in handling nonlinear relationships and high-dimensional data. To ensure reproducibility, the model was initialised with 100 estimators and a fixed random state. The dataset was partitioned into training (80%) and testing (20%) subsets to facilitate model evaluation. Additionally, a holdout dataset, which was not utilised during training, was reserved for assessing the model's ability to generalise to unseen data. Figure 2 shows the adopted methodology for the model development.

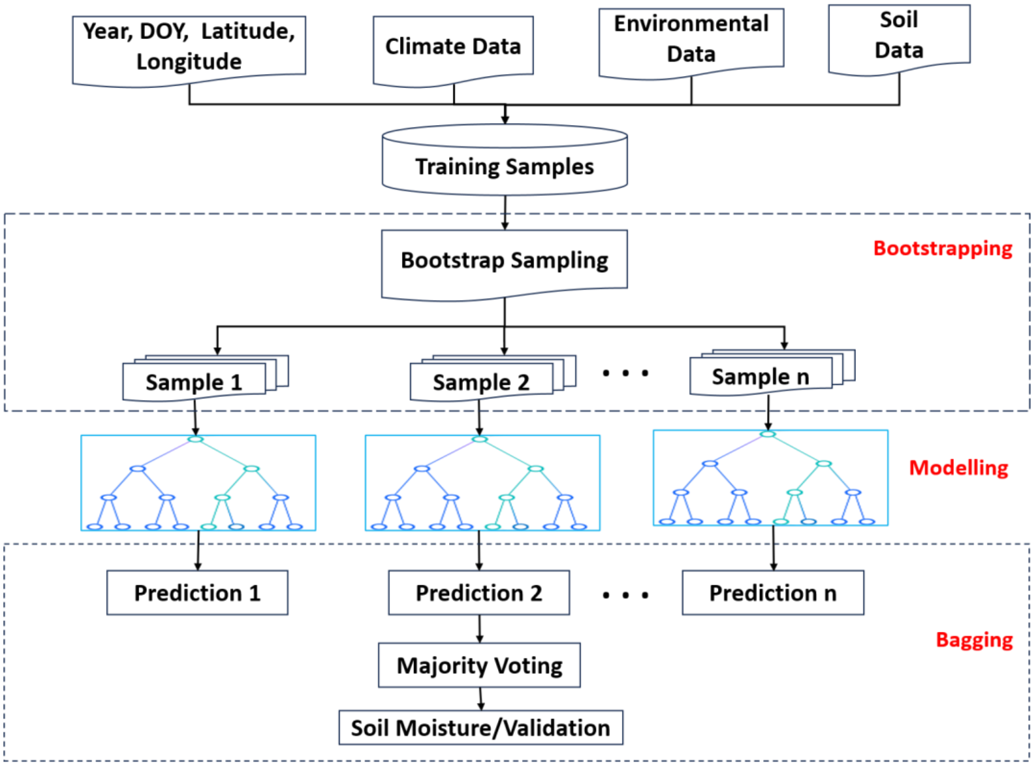

Figure 2. Methodological workflow for soil moisture prediction model development, illustrating the sequential steps from data acquisition through preprocessing, feature selection, Random Forest model training with hyperparameter tuning, to final model validation and spatial-temporal performance assessment.

To enhance the predictive performance of the model, hyperparameter tuning was conducted to identify the optimal model configuration. Key hyperparameters were systematically changed during the tuning process, and the model's performance on training and validation datasets was assessed. For optimization, three important hyperparameters were taken into account. First, the number of decision trees in the ensemble is determined by the number of estimators (n_estimators). The impact of this parameter on generalization was evaluated by varying it between 100 and 200. The maximum depth of each decision tree (max_depth), which regulates model complexity and prevents overfitting, was the second hyperparameter. Three values were examined: 10, 20, and None (signalling unbounded tree depth). The minimal number of samples needed to split an internal node (min_samples_split), the third hyperparameter, helps avoid overfitting by making sure splits only take place when there are enough data points available. Two values were considered: 2 and 5. The combination of these hyperparameters resulted in a total of 12 different model configurations, which were iteratively tested to identify the most effective setup for soil moisture prediction.

The dataset was randomly partitioned into training (80%) and testing (20%) subsets to facilitate performance assessment. Each hyperparameter combination was used to train an RFR model on the training dataset, after which the trained model generated soil moisture predictions for both training and test datasets. Model performance was assessed using three primary evaluation metrics. Mean Absolute Error (MAE) was used to quantify the average absolute difference between observed and predicted soil moisture values. Root Mean Squared Error (RMSE) was employed to emphasize larger deviations by squaring errors before averaging. Additionally, the coefficient of determination (R^2) was computed to assess the proportion of variance explained by the model.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y_i}| \tag{1}$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y_i})^2}{\sum(y_i - \bar{y})^2} \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2} \tag{3}$$

These metrics were calculated separately for each soil moisture layer. For each hyperparameter combination, the primary selection criterion was the average Test MAE across all soil moisture layers.

2.6 Performance Evaluation

To assess the model's temporal consistency, performance evaluation was conducted separately for each year within the dataset, spanning from 1983 to 2023. The methodology comprised the extraction of pertinent geographical, climatic, and soil property variables together with daily soil moisture data. To ensure that only complete records were used for model evaluation, and to preserve data integrity and avoid biases in performance evaluation, missing values in the dataset were dropped by removing incomplete observations before making predictions. The trained model was used to forecast soil moisture at the two respective layers for 11 independent years (2013-2023). The validation metrics, R, R^2, RMSE, and MAE were computed for each year to evaluate the model's performance over time. Also, the metrics were computed at each grid over the country to assess the model's spatial performance over the entire country.

3. RESULTS AND DISCUSSION

3.1 Feature importance for soil moisture prediction

Feature selection was performed using a feature importance ranking approach to identify and retain only the most relevant predictors for soil moisture prediction across two distinct layers. As illustrated in Figure 3, the analysis revealed that the volumetric soil water content in layer 1 emerged as the most influential predictor, registering importance scores of 0.59876 for layer 1 and 0.54555 for layer 2. This finding underscores the dominant role of soil water content in influencing the model's performance across both layers.

