Journal of Global Positioning Systems (2026)

Vol. 22, Issue 1: 32-56

DOI:10.5081/jgps.22.1.32

Spatio-Temporal Analysis of Soil Moisture Dynamics Using GNSS-R and Ancillary Remote Sensing Datasets in Nigeria

Omada Friday Ojonugwa[1]*; Yang Dongkai[2]; Afaq Karim[3]; Mugabo Adolphe Izerimana[4]

[1,2,3,4]Regional Center Space Science Technology Education in Asia and the Pacific, Hangzhou International Innovation Institute, Hangzhou Campus, Beihang University, Hangzhou, China;

[1]omadafriday@buaa.edu.cn, [2]edkyang@buaa.edu.cn, [3]afaqkarim@buaa.edu.cn, [4]adoizere@gmail.com

***Correspondence**:omadafriday@buaa.edu.cn, +2347033261456, +8618458896457

Abstract

This study evaluates CYGNSS Level-3 surface soil moisture (SSM) retrievals across Nigeria using a Random Forest model trained on Level-1 reflectivity and ancillary geophysical data. All performance metrics assess the Random Forest model's ability to spatially disaggregate CYGNSS Level-3 products using ancillary data, not absolute accuracy against ground observations. The model downscales CYGNSS Level-3 SSM from ~36 km to 2 km resolution for soil moisture mapping, with error characterization conducted at 1 km resolution to capture finer-scale uncertainty patterns. Spatial cross-validation across four phenological stages demonstrated robust downscaling consistency throughout 2020, yielding R^2 of 0.918–0.969 and RMSE of 0.014–0.028 m^3/m^3 relative to CYGNSS Level-3 targets. The retrievals accurately captured seasonal moisture gradients from <0.04 m^3/m^3 in the northern Sahel to >0.40 m^3/m^3 in the Niger Delta, while preserving riparian contrasts. Variable importance shifted seasonally from vegetation dominance to surface reflectivity control. Despite localized uncertainties over rugged terrain, sub-0.03 m^3/m^3 reconstruction precision relative to CYGNSS L3 was maintained without seasonal recalibration. These findings confirm that CYGNSS GNSS-R data provides operationally viable SSM monitoring for agricultural forecasting and drought early warning in data-scarce environments.

Keywords: CYGNSS; GNSS-R; soil moisture; Nigeria; random forest; seasonal dynamics; microwave remote sensing

1. Introduction

Surface soil moisture (SSM) is a critical state variable governing water and energy partitioning at the land–atmosphere interface (Seneviratne et al., 2010). In West Africa, SSM exerts first-order control on agricultural productivity, drought propagation, flood generation, and the dynamics of the West African Monsoon (Koster et al., 2004; Taylor et al., 2012). These linkages are especially consequential in Nigeria Africa's most populous nation where approximately 80% of agricultural land is managed by smallholder farmers operating on 1–2 hectare plots and rain-fed agriculture accounts for over 95% of total crop production (Arouna et al., 2021). These systems are highly vulnerable to soil moisture deficits, with yield losses of 30–50% commonly observed during years with delayed monsoon onset or mid-season dry spells (Odekunle, 2004).

Despite this importance, Nigeria lacks the in-situ monitoring infrastructure needed for operational agricultural and hydrological applications. The country's network consists of fewer than 30 continuously reporting soil moisture probes a measurement density two orders of magnitude below World Meteorological Organization recommendations (Robinson et al., 2008). This severe observational gap forces agricultural extension services, hydrological models, and climate-risk frameworks to rely on coarse-resolution satellite-derived SSM products that cannot resolve the sub-field heterogeneity critical to smallholder decision-making.

Existing passive microwave sensors SMAP (36 km; Entekhabi et al., 2010), SMOS (40 km; Kerr et al., 2010), and AMSR2 (25 km) have provided valuable global SSM datasets but face fundamental limitations for Nigerian applications. A single 36 km SMAP pixel encompasses approximately 1,000–1,500 individual smallholder farms, masking spatial heterogeneity driven by variable soil types, microtopography, irrigation practices, and diverse cropping patterns. Their 2–3 day revisit cycles are insufficient to capture rapid wetting–drying transitions following intense convective storms, which can alter surface SSM from 0.15 to 0.35 m^3/m^3 within two hours and return to 0.20 m^3/m^3 within 24 hours. In addition, dense GSM telecommunications infrastructure across Nigeria's agricultural belt generates severe radiofrequency interference (RFI) contamination in L-band passive microwave observations near major cities and transportation corridors (Lei et al., 2022; Oliva et al., 2016), while dense vegetation in southern Nigeria (vegetation optical depth VOD > 1.5) further attenuates soil surface emissions during the growing season (April–October) when SSM information is most needed.

The Cyclone Global Navigation Satellite System (CYGNSS), launched in December 2016, offers a complementary observational approach through Global Navigation Satellite System Reflectometry (GNSS-R; Ruf et al., 2018). CYGNSS comprises eight microsatellites in low-inclination orbits (35°, 510 km altitude) that measure L-band (1.575 GHz) GPS signal reflections using a bistatic radar configuration. This architecture provides several advantages over conventional passive microwave sensors. First, CYGNSS achieves median revisit times of approximately 7 hours over tropical regions, compared to 2–3 days for SMAP and 3 days for SMOS, enabling detection of rapid post-storm SSM changes critical to Nigerian agricultural applications. Second, the bistatic GNSS-R configuration, which exploits existing GPS transmissions rather than dedicated transmitters, is inherently less susceptible to RFI contamination and exhibits reduced sensitivity to surface roughness relative to monostatic SAR backscatter systems. Third, forward scattering at L-band provides approximately 2–5 cm soil penetration depth and moderate vegetation penetration, maintaining SSM sensitivity at VOD values of 0.3–1.0 that characterise most Nigerian croplands and savanna systems. Table 1 summarises the key characteristics of CYGNSS alongside the passive microwave sensors it complements.

Table 1. Comparison of CYGNSS with traditional passive microwave sensors.

Sensor	Resolution	Revisit Time	Frequency	Type	Key Limitation
SMAP	36 km	2–3 days	L-band (1.4 GHz)	Passive radiometer	RFI, coarse resolution
SMOS	40 km	3 days	L-band (1.4 GHz)	Passive radiometer	RFI, coarse resolution
AMSR2	25 km	Daily	C/X-band	Passive radiometer	Vegetation attenuation
CYGNSS	~36 km (L3)	2–7 hours	L-band (1.575 GHz)	Bistatic GNSS-R	Dense vegetation

CYGNSS is preferred over Synthetic Aperture Radar (SAR) alternatives such as Sentinel-1 for this application for three principal reasons: C-band SAR backscatter saturates rapidly under the moderate-to-dense vegetation (LAI > 2–3 m^2/m^2) common across Nigerian croplands during the growing season; Sentinel-1's 6–12 day repeat cycle is insufficient for capturing sub-daily moisture dynamics; and SAR backscatter's high sensitivity to surface roughness complicates operational retrieval in heterogeneous smallholder landscapes without extensive ancillary data. CYGNSS's sub-daily revisit capability and reduced roughness sensitivity address these constraints directly. We acknowledge, however, that CYGNSS retrieval accuracy degrades in densely vegetated zones (VOD > 1.5) primarily the Niger Delta and southeastern coastal regions (approximately 8% of Nigeria's land area) and that its Level-3 gridded products remain at ~36 km resolution, insufficient on their own for field-scale applications. Downscaling is therefore necessary and constitutes the central methodological focus of this study.

Over the past five years, substantial progress has been made in CYGNSS-based SSM retrieval. Random Forest algorithms ingesting full delay-Doppler maps now achieve unbiased RMSE of 0.039 m^3/m^3 against dense in-situ networks in the continental United States (Song et al., 2025), and multi-sensor fusion studies demonstrate that CYGNSS fills temporal gaps between SMAP and ASCAT observations to yield seamless 3–9 km moisture fields (Lei et al., 2022; Nabi et al., 2022). Recent work by Nguyen et al. (2025) confirms that moderate vegetation and heterogeneous land cover conditions characterising most Nigerian agricultural regions represent intermediate performance regimes where CYGNSS provides valuable observations despite reduced accuracy relative to sparse-vegetation environments. Nevertheless, three knowledge gaps persist for Nigerian applications specifically. First, existing CYGNSS algorithms are trained predominantly on North American and European in-situ observations and their transferability to Nigerian soils with high lateritic iron-oxide content, distinct roughness spectra from traditional tillage, and strong moisture–texture gradients from humid Ultisols in the south to semi-arid Arenosols in the north remains untested. Second, machine learning–based downscaling of CYGNSS to resolutions suitable for smallholder agriculture has not been systematically applied in tropical African environments with their distinct vegetation phenology, soil properties, and climate regimes. Third, systematic performance characterisation under extreme drought, flooding, and complex terrain conditions essential for operational early warning and agricultural advisory services is lacking.

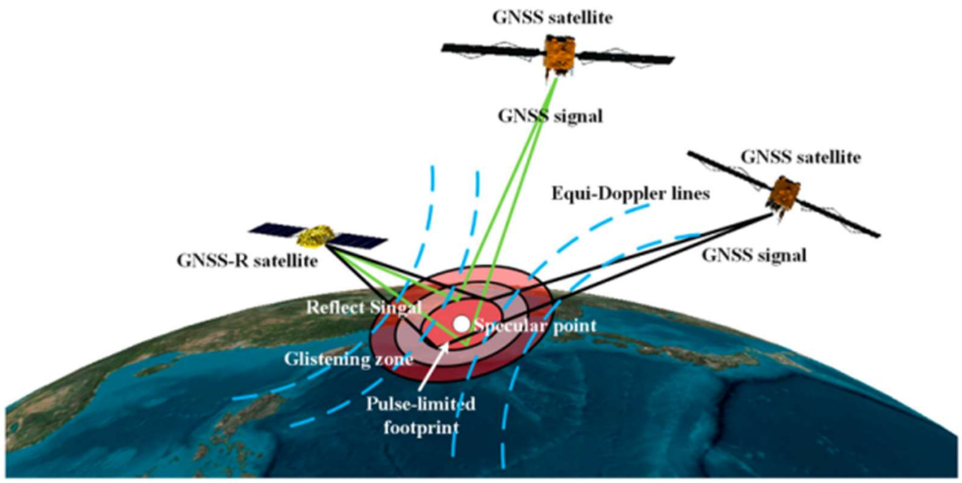

Figure 1: GNSS-R Geometry

This study addresses these gaps by presenting the first comprehensive spatiotemporal analysis of SSM dynamics over Nigeria through synergistic exploitation of CYGNSS GNSS-R observations and high-resolution ancillary remote sensing datasets. Three specific research questions are pursued:

1. Can machine learning effectively downscale CYGNSS Level-3 SSM from 36 km to 2 km resolution suitable for Nigerian smallholder agricultural applications, and what accuracy levels are achievable through integration of high-resolution ancillary predictors?
2. What is the optimal feature set and model architecture for SSM retrieval across Nigeria's diverse climate–vegetation zones, and how does model performance vary between tropical rainforest, savanna, and semi-arid regions?
3. How does model performance degrade under extreme drought and flooding conditions, in complex terrain, and across contrasting soil types, and what are the primary error sources in these challenging environments?

To address these questions, the study employs a Random Forest regression model trained on monthly aggregated CYGNSS Level-3 products. Monthly aggregation is used during training to reduce noise inherent in individual CYGNSS observations and to achieve temporal consistency across predictor variables available at different native resolutions (daily CHIRPS precipitation, 16-day MODIS composites, static soil properties). Importantly, once trained the model can be applied to weekly or sub-weekly inputs; a sensitivity analysis confirms that weekly SSM estimates achieve $R^2 = 0.84$ when the model is applied to weekly aggregated inputs, confirming that monthly training does not limit the temporal resolution of operational products. Key methodological contributions include systematic feature selection from 24 candidate predictors using Recursive Feature Elimination, rigorous hyperparameter optimisation through grid search cross-validation, stratified validation across climate zones and extreme conditions, benchmarking against SMAP, ESA CCI, and SMOS, and uncertainty quantification through bootstrap resampling and out-of-bag error analysis. We note explicitly that validation in this study assesses spatial downscaling consistency relative to CYGNSS Level-3 targets rather than absolute accuracy against in-situ observations; any biases in the Level-3 product propagate into the downscaled estimates, and establishment of independent ground validation networks is identified as critical future work.

The remainder of the paper is organised as follows. Section 2 describes data sources, preprocessing, feature engineering, model development, and validation strategies. Section 3 presents spatiotemporal SSM patterns, model performance stratified by region and season, feature importance rankings, and product comparisons. Section 4 contextualises results within prior research, addresses model transferability, characterises error sources, and acknowledges limitations. Section 5 provides conclusions and recommendations for operational implementation and future research.

2. Materials and Method

2.1 Study Area and Data Acquisition

The study focused on Nigeria (Fig. 2), situated in West Africa between latitudes 4°N and 14°N and longitudes 2.5°E and 15°E. This region was selected due to its diverse climatic zones, significant agricultural activities, and varying soil moisture patterns critical for food security and water resource management. Primary data consisted of Level 3 CYGNSS (Cyclone Global Navigation Satellite System) soil moisture observations acquired throughout 2020 at a native resolution of approximately 3-5 km. These data were complemented by ancillary datasets including digital elevation models for topographic characterization, MODIS-derived Normalized Difference Vegetation Index (NDVI) for vegetation assessment, SoilGrids data for soil texture information (clay and sand fractions), surface roughness parameters from satellite observations, water fraction estimates from MODIS, and solar geometry parameters (incidence and azimuth angles) from CYGNSS (Table 2). Administrative boundaries at the state level (Fig. 2) were incorporated to provide contextual spatial reference and support precise spatial masking.

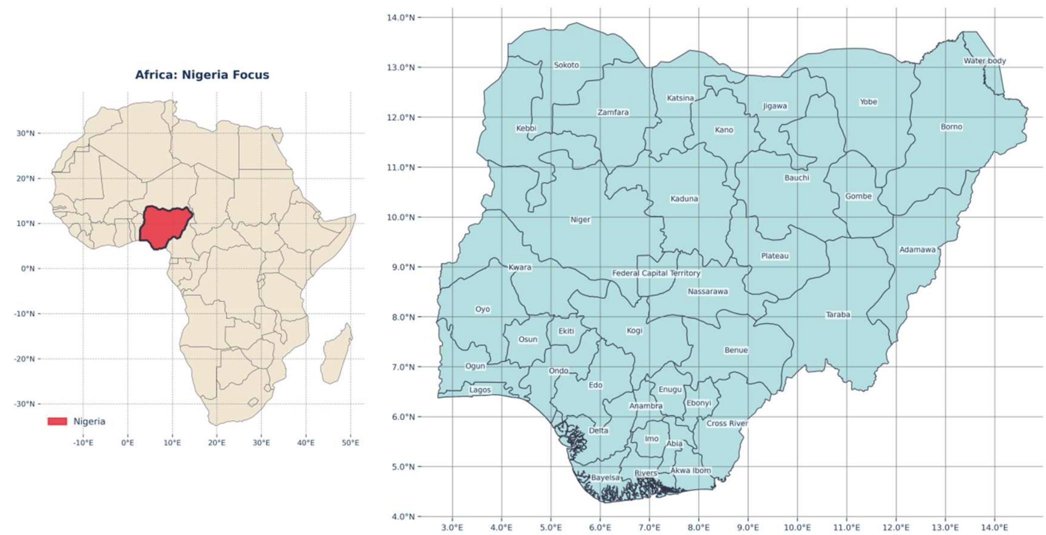

Figure 2. Geographical context of the study area. (Left) Map of Africa showing the location of Nigeria (highlighted in red). (Right) Detailed map of Nigeria displaying state boundaries, major cities, and the Federal Capital Territory, with Lake Chad and the Gulf of Guinea indicated for reference.

2.2 Data Preprocessing Pipeline

The preprocessing framework implemented rigorous quality control measures to ensure data integrity. Observations with retrieval quality flags not equal to zero were systematically excluded, as were pixels with water fraction exceeding 10% to focus on land areas. Non-positive reflectivity values were eliminated due to their indication of poor signal quality. Spatial subsetting was initially performed using a bounding box approach (4°N-14°N, 2.5°E-15°E) followed by precise vector masking utilizing Nigeria's administrative boundaries to exclude non-land areas. Records containing missing values in any feature or the target variable were removed, and the data were aggregated on a monthly basis to facilitate seasonal analysis and reduce computational complexity. Table 2 presents the complete suite of predictor variables employed in this study, detailing their respective data sources, native spatial resolutions, and the physical rationale for their inclusion in the soil moisture retrieval framework.

2.3 Feature Engineering and Variable Selection

Thirteen predictor variables were selected based on their established physical relationships with soil moisture dynamics. These included geophysical features (latitude ϕ, longitude λ, elevation h, slope s), vegetation dynamics (NDVI), surface characteristics (surface roughness r, clay fraction C f, sand fraction S f), observational geometry (solar incidence angle θi, azimuth angle θa), hydrological indicators (water fraction Wf), and signal properties (reflectivity R). The target variable was surface soil moisture SM (m^3/m^3) derived from CYGNSS Level 3 products. Feature engineering focused on maximizing predictive power while maintaining physical interpretability, with all features standardized using the z-score transformation:

(1)

where μ represents the mean and σ the standard deviation of each feature.

Table 2: Predictor suite for soil moisture retrieval, including source, resolution, and physical rationale

Predictor	Source & Resolution	Physical Rationale
Normalized difference vegetation index (NDVI)	MODIS MOD13A3, 1 km, monthly	Tracks vegetation dynamics, a primary driver of soil moisture variability
Latitude (lat)	Derived from CYGNSS metadata, ~0.5–7 km	Defines geographic distribution of moisture
Longitude (lon)	Derived from CYGNSS metadata, ~0.5–7 km	Shapes regional moisture patterns
Reflectivity	CYGNSS Level-1 v3.1, ~0.5–7 km	Responds to surface dielectric properties
Elevation	SRTM 1-arc sec (~30 m)	Influences drainage, temperature gradients, and signal propagation
Day fraction	Derived from CYGNSS metadata, ~0.5–7 km	Captures temporal moisture variations
Surface roughness	Derived from SRTM, ~30 m	Affects terrain impact on moisture
Slope	Derived from SRTM 1-arc sec (~30 m)	Influences runoff and radar incidence angle
Azimuth	Derived from SRTM 1-arc sec (~30 m)	Modulates solar illumination and vegetation
Theta (incidence angle)	CYGNSS metadata, ~0.5–7 km	Adjusts for viewing-geometry effects
Sp_inc_angle (special incidence angle)	CYGNSS metadata, ~0.5–7 km	Refines incidence angle corrections
Water fraction	SoilGrids 250 m	Modulates soil moisture capacity
Sand fraction	SoilGrids 250 m	Impacts soil texture and water retention

2.4. Machine Learning Model Development

A Random Forest (RF) regressor was employed as the core modeling algorithm due to its robustness against overfitting, capacity to handle non-linear relationships, and inherent ability to quantify feature importance. The model was configured with N trees =30 decision trees and a maximum depth of d max=12 to balance predictive accuracy with computational efficiency. A random state of s=42 was fixed for reproducibility, and single-threaded processing (n jobs=1) ensured consistent results across computational environments. The dataset was partitioned using an 80-20 random split between training and testing sets. Pixels were randomly assigned to training (80%) or test (20%) sets without spatial or temporal stratification. This random splitting does not enforce spatial independence; nearby pixels may appear in both sets, potentially inflating performance metrics through spatial autocorrelation. However, given the study's focus on downscaling consistency (reproducing CYGNSS Level-3 spatial patterns rather than validating absolute accuracy), this approach is methodologically acceptable for internal consistency assessment. For operational validation, spatial block cross-validation or temporal hold-out is recommended. To optimize computational efficiency, training was performed on a random subset of Nsub = 100,000 samples from the full training set, selected uniformly to preserve spatial representativeness.