Temporal and geographic variables also demonstrated significant contributions to capturing the inherent variability in soil moisture. In particular, the day of the year (DOY) exhibited importance scores of 0.11451 for layer 1 and 0.09967 for layer 2, suggesting that seasonal patterns are vital for explaining moisture dynamics. Latitude further enhanced the model's predictive capability, with importance values of 0.06066 for layer 1 and 0.15797 for layer 2, highlighting the role of spatial variation in influencing soil moisture distribution.

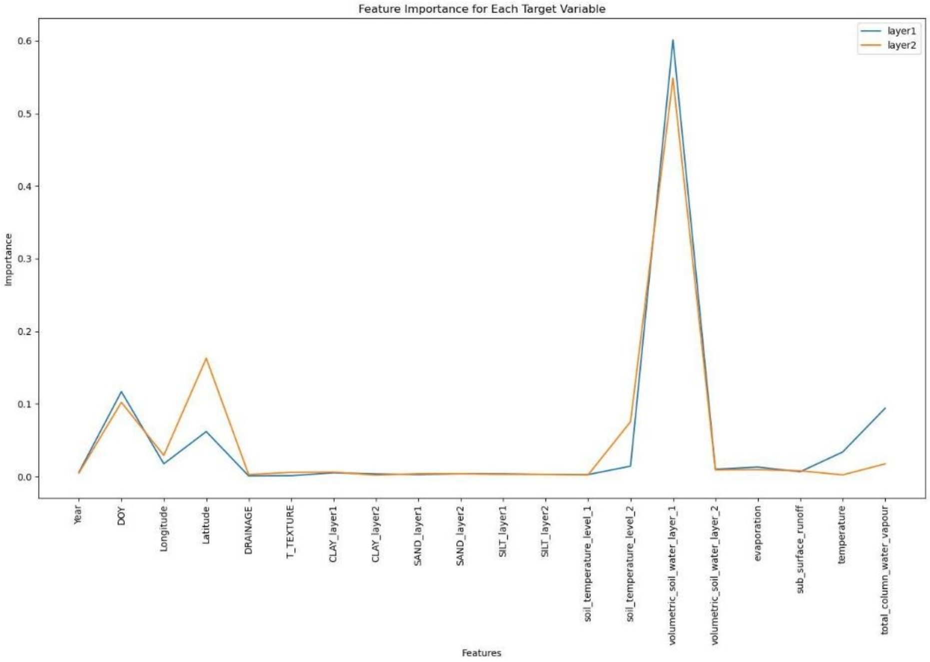

Figure 3. Feature importance scores for each target soil moisture layer (Random Forest).

Layer-specific influences were evident in the analysis. For soil moisture in layer 1, additional predictors such as total column water vapor (0.06729), temperature (0.02809), and total column water (0.02102) contributed substantially to the model, thereby reinforcing the importance of atmospheric and thermal conditions in this layer. Conversely, soil moisture in layer 2 was notably affected by soil temperature at level 2 (0.07259) and, to a lesser extent, by longitude (0.02616), indicating that these factors play a more critical role in this particular layer.

Notably, several features typically associated with soil physical properties, such as various clay, sand, and silt fractions, as well as bulk density measurements, consistently exhibited very low importance scores (generally below 0.01). Although these features may be intrinsically valuable from a soil science perspective, their minimal contribution to the model's performance suggests that excluding them could reduce model complexity and lower the risk of overfitting without compromising predictive accuracy.

Based on these findings, a joint model was developed that integrates a common set of high-importance features. This model incorporates the dominant predictors, volumetric soil water content in layer 1, DOY, and latitude, while also including layer-specific variables, such as total column water vapor and temperature for layer 1, and soil temperature at level 2 and longitude for layer 2. This refined feature set not only simplifies the model but also ensures that critical temporal, spatial, and atmospheric factors are adequately represented, thereby enhancing the overall robustness and interpretability of the soil moisture predictions.

3.2 Optimal model hyperparameters

The hyperparameter tuning of the Random Forest model for soil moisture prediction was carried out separately for two soil layers. The performance was evaluated based on the average Test Mean Absolute Error (MAE), along with other metrics such as RMSE and R-squared. Various combinations of hyperparameters were tested, including changes in the number of trees (n_estimators), maximum tree depth (max_depth), and the minimum number of samples required to split a node (min_samples_split).

Table 1: Hyperparameter Tuning Results for Random Forest Soil Moisture Prediction.

n_estimators	max_depth	min_samples_split	Layer	Train MAE (kg/m²)	Train RMSE kg/m²	Train R^2	Test MAE kg/m²	Test RMSE (kg/m²)	Test R^2
100	None	2	1	0.476	0.749	0.996	1.280	1.995	0.972
100	None	2	2	0.755	1.137	0.997	2.033	3.055	0.979
100	None	5	1	0.623	0.973	0.993	1.284	1.998	0.972
100	None	5	2	0.924	1.393	0.996	2.039	3.060	0.979
100	10	2	1	1.870	2.802	0.945	1.904	2.875	0.942
100	10	2	2	3.419	4.784	0.948	3.487	4.919	0.945
100	10	5	1	1.870	2.802	0.945	1.904	2.875	0.942
100	10	5	2	3.419	4.784	0.948	3.487	4.919	0.945
100	20	2	1	0.717	1.050	0.992	1.297	2.006	0.972
100	20	2	2	1.174	1.708	0.993	2.083	3.094	0.978
100	20	5	1	0.796	1.176	0.990	1.300	2.009	0.971
100	20	5	2	1.263	1.833	0.992	2.087	3.099	0.978
200	None	2	1	0.473	0.741	0.996	1.277	1.990	0.972
200	None	2	2	0.749	1.124	0.997	2.029	3.048	0.979
200	None	5	1	0.621	0.967	0.993	1.282	1.995	0.972
200	None	5	2	0.919	1.383	0.996	2.036	3.055	0.979
200	10	2	1	1.868	2.802	0.945	1.903	2.874	0.942
200	10	2	2	3.422	4.787	0.948	3.491	4.921	0.945
200	10	5	1	1.868	2.803	0.945	1.903	2.874	0.942
200	10	5	2	3.422	4.787	0.948	3.491	4.921	0.945
200	20	2	1	0.715	1.045	0.992	1.294	2.003	0.972
200	20	2	2	1.171	1.701	0.993	2.080	3.090	0.978
200	20	5	1	0.794	1.172	0.990	1.298	2.007	0.972
200	20	5	2	1.259	1.827	0.992	2.085	3.095	0.978

Across all the experiments, models with n_estimators set to 100 generally yielded slightly higher Test MAE values than those with 200 trees. For instance, with n_estimators of 100, max_depth set to None, and min_samples_split of 2, the Test MAE was 1.2803 kg/m² for layer1 and 2.0330 kg/m² for layer2. In contrast, when n_estimators was increased to 200 while maintaining max_depth at None and min_samples_split at 2, the Test MAE improved marginally to 1.2775 kg/m² for layer1 and 2.0286 kg/m² for layer2. This slight improvement indicates that increasing the ensemble size enhances model stability and reduces variance, a finding supported by Breiman's seminal work on Random Forests (Breiman, 2001) and further elaborated by Probst *et al.* (2019).

The influence of tree depth was particularly notable. The model's performance significantly declined when the max_depth parameter was restricted (for example, set to 10). The Test MAE rose to 1.9039 kg/m^2 for layer 1 and 3.4871 kg/m^2 for layer 2 when the n_estimators were set to 100, the max_depth to 10, and the min_samples_split to 2. With max_depth still set at 10 and n_estimators raised to 200, the Test MAE stayed high at 1.9027 kg/m^2 for layer 1 and 3.4905 kg/m^2 for layer 2. This implies that in order to capture the intricate, nonlinear interactions present in soil moisture dynamics, trees must be allowed to grow without depth limitations (max_depth=None). Model performance was also influenced by the parameter min_samples_split, which regulates the bare minimum of samples needed to split an internal node. While a value of 2 generally yielded lower errors, raising this number to 5 caused both layers' Test MAE to slightly increase. For instance, under the optimal configuration with n_estimators=200 and max_depth=None, switching min_samples_split from 2 to 5 increased the Test MAE for layer2 from 2.0286 to 2.0357 kg/m^2. This indicates that finer splits (i.e., using a lower min_samples_split) allow the model to better capture subtle variations in the data. The optimal model was configured with the following parameters: bootstrap enabled (True), using the squared error criterion for node splitting, no restriction on tree depth (max_depth=None), all features considered at each split (max_features=1), no limits on leaf nodes (max_leaf_nodes=None) or samples per split (min_samples_split=2, with min_samples_leaf set to 1), and a single ensemble comprising 200 estimators. Additional settings included ccp_alpha set to 0 (no post-pruning), no subsampling of the training set (max_samples=None), and all other parameters at their default values with a fixed random state of 42 for reproducibility.

Under these settings, the model achieved Test MAEs of approximately 1.2775 kg/m^2 for layer 1 and 2.0286 kg/m^2 for layer 2, with high Test R-squared values (0.972 and 0.979, respectively). Figure 4 shows the scatter plots for the training and testing data for both layers. These results indicate that the model generalizes well to unseen data, effectively capturing both temporal and spatial variations in soil moisture. The observed performance improvements with increased n_estimators and unrestricted tree depth are consistent with established literature, which emphasizes that larger, more flexible ensembles can better accommodate the complexity of environmental data (Breiman, 2001; Probst *et al.*, 2019).

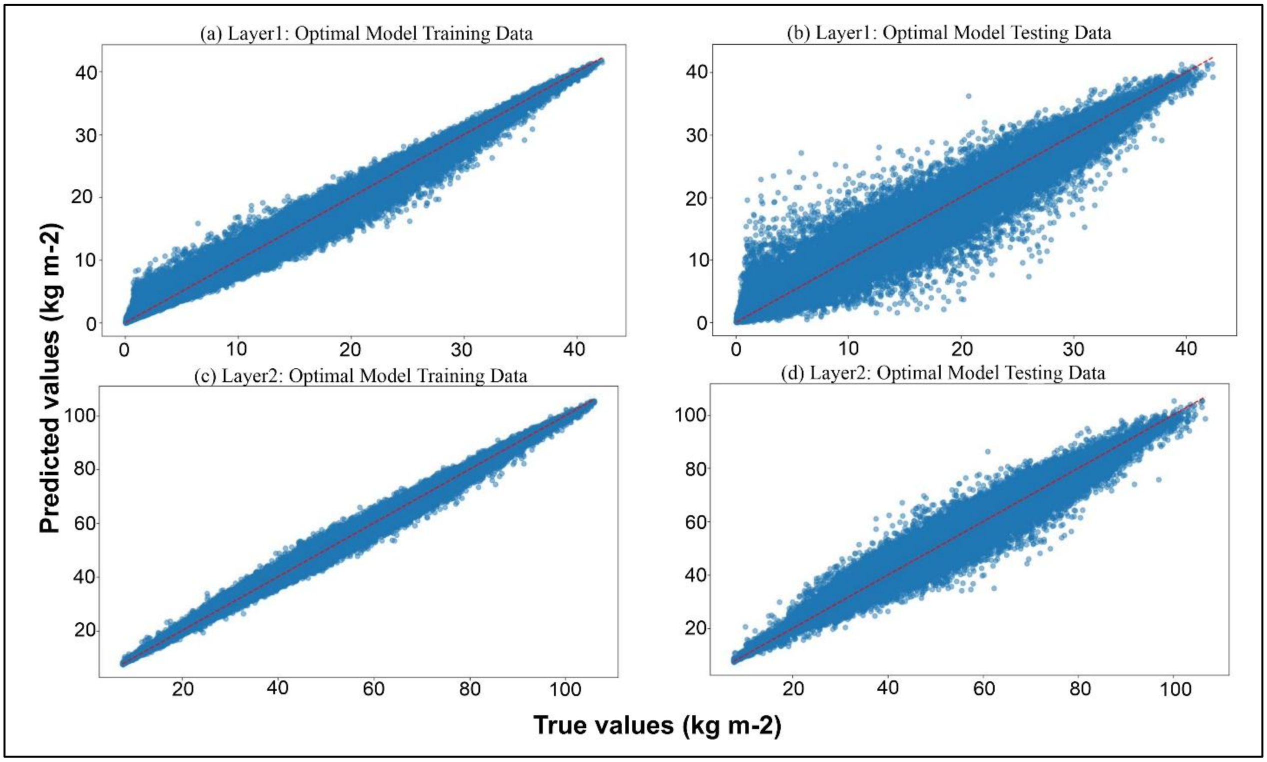

Figure 4. Scatter plots of observed vs predicted soil moisture for training and test sets (layers 1 and 2). *Validation period: 2013–2023; units: kg m⁻². The diagonal line shows 1:1 agreement.*