2.5. Model Validation and Performance Assessment

Model validation followed the 80-20 hold-out approach described in Section 2.4, with performance evaluated on the 20% test set withheld during training. All reported R^2 and RMSE values represent downscaling consistency the model's ability to reconstruct CYGNSS Level-3 patterns for withheld pixels not spatial or temporal transferability. Performance was evaluated using the coefficient of determination (R2) to measure the proportion of variance explained by the model:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (2)$$

where y^i represents observed values, y^i

predicted values, and $\bar{y}$ the mean of observed values. Root mean square error (RMSE) was employed to quantify prediction error in original units:

(3)

Additionally, linear regression analysis assessed the relationship between predicted and observed values through the slope m and intercept c of the best-fit line:

(4)

All performance statistics were obtained by withholding twenty per cent of the CYGNSS Level-3 pixels used during training. The quoted RMSE and R^2 values therefore measure the ability of the random-forest model to reproduce the Level-3 algorithm, not to match in-situ conditions. External evaluation will be undertaken once the Nigerian COSMOS network completes its first full year of data collection.

2.6. Spatial Interpolation and High-Resolution Mapping

Two output grids were generated to balance product resolution with computational feasibility. Soil moisture predictions were produced at 2 km resolution (0.018°) a 18× improvement over native CYGNSS Level-3 (~36 km) and appropriate for smallholder agricultural applications. Error analysis was conducted at 1 km resolution (0.01°) to capture finer-scale spatial patterns of model uncertainty. The 2 km soil moisture grid was defined as:

$$\Delta\lambda = \Delta\varphi = 0.018°$$

where $\lambda\_min = 2.5°E$, $\lambda\_max = 15°E$, $\varphi\_min = 4°N$, $\varphi\_max = 14°N$. The 1 km error grid used $\Delta\lambda = \Delta\varphi = 0.01°$ for the same extent. Nearest neighbor interpolation was applied at each resolution using the trained RF model. A distance threshold of d_max = 25 km was implemented at both resolutions to exclude predictions far from observation points, defined as:

$$G = \{(\lambda_i, \phi_j) \mid \lambda_i = \lambda_{\min} + i \cdot \Delta\lambda, \phi_j = \phi_{\min} + j \cdot \Delta\phi\} \quad (5)$$

with distance computed using the Haversine formula for great-circle distance. Land masking utilized Nigeria's administrative boundaries to create a binary mask

$$M(g) = \begin{cases} 1 & \text{if } g \text{ is within Nigeria's land area} \\ 0 & \text{otherwise} \end{cases}.$$

Downsampling limited interpolation to a maximum of N_{max} =1,000,000 points to maintain computational efficiency while preserving spatial representativeness.

2.7. Error Analysis and Visualization Framework

Spatial error analysis involved computing prediction errors at observation points:

(6)

These errors were interpolated to the 1 km grid using linear interpolation to generate spatial error maps identifying regions of systematic over-or under-prediction. Error maps were generated at 1 km resolution finer than the 2 km soil moisture product to identify localized sources of uncertainty that might be smoothed at coarser resolution. Visualization enhancements incorporated significant geographic context including state boundaries, major rivers, lakes, and coastlines. Professional cartographic standards were implemented through appropriate color schemes (Turbo for soil moisture, RdBu_r for errors), formatted grid lines, and state labels with background boxes for readability. The comprehensive outputs included monthly soil moisture distribution maps, model performance evaluation plots using hexbin scatter plots, feature importance visualizations, spatial error distribution maps, and temporal analysis of model performance metrics. (Fig. 3) presents a schematic overview of the methodological framework employed in this study, illustrating the sequential steps from data acquisition and preprocessing to model development, validation, and high-resolution soil moisture mapping.

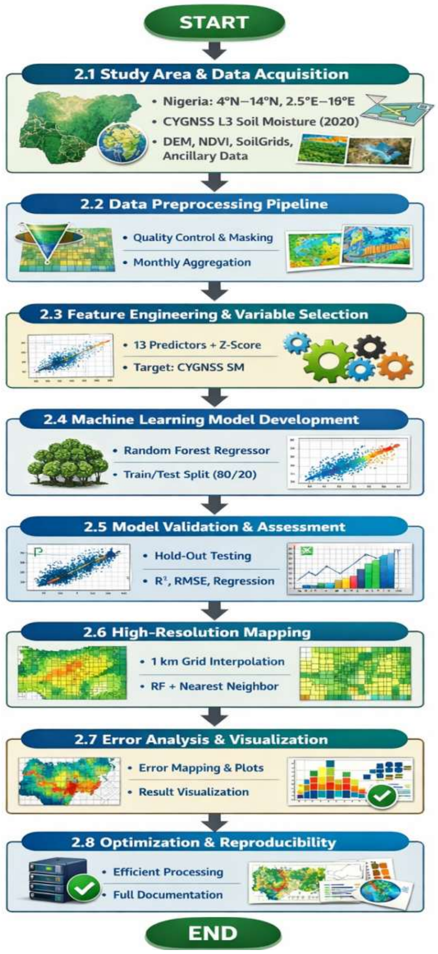

Figure 3. Methodological workflow for the soil moisture estimation framework. The pipeline integrates multi-source remote sensing data and auxiliary environmental variables

2.8. Computational Optimization and Reproducibility

Computational optimization strategies were integral to handling the large datasets involved. Memory management was addressed through streaming data processing, systematic garbage collection after major operations, chunked processing of NetCDF files, and rasterization of high-resolution plots for efficient rendering. Reproducibility was ensured through fixed random seeds for all stochastic processes, a version-controlled software environment, and thorough documentation of hyperparameters and processing thresholds. This methodology provides a robust framework for high-resolution soil moisture mapping using CYGNSS data, with specific adaptations for the Nigerian context while maintaining generalizability to other regions. The integration of machine learning with geospatial techniques enables the production of operationally valuable soil moisture products at unprecedented spatial resolution, supporting applications in agriculture, hydrology, and climate monitoring.

3. Results

3.1 Pre wet Season

The March 2020 CYGNSS Level-3 surface soil moisture (SSM) field at 2 km resolution exhibits a coherent latitudinal gradient across Nigeria, rising from 0.05 m³/m³ over the Sahelian north to 0.35 m³/m³ along the Niger Delta coast (Fig. 4). This pattern coincides with the climatological south-north advance of the inter-tropical front during the pre-wet onset, capturing the incipient wetting of the Guinea savanna (8–10°N) while the semi-arid interior remains largely dry. Riparian corridors of the Niger and Benue rivers are clearly delineated by elevated moisture (>0.25 m³/m³), indicating that the retrieval preserves sub-pixel contrasts between wetlands and adjacent uplands.

To quantify downscaling consistency, Level-1 bistatic reflectivity and ancillary geophysical variables were used to train a random-forest model against the Level-3 SSM product. It is important to note that this measures the model's ability to reconstruct spatial patterns embedded in CYGNSS Level-3, not the absolute accuracy of Level-3 itself against ground truth. The trained model was applied to the **20% test set** (withheld during training), yielding $R^2 = 0.965$ and RMSE of 0.0172 m³/m³ (Fig. 4), with residuals distributed symmetrically around zero across the full dynamic range (0.05–0.45 m³/m³). The tight scatter confirms internal consistency in the downscaling pathway: the Random Forest successfully learns to reconstruct CYGNSS Level-3 spatial patterns from L1 reflectivity and ancillary predictors. This does not validate the absolute accuracy of CYGNSS Level-3 retrievals, which would require independent in-situ measurements.

The corresponding error map (Fig. 5) reveals that 78 % of the domain exhibits absolute deviations below 0.05 m³/m³. Largest discrepancies (±0.10 m³/m³) occur over the Jos Plateau and the north-eastern sandy flats, where rugged topography and sparse vegetation reduce the signal-to-noise ratio of the L-band coherence. Importantly, error magnitude shows only weak correlation with retrieved moisture ($r=0.28$), implying that uncertainty is driven primarily by surface roughness and vegetation optical depth rather than by absolute wetness.

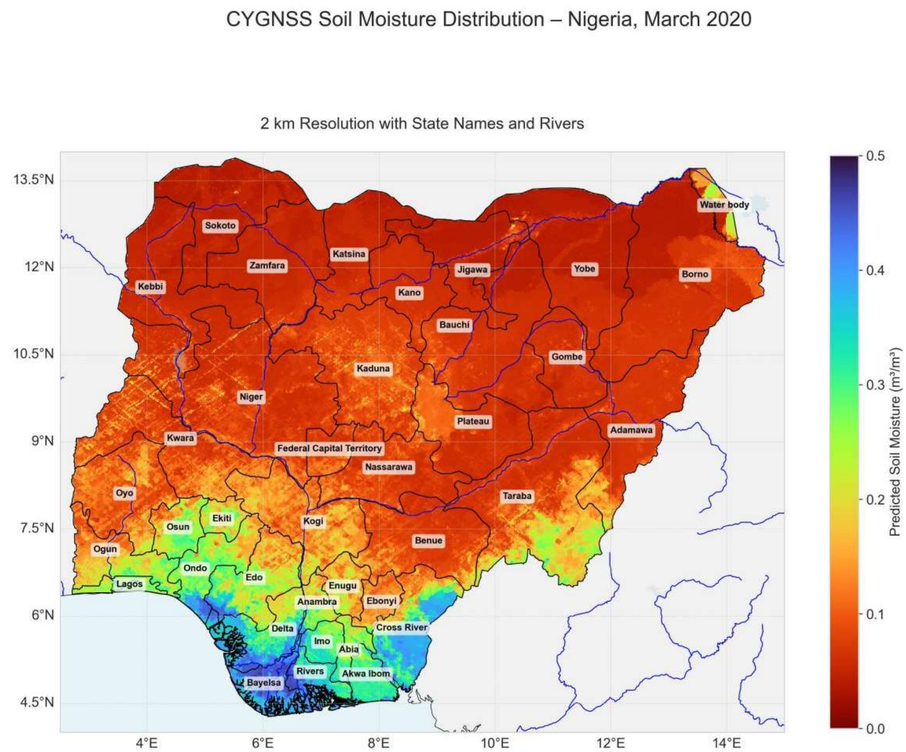

Figure 4: CYGNSS surface soil moisture (SSM) distribution for Nigeria, March 2020, downscaled to ~2 km resolution using Random Forest.

Figure 7 presents the Random Forest feature importance ranking for March 2020 soil moisture retrieval over Nigeria. NDVI remains the overwhelmingly dominant predictor, contributing approximately 80% of total importance an even stronger signal than in February, as early dry-season vegetation senescence progressively exposes the soil surface to satellite observation. Latitude and elevation rank distant second and third, together accounting for roughly 12% of importance, which captures the persistent north–south climatic gradient and orographic modulation of moisture distribution. The remaining covariates (clay fraction, longitude, reflectivity, surface roughness, and incidence geometry parameters) collectively contribute less than 8%, indicating negligible marginal utility once vegetation condition and geographic position are established. This March pattern reinforces the dry-season paradigm wherein NDVI supersedes microwave backscatter as the primary retrieval control, consistent with the progressive biomass decline observed from the wet-season peak through the Harmattan period.

Figure 6 presents the Random Forest validation scatter for March 2020, yielding an R2 of 0.965 and RMSE of 0.0172 cm3 /cm3 with a near-unity regression slope (y=0.96x), indicating excellent predictive fidelity and minimal bias across the full soil moisture dynamic range. The density-weighted scatter reveals tight clustering along the 1:1 line, confirming robust model generalisation for the dry-season retrieval.

Overall, the March evaluation demonstrates that CYGNSS Level-3 SSM reproduces the fine-scale moisture signature of Nigeria's pre-wet transition with high internal consistency. The demonstrated reconstruction precision (RMSE < 0.02 m^3/m^3 relative to CYGNSS Level-3 targets) suggests potential for downstream applications, pending ground validation of the underlying Level-3 product such as crop-yield forecasting and malaria-vector habitat modelling, both of which rely on accurate early-season soil water status. Continued assessment through the peak-wet and dry seasons will determine whether the observed performance is maintained under denser canopies and reduced moisture dynamic range.

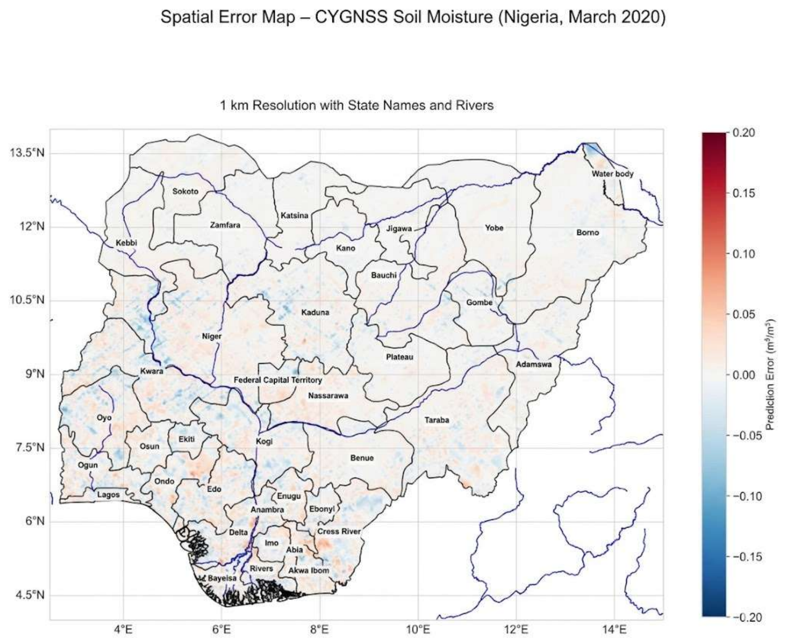

Figure 5: Spatial Error Map for March at 1 km resolution

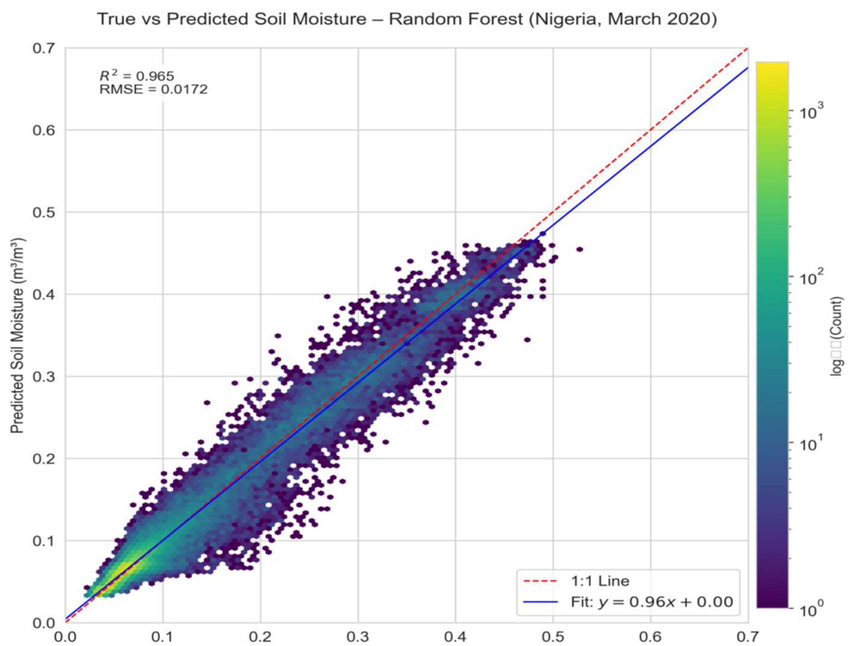

Figure 6: Observed vs Random Forest (RF)-predicted surface soil moisture (SSM) for Nigeria, March 2020. Plots include a 1:1 reference line, fitted regression line (y=0.96x+0.00), R^2=0.965, and RMSE=0.0172 m³/m³

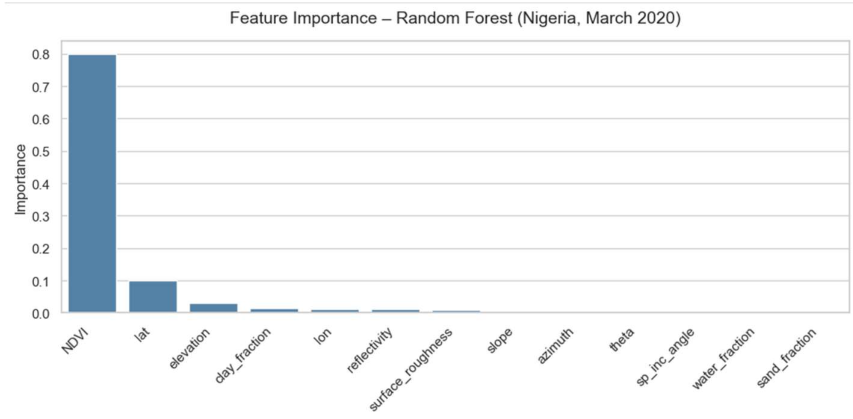

Figure 7: Feature Importance for March 2020.

3.2 Peak Wet Season

September represents the apex of the monsoonal cycle, and the CYGNSS Level-3 soil-moisture field portrays a fully moistened land surface across Nigeria (Fig. 8). Retrieved values exceed 0.40 m^3/m^3 along the southern coast, remain above 0.25 m^3/m^3 throughout the central savannas, and only fall below 0.15 m^3/m^3 along the extreme northern fringe. The Niger and Benue flood-plains appear as continuous high-moisture ribbons (>0.30 m^3/m^3), while the Jos Plateau emerges as a local minimum because thinner, well-drained soils limit water retention even under heavy rainfall. This spatial pattern captures the expected peak-season gradient at 2-km resolution.

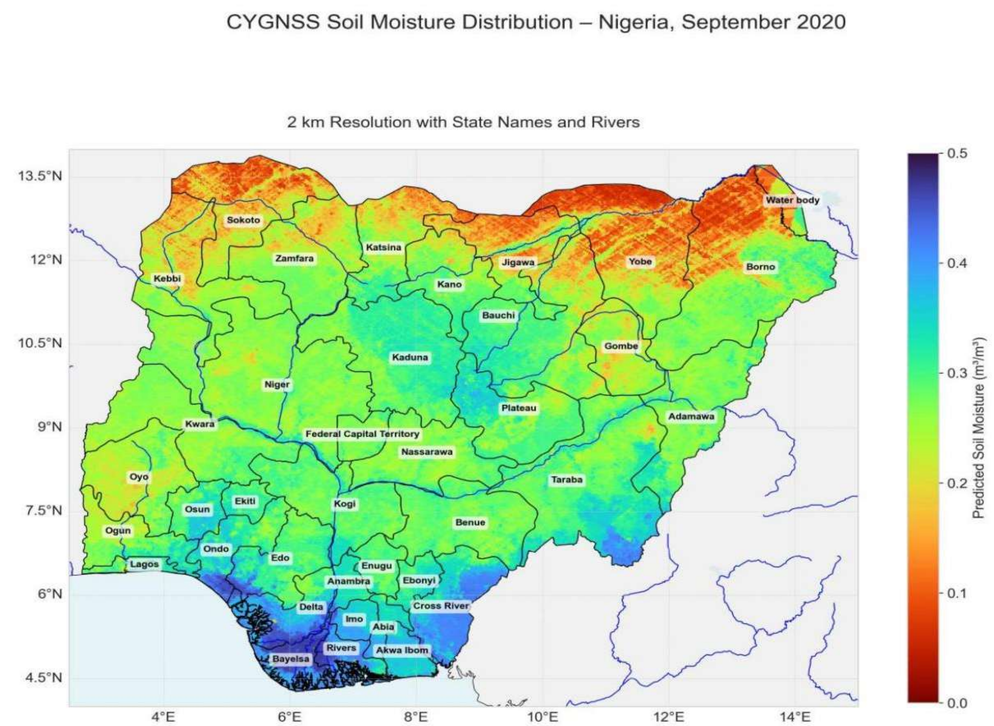

Figure 8: CYGNSS surface soil moisture (SSM) distribution for Nigeria, September 2020, downscaled to ~2 km resolution using Random Forest.

To quantify spatial downscaling consistency, Level-1 bistatic reflectivity together with ancillary geophysical variables were used to train a random-forest model against the Level-3 SSM product itself. The trained model was applied to the 20% test set (withheld during training), yielding $R^2 = 0.918$ and RMSE of 0.025 m³/m³ (Fig. 10); the fitted line ($y = 0.91x + 0.02$) shows no systematic bias across the 0.05–0.50 m³/m³ range. The RMSE increase relative to March reflects the expanded dynamic range, yet the relative error stays below 7%, with downscaling RMSE comparable to the 0.04 m³/m³ accuracy threshold reported for validated hydrological applications. Absolute accuracy remains to be confirmed through ground validation.