3.3 Model Performance

3.3.1 Evaluation of Regional Performance of the Model

The model's evaluation relies on Mean Absolute Error, Root Mean Square Error, Pearson's Correlation Coefficient (R), and the Coefficient of Determination (R^2). The MAE (1.144 kg/m²) and RMSE (1.755 kg/m²) for layer 1 are relatively low, with a high R (0.971) and R^2 (0.944). This suggests the model explains over 94% of surface moisture variance. Layer 2, which is the subsurface soil moisture content, shows higher MAE (1.824 kg/m²) and RMSE (2.698 kg/m²), but still maintains a strong R (0.977) and R^2 (0.952), capturing 95.2% of variability. While absolute errors increase with depth, the model demonstrates good performance across both layers.

Figures 5 and 6 show that the model's performance varies slightly across the country’s agro-ecological zones. Due to more consistent soil moisture levels, solid land features, and less precipitation variability, the northern Sahel and Sudan savanna zones show the lowest errors and highest correlation coefficients. On the other hand, performance declines with larger MAE and RMSE values in the southern coastal regions and the vicinity of Lake Chad. These variations can be attributed to the complexity introduced by water bodies, mixed land-water pixels, tidal influences, high precipitation, and dynamic river systems. The spatial discrepancies align with the understanding that surface moisture is more directly influenced by predictable meteorological factors such as rainfall and evaporation, which leads to lower prediction errors in these regions. Subsurface moisture is subject to more complex processes such as infiltration, groundwater interactions, and heterogeneous soil properties, increasing modelling difficulty (Torre *et al.*, 2019). However, the consistently high correlation values for Layer 2 indicate the model effectively captures spatial distribution and trends, even with larger absolute errors. This observation echoes findings in other studies, where models may reproduce temporal variations in wetter areas more effectively than in drier regions with lower precipitation variance (Koster *et al.*, 2009).

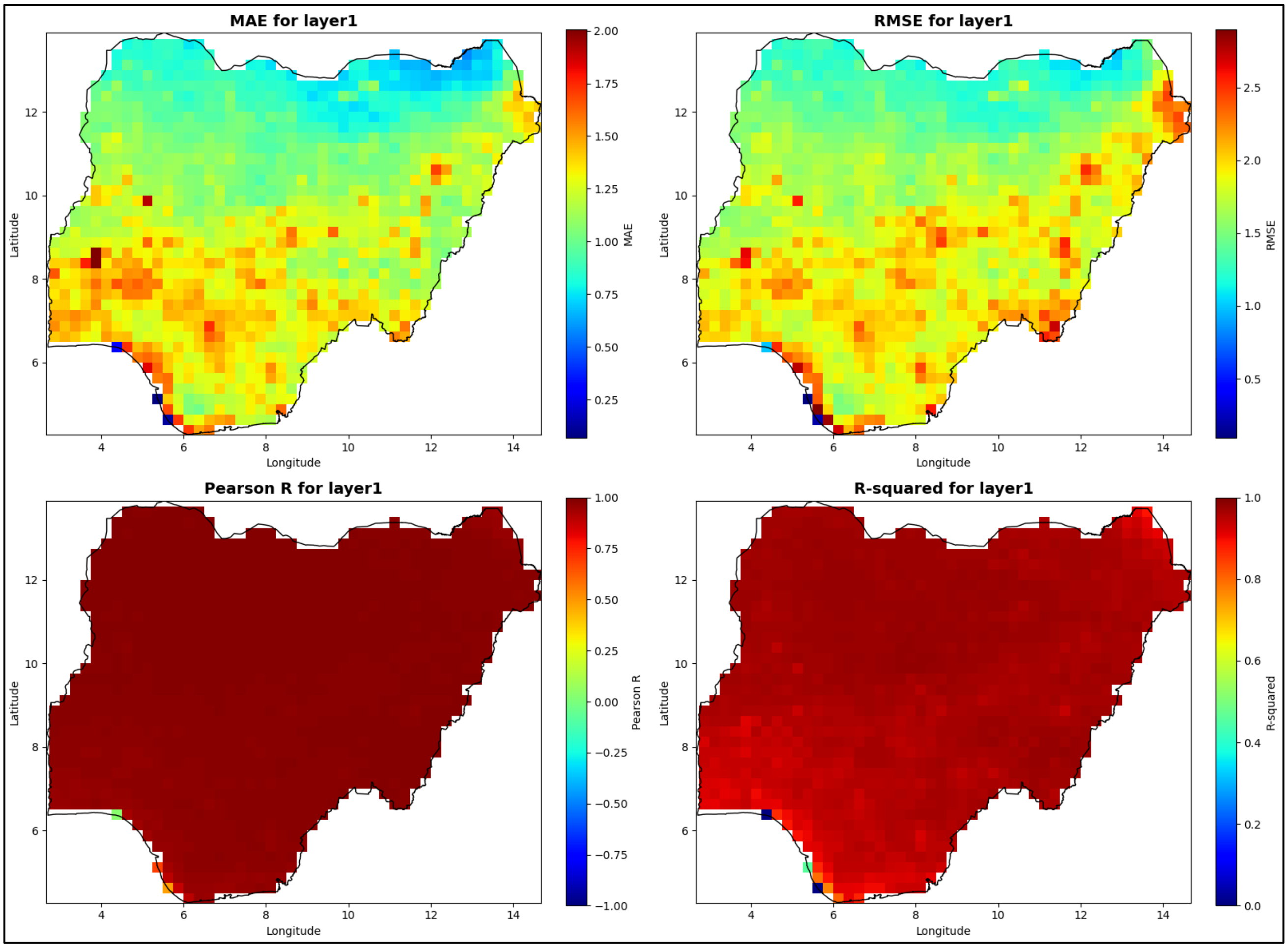

Figure 5. Spatial distribution of model validation metrics (MAE in kg/m², RMSE in kg/m², R, and R²) across Nigeria for Surface Soil Moisture (Layer 1) over the 2013–2023 validation period.