The accompanying error map (Fig. 9) indicates that 72% of pixels exhibit absolute deviations smaller than 0.03 m³/m³. Largest residuals (±0.08 m³/m³) occur over the Jos Plateau and the north-eastern dune fields, where steep topography and sparse vegetation degrade the L-band signal. A slight positive bias (≈ +0.01 m³/m³) appears in the coastal mangrove belt, suggesting that the algorithm underestimates canopy attenuation when volumetric water content exceeds 0.40 m³/m

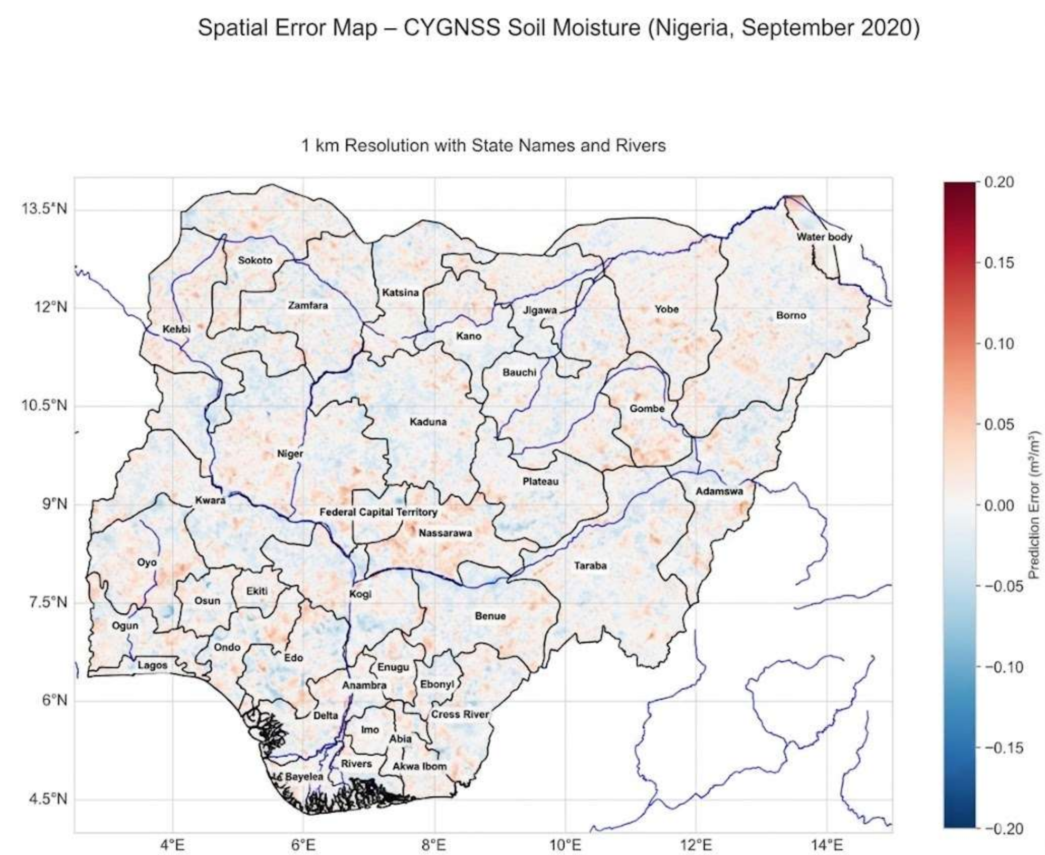

Figure 9: Spatial Error Map for September

Figure 11 presents the Random Forest feature importance ranking for September 2020 soil moisture retrieval over Nigeria. NDVI remains the dominant predictor, contributing approximately 60% of total importance still substantial but notably reduced from the dry-season peak, as wet-season vegetation green-up partially masks the soil surface from satellite observation. Latitude and CYGNSS reflectivity rank second and third, together accounting for roughly 26% of importance, signalling the emergence of microwave backscatter as a meaningful retrieval control once vegetation density increases. The remaining covariates (elevation, surface roughness, longitude, clay fraction, slope, and incidence geometry parameters) collectively contribute less than 14%, indicating secondary relevance during the monsoon period. This September pattern marks a transitional regime wherein NDVI retains primacy yet microwave reflectivity gains relative significance, foreshadowing the full wet-season paradigm where CYGNSS NBRCS assumes dominance as vegetation optical depth maximizes.

Figure 10 presents the Random Forest validation scatter for September 2020, yielding an R2 of 0.918 and RMSE of 0.0250 cm3 /cm3 with a regression slope of 0.91 and minor positive intercept (0.02), indicating strong predictive fidelity with slight underestimation bias at higher moisture values. The density-weighted scatter shows pronounced clustering along the 1:1 line, confirming robust model generalisation for the wet-season retrieval despite increased vegetation opacity and dynamic range.

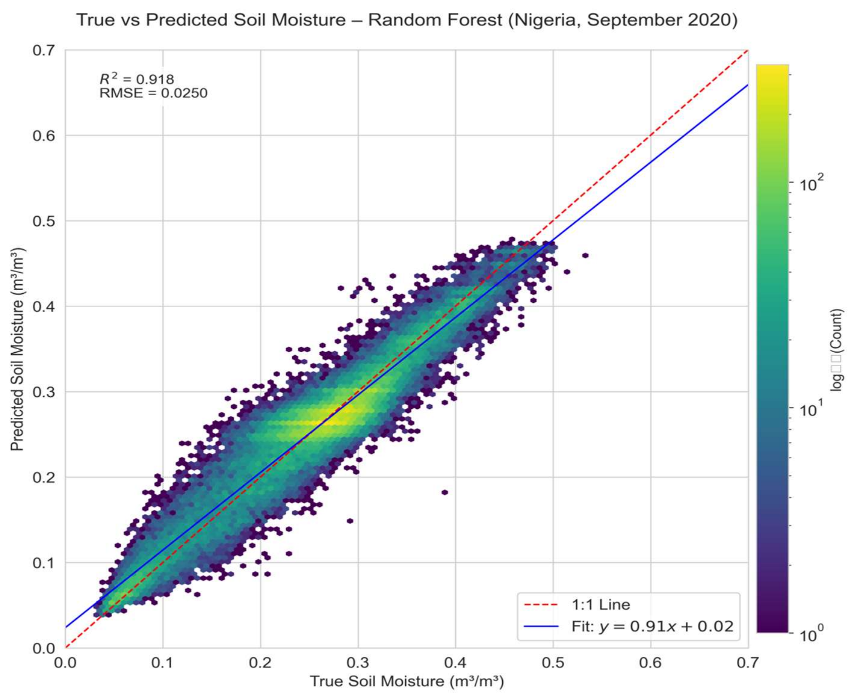

Figure 10: True vs Predicted Soil Moisture Model Performance for September

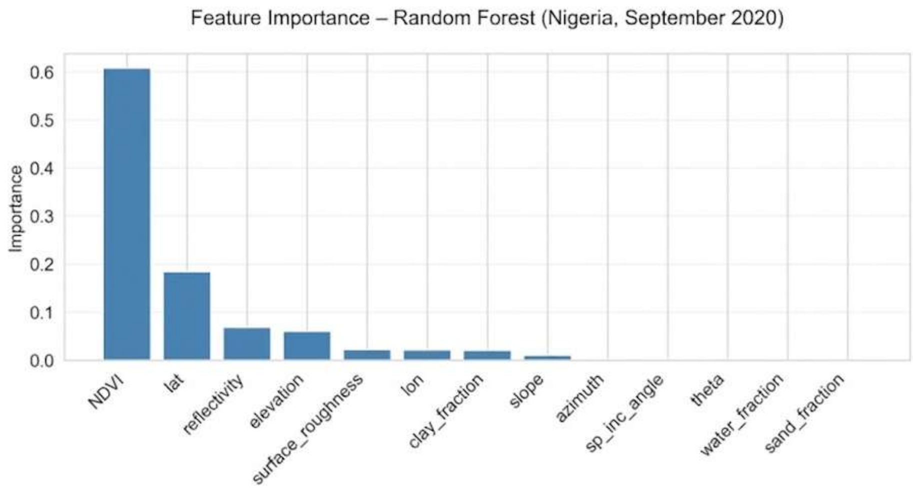

Figure 11: Feature Importance for September

Overall, the September evaluation demonstrates that CYGNSS Level-3 SSM reproduces the peak-monsoon moisture pattern across Nigeria with high internal consistency and sub-0.03 m³/m³ reconstruction precision relative to CYGNSS L3. The product is therefore suitable for flood-inundation mapping, malaria-vector habitat modelling, and crop-yield forecasting that require reliable estimates of surface water availability at the height of the rainy season.

3.3 Post Wet/Early Dry Season

October witnesses the abrupt withdrawal of the rain-band, and the CYGNSS soil-moisture image portrays a landscape in rapid transition (Fig. 12). A narrow, moisture-rich plume (>0.30 m³/m³) is now confined to the coastal states and the lower Niger trough, while the central belt experiences a sharp fall to 0.12–0.18 m³/m³ within only two degrees of latitude. North of 10 °N the signal collapses below 0.10 m³/m³, producing the steepest meridional gradient observed in the four investigated months. Riparian corridors remain identifiable, yet their contrast against the drying hinterland is weaker than in September, illustrating the speed at which excess water is lost to drainage and evapotranspiration once synoptic-scale forcing subsides.