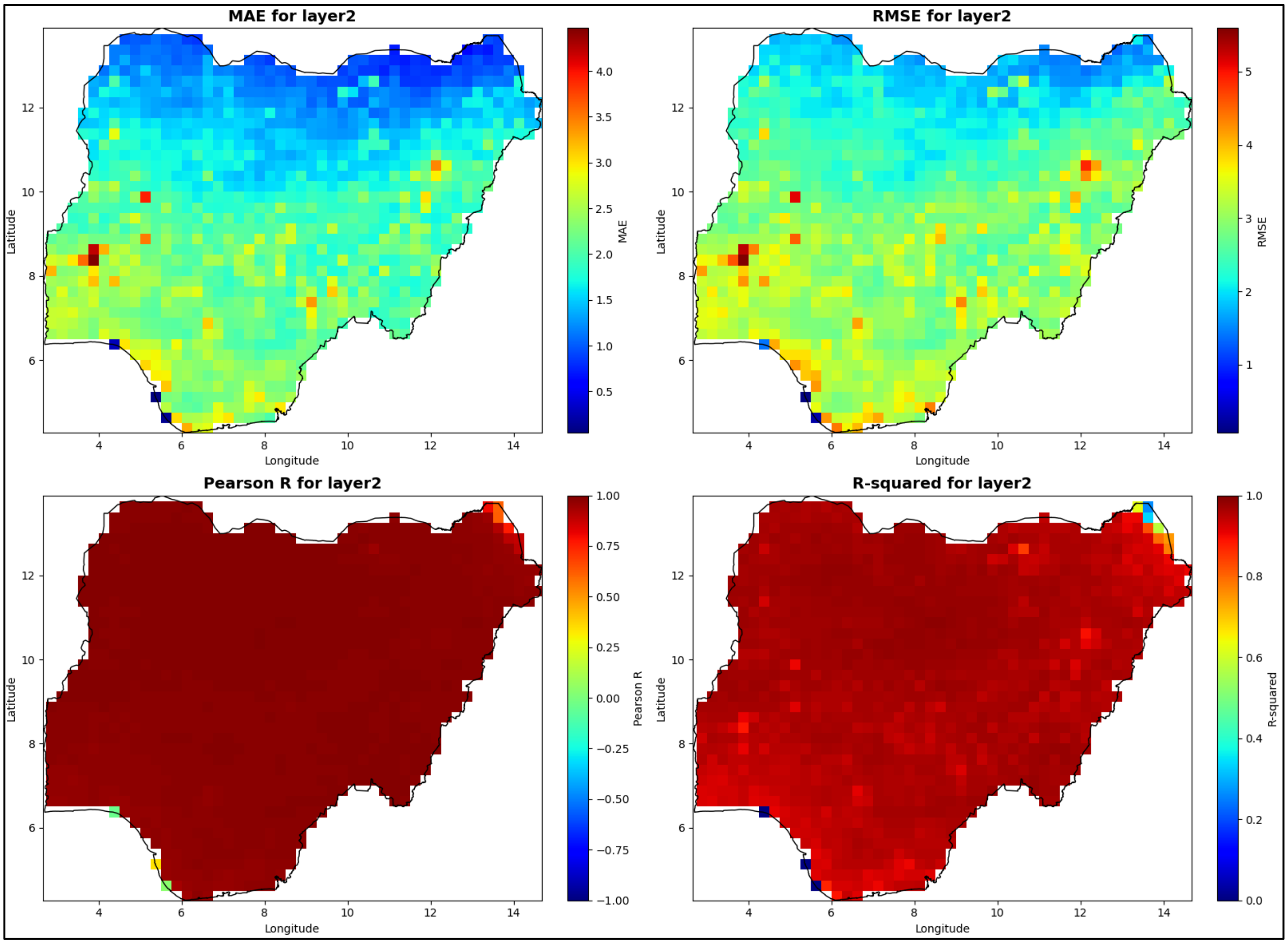

Figure 6. Spatial distribution of model validation metrics (MAE in kg/m², RMSE in kg/m², R, and R²) across Nigeria for Surface Soil Moisture (Layer 2) over the 2013–2023 validation period.

A detailed comparison between the two layers further underscores these spatial differences. For Layer 1, which represents surface soil moisture directly impacted by meteorological variables such as rainfall and evaporation, the model consistently shows lower error metrics. Because these factors typically exhibit predictable seasonal trends, the model captures their effects more effectively. However, additional factors, including soil infiltration, groundwater interactions, and heterogeneous soil qualities, affect the subsurface moisture in Layer 2, making it more challenging to precisely model. Despite larger absolute errors in Layer 2, the model's strong correlation and high R² values demonstrate that it reliably reproduces both the spatial distribution and spatiotemporal variations of subsurface soil moisture across Nigeria. Also, the accuracy of the northern region demonstrates how well the model works in stable hydrological circumstances. However, higher error metrics in coastal and water-dominated regions reveal limitations in capturing complex land-water interactions.

3.3.2 Seasonal Variability and Model Responsiveness

The model demonstrates strong temporal performance for both surface and subsurface soil moisture predictions from 2013 to 2023, used for independent validation as shown in Figures 7 and 8. Layer 1 shows stable MAE (1.085-1.194 kg/m²) and RMSE (1.739-1.931 kg/m²), with high R (0.986-0.990) and R² (0.973-0.979) values, capturing over 94% of the variance. Layer 2 exhibits slightly higher errors (MAE: 1.688-1.919 kg/m²; RMSE: 2.629-2.890 kg/m²) but retains excellent correlation (R: 0.991-0.993; R²: 0.981-0.985). This minimal bias over long timescales suggests good model calibration, despite the slightly increased error magnitudes in subsurface predictions due to inherent process complexity, a common challenge in soil moisture modelling (Han *et al.*, 2021).

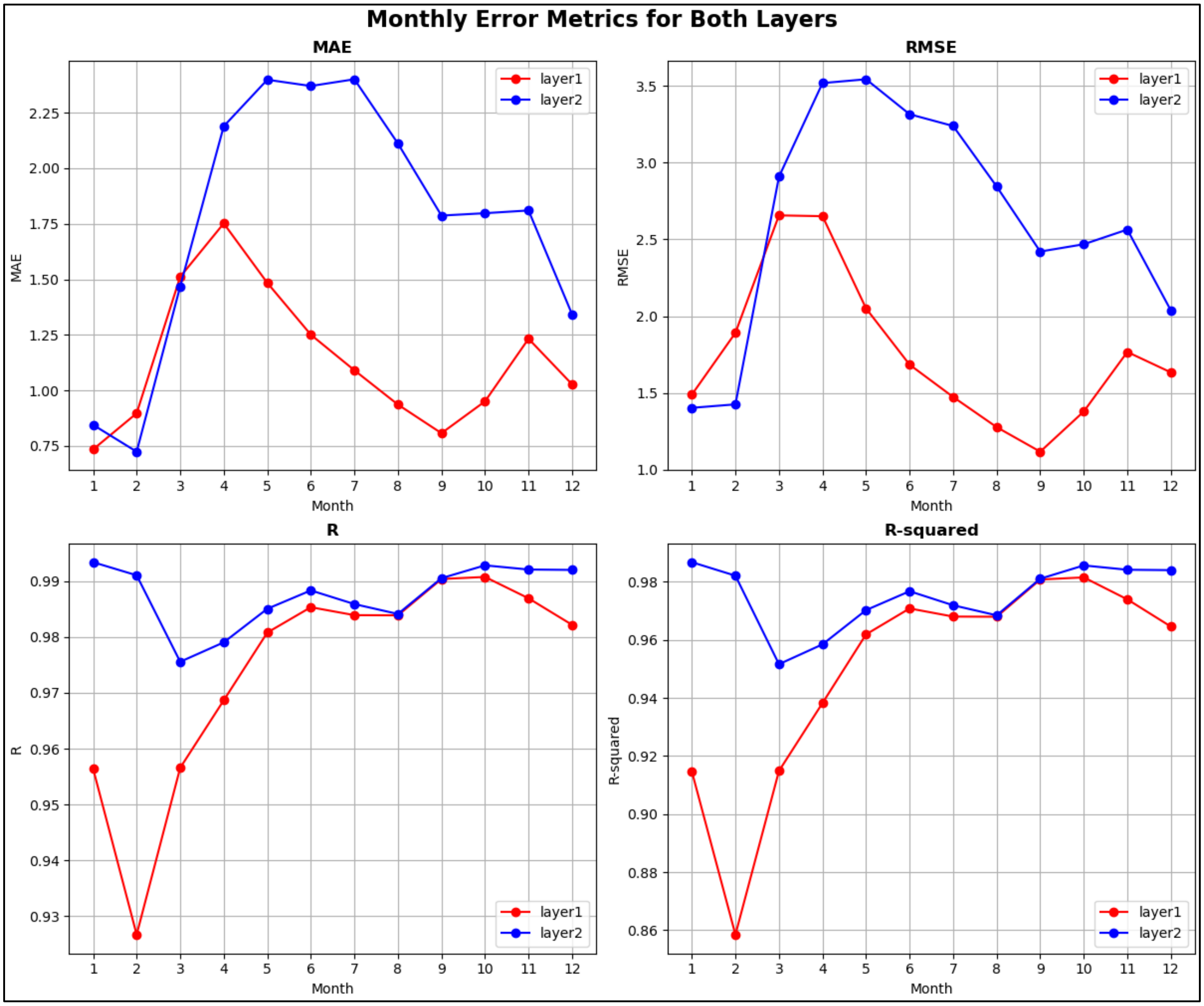

Figure 7. Monthly variation of model validation metrics (MAE, RMSE, R, and R²) across Nigeria, aggregated from 2013 to 2023, illustrating seasonal model responsiveness.