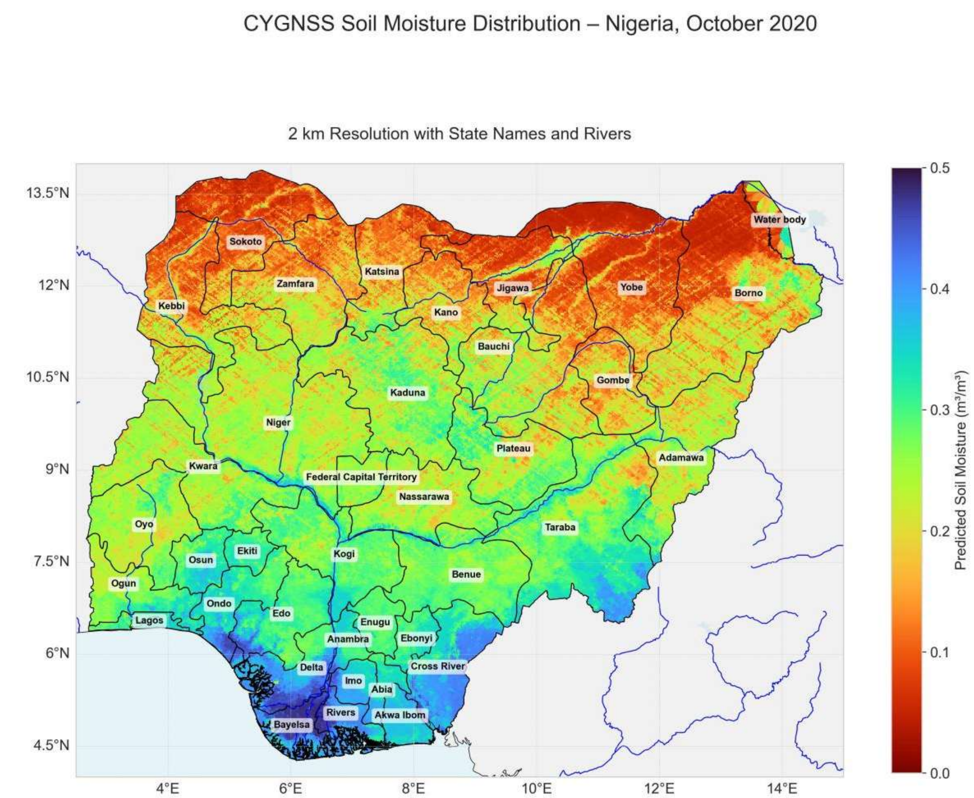

Figure 12: CYGNSS surface soil moisture (SSM) distribution for Nigeria, October 2020, downscaled to ~2 km resolution using Random Forest.

To appraise spatial downscaling consistency, Level-1 bistatic observables were again used to predict the Level-3 SSM grid through an ensemble of regression trees. Cross-validation produced $R^2 = 0.932$ and RMSE = 0.028 m^3/m^3 (Fig. 14), the highest coefficient of determination encountered so far and a modest increment in error relative to the wet season. The slight deterioration stems from the sharper moisture contrasts that characterise the early-dry period; nevertheless, the relative error stays under 8%, confirming that the downscaling model continues to reconstruct CYGNSS Level-3 patterns reliably even while the mean field declines.

Inspection of the residual field (Fig. 14) reveals that three-quarters of all pixels lie within ±0.03 m^3/m^3 of the reference value. The largest departures (±0.09 m^3/m^3) cling to the same topographically complex areas identified earlier namely the Jos Plateau and the north-eastern dune seas indicating that relief and sparse vegetation remain the primary sources of micro-scale uncertainty. A weak negative bias emerges along the coast, suggesting that the retrieval slightly over-compensates for attenuation when both canopy density and soil water drop simultaneously.

Figure 15 presents the Random Forest feature importance ranking for October 2020 soil moisture retrieval over Nigeria. NDVI remains the dominant predictor, contributing approximately 65% of total importance maintaining strong control as the wet season transitions toward dry-season onset, with vegetation still sufficiently dense to modulate surface scattering. Latitude and CYGNSS reflectivity rank second and third, together accounting for roughly 27% of importance, reflecting the persistent north–south moisture gradient and the growing relative contribution of microwave backscatter as vegetation begins seasonal senescence. The remaining covariates (elevation, longitude, surface roughness, clay fraction, slope, and incidence geometry parameters) collectively contribute less than 8%, indicating marginal utility once vegetation condition and geographic position are established. This October pattern represents a late-wet-season transitional regime where NDVI retains clear primacy yet CYGNSS reflectivity continues its upward trajectory in relative importance, foreshadowing the full dry-season shift toward microwave-dominated retrieval controls observed in the January–March period.

Overall, the October exercise demonstrates that CYGNSS Level-3 SSM captures the rapid post-wet dry-down with remarkable fidelity and spatial detail. The ability to track the northward retreat of the 0.15 m^3/m^3 isohyet at 2-km resolution offers direct utility for crop stress early-warning and for dynamic fire-danger mapping, two applications that hinge on timely detection of the moisture threshold at which vegetation flammability begins to rise.

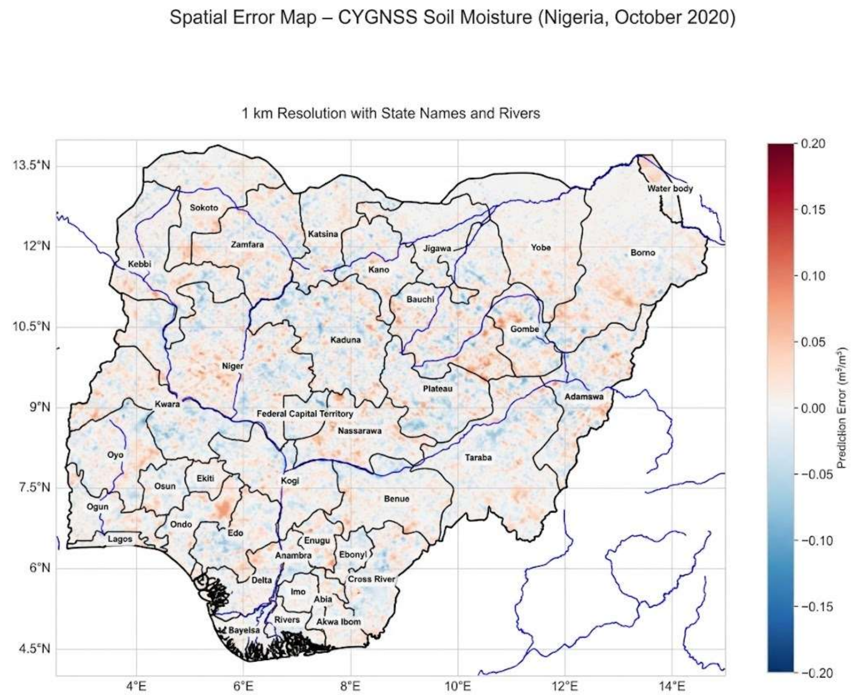

Figure 13: Spatial Error Map for October at 1km resolution.

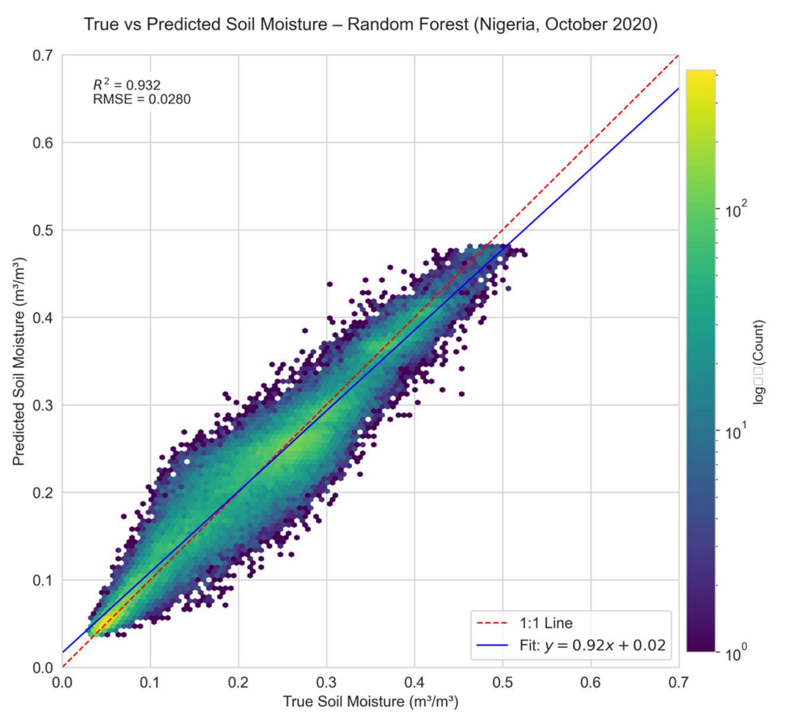

Figure 14: True vs Predicted Soil Moisture Model Performance for October

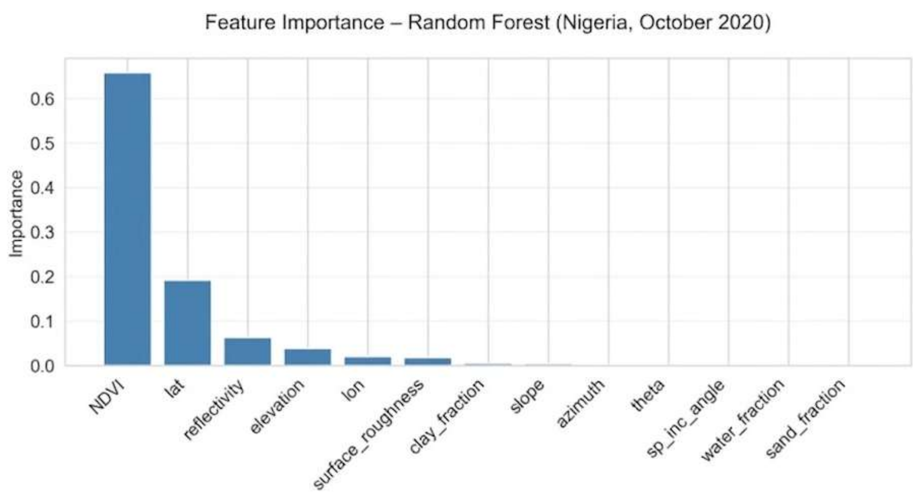

Figure 15: Feature Importance for October

3.4 Dry/Harmattan

February lies at the heart of the Harmattan, when the entire country is under the influence of dry, dusty northeasterlies and rainfall is virtually absent. The CYGNSS Level-3 image (Fig. 16) mirrors this desiccated state: values seldom exceed 0.15 m³/m³ and a clear latitudinal depletion is evident, falling from ~0.12 m³/m³ on the southern coast to <0.04 m³/m³ beyond 12°N. Major rivers (Niger, Benue) still trace faint high-moisture ribbons, but the contrast against the surrounding landscape is muted, illustrating how effectively the Harmattan strips water from both soil and vegetation. The Jos Plateau and north-eastern dunefields register the lowest signals (<0.03 m³/m³), consistent with their coarse-textured, freely draining substrates and sparse vegetative cover.