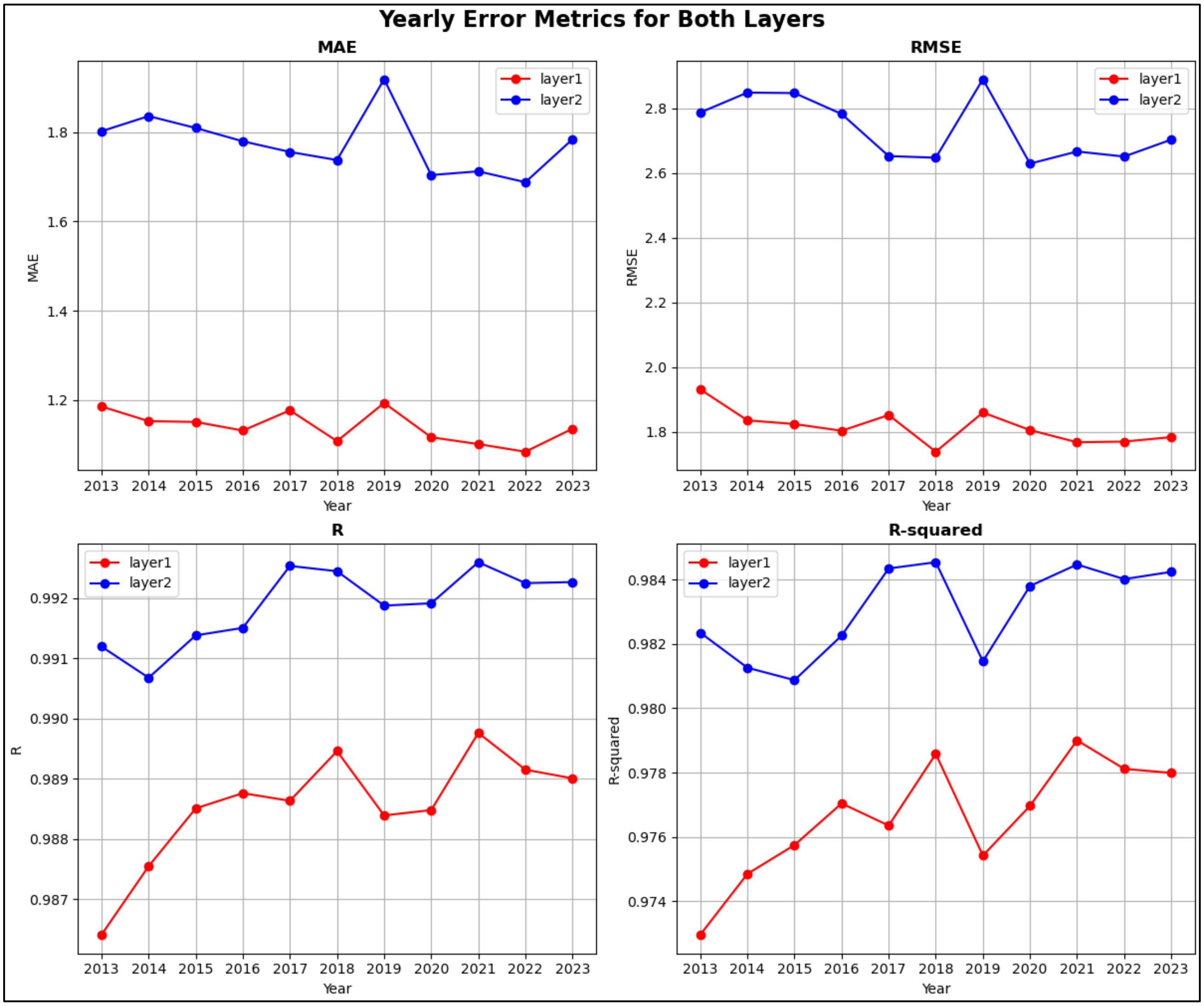

Figure 8. Annual model validation metrics across Nigeria (2013-2023)

The model's performance reflects Nigeria's seasonality. During the dry season, error metrics are lower, as seen in January's low MAE (0.735 kg/m²) and RMSE (1.489 kg/m²) for Layer 1. Errors increase in the early rainy season due to rapid moisture fluctuations. Notably, performance improves during peak rainy months, possibly due to stabilized moisture conditions despite high rainfall. A study on soil moisture variation in Nigeria (Fuwape *et al.*, 2019) also emphasizes the influence of seasonal factors. Layer 2 exhibits a similar seasonal pattern with higher absolute errors, performing best in the cooler, drier months and reaching peak errors in the early to mid-rainy season. This sensitivity to seasonal hydrological processes like infiltration and groundwater recharge is consistent with findings in (Liu & Zhu, 2013) which discusses groundwater's role in root zone moisture.

4. CONCLUSION

The present research contributes a robust soil moisture prediction model that has demonstrated strong and stable performance across Nigeria over 11 years. The model’s ability to accurately capture both interannual trends and seasonal fluctuations, as evidenced by consistently high correlation coefficients and R² values, confirms its effectiveness in representing the spatial and temporal dynamics of both surface and subsurface soil moisture. This capability is especially significant for agriculture and water management in Nigeria, where accurate soil moisture estimates can guide critical decisions such as irrigation scheduling, fertilizer application, and drought monitoring. For farmers, these insights translate into optimized crop management and enhanced water use efficiency, while policymakers can leverage the model's outputs to target interventions in water-stressed regions and develop strategies to improve food security. Key findings from the study indicate that while surface soil moisture is predicted with relatively lower error margins due to its direct responsiveness to meteorological variables, the subsurface layer poses greater challenges owing to complex hydrological processes such as infiltration and groundwater recharge. The model's performance in regions with stable, homogeneous conditions contrasts sharply with the higher uncertainties observed in coastal and water-influenced areas, where dynamic environmental factors and data limitations hinder prediction accuracy. Future research will focus on enhancing the model by incorporating additional hydrological parameters, including groundwater dynamics, topographical influences, and high-resolution vegetation indices, as well as integrating sensor-derived soil moisture products. Such refinements will be essential for improving predictive accuracy in areas with rapid moisture fluctuations and complex water–soil interactions. A potential limitation of this study is the reliance on TAMSAT satellite-derived soil moisture as both the primary training feature and the reference target. While TAMSAT has been validated for the region, this dependency may lead to error propagation from the original satellite product into the machine learning model. Future research should incorporate independent in-situ measurements from local meteorological stations to further validate the model's independence and absolute accuracy.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the Tropical Applications of Meteorology using Satellite for providing the Network Common Data Form (NetCDF) satellite data used to derive the soil moisture content in this study. The author also acknowledges financial support from the National Information Technology Development Agency under the Nigeria Artificial Intelligence Research Scheme.

REFERENCES

Adedolapo, K. A., & Ajetomobi, J. O. (2020). Climate Change Influence on Crop Yields and Crop Yield Variability in Nigeria. *Natural Science Edition, 16*(8), 113–121. https://www.xisdxjxsu.asia/V16-8-15.pdf

Babaeian, E., Sadeghi, M., Jones, S. B., Montzka, C., Vereecken, H., & Tuller, M. (2019). Ground, Proximal, and Satellite Remote Sensing of Soil Moisture. In *Reviews of Geophysics 57*(2), 530–616). Blackwell Publishing Ltd. https://doi.org/10.1029/2018RG000618

Bauer-Marschallinger, B., Cao, S., Navacchi, C., Freeman, V., Reuß, F., Geudtner, D., Rommen, B., Vega, F. C., Snoeij, P., Attema, E., Reimer, C., & Wagner, W. (2021). The normalised Sentinel-1 Global Backscatter Model Mapping Earth's land surface with C-band microwaves. *Scientific Data, 8*(1). https://doi.org/10.1038/S41597-021-01059-7

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

Ekwealor, K. U., Iroka, C. F., Ukpaka, G. C., Okeke, P. N., Okafor, P. N., & Okereke, K. E. (2020). An Investigation on the Forest and Savanna Vegetation Types in Nnamdi Azikiwe University, Awka Campus, in Anambra State of Nigeria. *Natural Resources and Conservation, 8*(1), 7–18. https://doi.org/10.13189/NRC.2020.080102

Entekhabi, D., Reichle, R. H., Koster, R. D., & Crow, W. T. (2010). Performance Metrics for Soil Moisture Retrievals and Application Requirements. *Journal of Hydrometeorology, 11*(3), 832–840. https://doi.org/10.1175/2010JHM1223.1

FAO. (2007). The world's mangroves 1980-2005. *FAO Forestry Paper*, 153, 89.