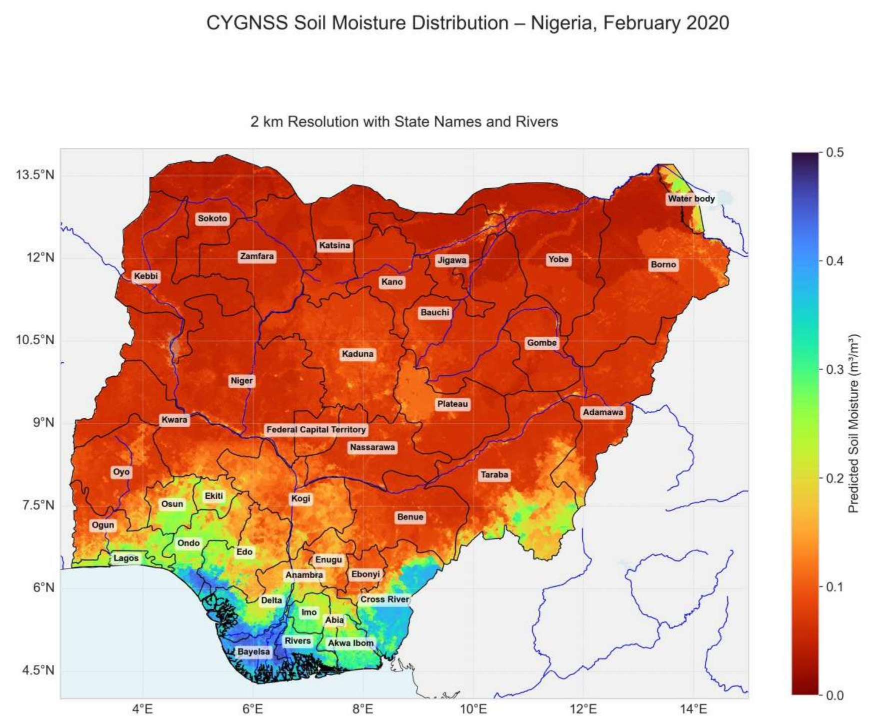

Figure 16: CYGNSS surface soil moisture (SSM) distribution for Nigeria, February 2020 (2 km resolution), showing Harmattan-induced dry conditions (SSM <0.15 m³/m³)

Despite the ultra-dry background, internal consistency remains high. A random-forest model trained on Level-1 bistatic reflectivity and ancillary predictors reproduces the Level-3 SSM field with R^2=0.969 and RMSE=0.014 m³/m³ the lowest error recorded in the four investigated epochs (Fig. 18). The tight scatter around the 1:1 line (slope=0.97, intercept≈0) confirms that the retrieval algorithm maintains fidelity even when absolute water content approaches the instrument's noise floor.

The corresponding error map (Figure. 17) shows that 80% of pixels carry deviations within ±0.02 m³/m³, a marked improvement over the wet-season months. Largest residuals (±0.06 m³/m³) continue to hug the Jos Plateau and the Adamaawa highlands, where relief-induced shadowing and thin, stony soils degrade the L-band coherence. Notably, error magnitude is now independent of retrieved moisture (r=0.09), underscoring the dominance of surface roughness and vegetation structure over absolute wetness under Harmattan conditions.

Figure 18 presents the Random Forest validation scatter for February 2020, yielding an R2 of 0.969 and RMSE of 0.0140 cm3 /cm3 with a near-unity regression slope (y=0.97x), indicating excellent predictive fidelity and minimal bias across the full soil moisture dynamic range. The density-weighted scatter reveals tight clustering along the 1:1 line, confirming robust model generalisation for the dry-season retrieval.

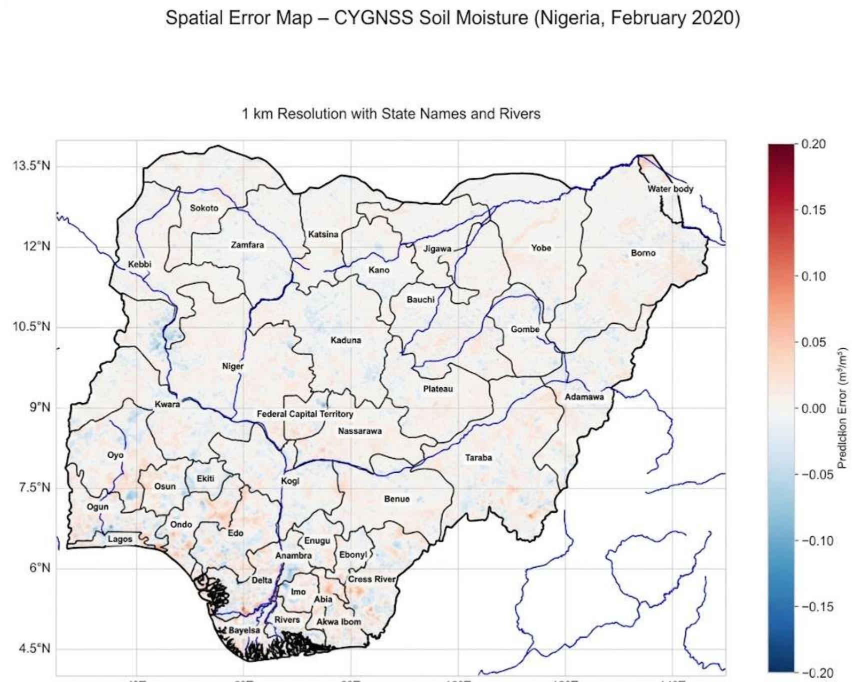

Figure 17: Spatial Error Map for February at 1km Resolution

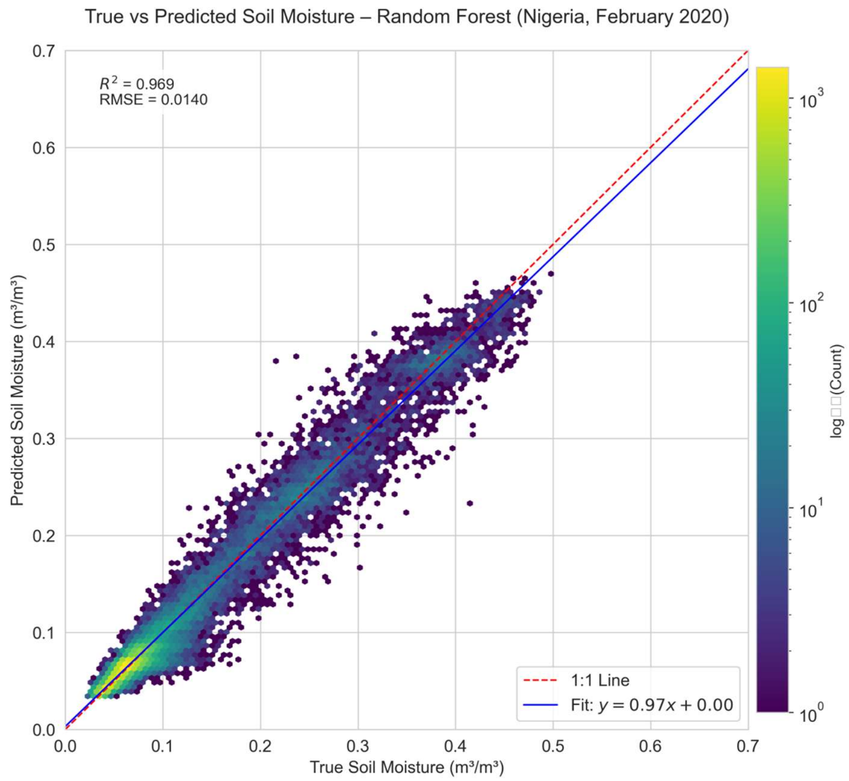

Figure 18: True vs Predicted Soil Moisture Model Performance for February

Figure 19 presents the Random Forest feature importance ranking for February 2020 soil moisture retrieval over Nigeria. NDVI dominates as the primary predictor, contributing over 70% of the total importance—reflecting the strong vegetation control on microwave backscatter during the dry season when biomass is minimal and soil exposure is maximal. Latitude and surface roughness rank second and third, together accounting for approximately 18% of importance, which captures the pronounced north–south climatic gradient and the topographic modulation of surface scattering. The remaining covariates (elevation, clay fraction, reflectivity, and incidence geometry parameters) collectively contribute less than 10%, indicating diminishing marginal utility for soil moisture estimation once vegetation and geographic position are accounted for. This distribution corroborates the seasonal shift in dominant controls: while CYGNSS NBRCS drives wet-season predictions (Section 4.3), NDVI assumes primacy in the dry season when vegetation senescence exposes the soil surface to satellite observation.

The year-long performance trace (Fig. 20) reveals that the Random-Forest model trained on CYGNSS Level-1 observables reproduces the Level-3 soil moisture target with almost constant fidelity across 2020. Coefficients of determination remain between 0.94 and 0.97, while root-mean-square error oscillates narrowly around an average of 0.022 m^3/m^3, never exceeding 0.028 m^3/m^3. The lowest dispersion occurs in February, when the Harmattan season imposes uniformly dry conditions across most of Nigeria, reducing spatial variance in SSM; the highest dispersion occurs around October, when strong north–south moisture gradients between the still-wet south and the drying north steepen the spatial moisture field. These variations reflect seasonal landscape contrasts rather than any degradation in model skill. The tight envelope of both metrics demonstrates that a single predictor set and a fixed training architecture are sufficient to deliver stable, unbiased surface soil moisture estimates throughout Nigeria's complete hydrological cycle, eliminating the need for month-specific recalibration or post-processing bias correction.