Fuwape, I. A., Ogunjo, S. T., & Owoola, E. O. (2019). Temporal variation of soil volumetric water content. *AIP Conference Proceedings, 2109*(1). https://doi.org/10.1063/1.5110131

Han, H., Choi, C. , Kim, J. , Morrison, R. R. , Jung, J. , & Kim, H. S. (2021). Multiple-Depth Soil Moisture Estimates Using Artificial Neural Network and Long Short-Term Memory Models. *Water, 13*(18), 2584, 13(18), 2584. https://doi.org/10.3390/W13182584

Hossain, M. R. H., & Kabir, M. A. (2023). Machine learning techniques for estimating soil moisture from smartphone captured images. *Agriculture, 13*(3), 574. https://doi.org/10.3390/agriculture13030574

Koster, R. D., Guo, Z., Yang, R., Dirmeyer, P. A., Mitchell, K., & Puma, M. J. (2009). On the nature of soil moisture in land surface models. *Journal of Climate, 22*(16), 4322–4335. https://doi.org/10.1175/2009JCLI2832.1

Liu, Y., & Zhu, Z. (2013). Calculating the flux at the bottom boundary of the root zone under shallow groundwater table based on SWAP model. In *Proceedings of the 2013 International Conference on Materials for Renewable Energy and Environment (ICMREE)*, 2, 617–620. https://doi.org/10.1109/ICMREE.2013.6893749

Martínez-de la Torre, A., Blyth, E. M., & Robinson, E. L. (2019). Evaluation of drydown processes in global land surface and hydrological models using flux tower evapotranspiration. *Water, 11*(2), 356. https://doi.org/10.3390/w11020356

Moses, M. (2025a). Machine learning-based calibration of satellite data for accurate soil moisture monitoring using the ISMN stations in Nigeria. *Nigerian Journal of Environmental Sciences and Technology, 9*(1), 125–136. https://nijest.com/wp-content/uploads/2025/04/125-136_20_Vol.-9-No.-1_NIJEST.pdf

Moses, M. (2025b). Mapping climatological trends in soil moisture variability across Nigeria. *Journal of Geomatics and Environmental Research,* 8(2), 125–141. https://doi.org/10.63745/joger.2025.12.30.012

Moses, M., Dodo, J. D., Ojigi, L. M., & Lawal, K. (2020). Regional TEC modelling over Africa using deep structured supervised neural network. *Geodesy and Geodynamics, 11*(5), 343–349. https://doi.org/10.1016/j.geog.2020.05.004

Numbere, A. O. (2018). Mangrove species distribution and composition, adaptive strategies and ecosystem services in the Niger River Delta, Nigeria. In *Mangrove ecosystem ecology and function*. IntechOpen. https://doi.org/10.5772/intechopen.79028

Ochsner, T. E., Cosh, M. H., Cuenca, R. H., Dorigo, W. A., Draper, C. S., Hagimoto, Y., Kerr, Y. H., Larson, K. M., Njoku, E. G., Small, E. E., & Zreda, M. (2013). State of the art in large-scale soil moisture monitoring. *Soil Science Society of America Journal,* 77(6), 1888–1919. https://doi.org/10.2136/sssaj2013.03.0093

Olasore, A. J., Olagbaiye, A. E., Ajayi, T. A., & Alabi, P. O. (2021). Drought monitoring in Northern Nigeria using four indices. *International Journal for Research in Applied Sciences and Biotechnology, 8*(1), 13–31. https://doi.org/10.31033/ijrasb.8.1.3

Osman, A. I. A., AlDahoul, N., Chong, K. L., Huang, Y. F., Ng, J. L., Elshafie, A., Sherif, M., & Ahmed, A. N. (2025). A review on machine learning models for drought monitoring and forecasting. *Climate Risk Management, 50*, 100758. https://doi.org/10.1016/j.crm.2025.100758

Pal, S., & Sharma, P. (2021). A review of machine learning applications in land surface modeling. *Earth, 2*(1), 174–190. https://doi.org/10.3390/earth2010011

Pallapothu, A. (2025). Development of Machine Learning model for Drought Prediction. *Journal of Student Research, 14*(1). https://doi.org/10.47611/jsrhs.v14i1.8737

Peng, J., Albergel, C., Balenzano, A., Brocca, L., Cartus, O., Cosh, M. H., Crow, W. T., Dabrowska-Zielinska, K., Dadson, S., Davidson, M. W. J., de Rosnay, P., Dorigo, W., Gruber, A., Hagemann, S., Hirschi, M., Kerr, Y. H., Lovergine, F., Mahecha, M. D., Marzahn, P., … Loew, A. (2021). A roadmap for high-resolution satellite soil moisture applications—Confronting product characteristics with user requirements. *Remote Sensing of Environment, 252*, 112162. https://doi.org/10.1016/j.rse.2020.112162

Probst, P., Wright, M. N., & Boulesteix, A. L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9*(3), e1301. https://doi.org/10.1002/widm.1301

Reddybattula, K. D., Nelapudi, L. S., Moses, M., Devanaboyina, V. R., Ali, M. A., Jamjareegulgarn, P., & Panda, S. K. (2022). Ionospheric TEC forecasting over an Indian low latitude location using long short-term memory (LSTM) deep learning network. *Universe, 8*(11), 562. https://doi.org/10.3390/universe8110562

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature, 566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Smirnov, A., Shprits, Y., Prol, F., Lühr, H., Berrendorf, M., Zhelavskaya, I., & Xiong, C. (2023). A novel neural network model of Earth's topside ionosphere. *Scientific Reports, 13*, Article 28034. https://doi.org/10.1038/s41598-023-28034-z

Srivastava, P. K., Pandey, V., Suman, S., Gupta, M., & Islam, T. (2016). Available data sets and satellites for terrestrial soil moisture estimation. In *Satellite soil moisture retrieval: Techniques and applications* (pp. 29–44). Elsevier. https://doi.org/10.1016/B978-0-12-803388-3.00002-4

World Bank Group. (2020). *Moving toward a middle-class society: Nigeria on the move—A journey to inclusive growth (Nigeria systematic country diagnostic)*. https://doi.org/10.1596/33347

Yusuf, D. O. A., Abdulhakeem, A., & Chatta, D. M. (2017). Floristic composition, vegetation structures and physiognomy of a typical Guinea savannah: A case study of Minna–Bida Road, Niger State. *International Journal of Biochemistry, Biophysics & Molecular Biology, 2*(4), 22–30. https://doi.org/10.11648/j.ijbbmb.20170204.11