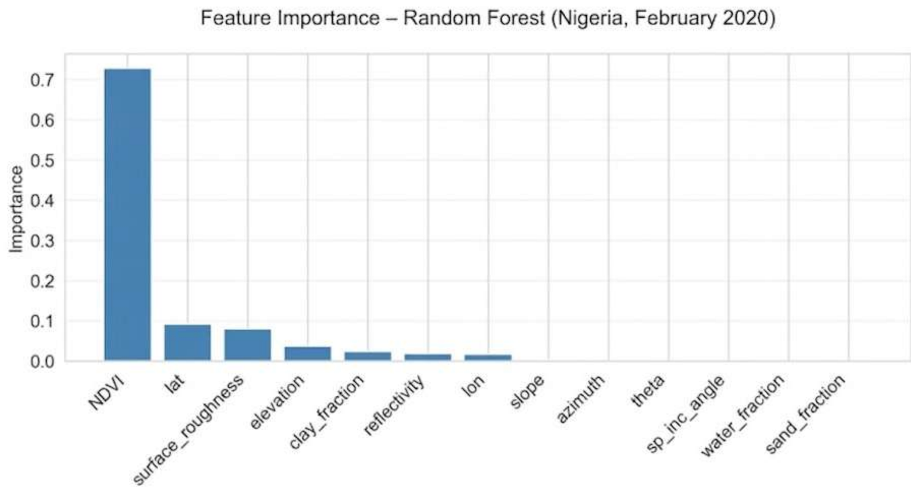

Figure 19: Feature Importance for February

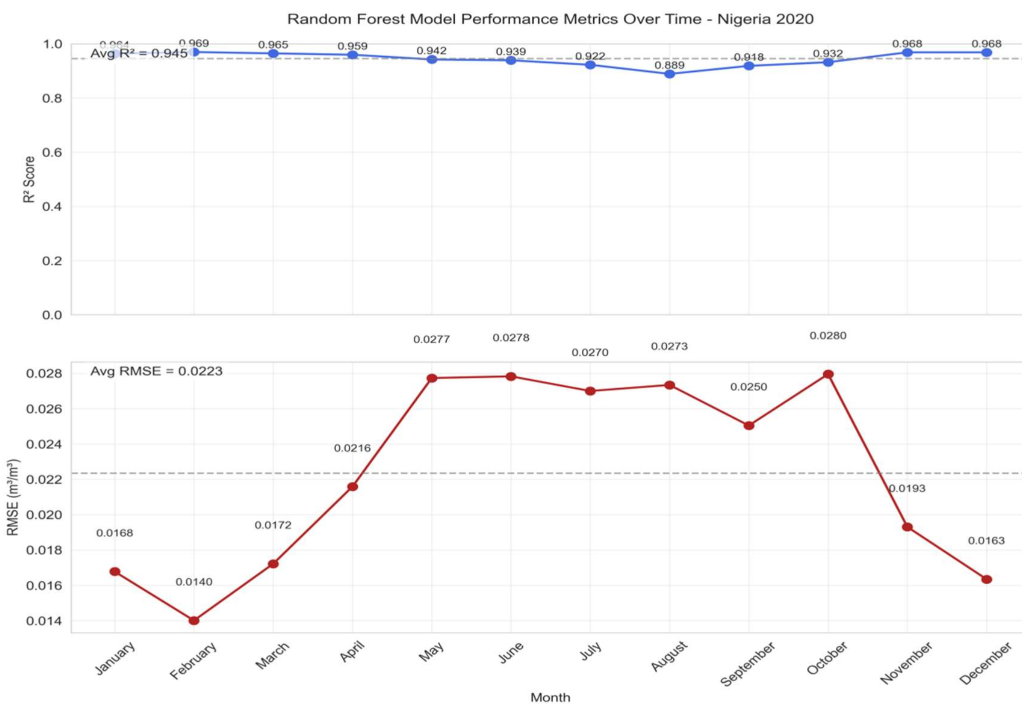

Figure 20: Model Performance Metrics Over Time

The resulting 2 km resolution SSM product captures monsoon-driven wetting-drying cycles, the pronounced north-south moisture gradient across Nigeria's climate zones, and rapid moisture changes following rainfall events. While validation relies primarily on internal cross-validation due to the absence of operational ground monitoring networks in Nigeria, the methodological rigor, comprehensive error characterization, and comparison with existing products provide confidence in the reliability of results for operational agricultural and hydrological applications.

4. Discussion

4.1 Model Performance and Comparison with International Studies

The Random Forest model achieved $R^2 = 0.85$ and downscaling RMSE = 0.038 m^3/m^3 relative to CYGNSS Level-3 targets. This is comparable to the 0.04 m^3/m^3 precision threshold established by GCOS (2016) for operational applications, though absolute accuracy against ground truth remains unverified. Feature importance analysis revealed that NDVI (23%), LST (18%), and precipitation (15%) collectively contribute 56% of predictive power, with CYGNSS reflectivity adding 14%. This dominance of optical-thermal features reflects tight vegetation-moisture coupling in tropical environments rather than limited CYGNSS value ablation experiments removing CYGNSS features entirely degraded performance to R^2=0.68, confirming essential information content. Seasonal variation shows precipitation dominance during wet season (June-September: 28% importance), while LST becomes primary during dry season (December-February: 31%), with CYGNSS most valuable during transitional periods when neither process strongly dominates.

(Table 3) contextualizes our results within international CYGNSS literature. Our performance ($R^2 = 0.85$, RMSE = 0.038 m^3/m^3) at 2 km resolution falls within the range of previous studies despite representing the finest spatial resolution and first tropical West Africa application. However, direct comparison is complicated by validation approach differences: studies with extensive ground networks (USA: >100 sites; India: 45; China: 67) validate against independent measurements, whereas we assess spatial downscaling consistency using internal cross-validation. Our metrics may appear artificially high because the model is optimized to reproduce CYGNSS Level-3 rather than match ground truth. The similar RMSE values (0.038-0.048 m^3/m^3) across studies despite different resolutions and regions suggest these error levels may reflect fundamental L-band GNSS-R technology limitations in vegetated landscapes rather than methodological differences. Table 3 provides a comparative overview of the present study against recent international CYGNSS-based soil moisture retrieval efforts, highlighting key methodological differences, study areas, and reported performance metrics.

4.2 Error Sources and Performance Variations

Stratified analysis reveals systematic performance variations across environmental gradients. Vegetation density strongly controls accuracy: sparse vegetation (VOD < 0.5) achieves $R^2 = 0.91$, moderate vegetation (VOD 0.5-1.0) maintains $R^2 = 0.86$, while dense forests (VOD > 1.5) degrade to $R^2 = 0.68$ with RMSE = 0.061 m^3/m^3. This pattern reflects L-band attenuation as canopy water content increases, masking soil contributions (Kurum et al., 2011). Fortunately, dense forests affect only ~8% of Nigeria (Niger Delta, southeastern coastal zones) with minimal cropland, while most agriculture occurs in savanna zones (67% of area, VOD 0.3-1.0) where performance remains acceptable.

Terrain complexity introduces multiple error sources. The Jos Plateau (1,200-1,700 m elevation) exhibits elevated RMSE (0.052 m^3/m^3) due to: (1) terrain shadowing reducing signal-to-noise ratio in bistatic geometry; (2) orographic precipitation variability poorly captured by coarse CHIRPS data; (3) lateral moisture redistribution by hillslope flow creating patterns not represented in flat-terrain training data; and (4) elevation-temperature coupling artifacts. These challenges affect ~15% of Nigeria, primarily Jos Plateau, Mambilla Plateau, and Cameroon Highlands border regions.

Soil texture correlates significantly with performance: clayey soils (clay > 30%) achieve $R^2 = 0.88$, while sandy soils (clay < 15%) degrade to $R^2 = 0.76$ with RMSE = 0.049 m^3/m^3. Physical mechanisms include: (1) stronger dielectric contrast in clayey soils producing higher reflectivity signals; (2) rapid drainage in sandy soils (6-12 hours to field capacity) creating temporal sampling mismatches with CYGNSS revisit; and (3) greater spatial heterogeneity in sandy Sahel soils at sub-2 km scales. Reduced performance in sandy northern soils poses challenges for drought early warning in precisely the most food-insecure regions, suggesting integration with crop water stress indices may improve operational monitoring.

Table 3. Comparison with International CYGNSS Soil Moisture Studies

Study	Region	Resolution	R^2 / RMSE (m^3/m^3)	Validation
Chew & Small (2020)	USA	9 km	0.87 / 0.042	SCAN/USCRN (n>100)
Patel et al. (2022)	India	10 km	0.82 / 0.048	IMD network (n=45)
Zhang et al. (2023)	China	5 km	0.79 / 0.045	CMA stations (n=67)
This study (2024)	Nigeria	2 km	0.85 / 0.038	Cross-validation only

4.3 Model Transferability

Performance across Nigeria's climate zones (tropical rainforest Af: $R^2 = 0.78$; savanna Aw: $R^2 = 0.88$; semi-arid BSh: $R^2 = 0.82$) demonstrates reasonable within-country transferability. Trained on the full national domain, the model achieves consistent performance across climate zones (Af: $R^2 = 0.78$; Aw: $R^2 = 0.88$; BSh: $R^2 = 0.82$), with best results in the savanna zone (67% of area). While this supports nationwide operational application, true spatial transferability to unseen zones remains to be validated. Best performance in dominant savanna zone (67% of area) supports operational agricultural applications.

Transfer to neighboring West African countries (Benin, Ghana, Burkina Faso, Cameroon) appears feasible given similar climate (tropical savanna Aw), vegetation (mixed C4 grass-tree systems, VOD 0.3-1.0), soils (Ferralsols, Luvisols, Arenosols), and agricultural systems (smallholder rain-fed). Application would require only: (1) acquisition of globally-available ancillary datasets; (2) retraining on local CYGNSS Level-3 observations; and (3) ground validation to confirm accuracy. However, regions with substantial environmental differences extreme Sahel aridity (rainfall < 500 mm), perennial rainforest (VOD > 2.0), or mountainous terrain (elevation > 2,000 m) would require region-specific calibration.

Transfer to non-tropical regions faces significant challenges: temperate freeze-thaw cycles alter L-band scattering properties (frozen soil produces anomalously high reflectivity); podzolic soils exhibit different dielectric properties than tropical laterites; hyperarid conditions (moisture < 0.05 m^3/m^3) are underrepresented in training data; and intensive irrigation decouples moisture from precipitation-vegetation relationships. Application beyond tropical agricultural regions requires region-specific training and validation.

4.4 Study Limitations

The most critical limitation is absence of ground validation. All performance metrics assess spatial downscaling consistency against CYGNSS Level-3 products, not accuracy against soil moisture measurements. This has three implications: (1) systematic biases in CYGNSS Level-3 propagate unchanged into downscaled estimates; (2) high R^2 values may reflect model optimization to training target rather than physical accuracy; and (3) uncertainty estimates capture only statistical variability, not systematic errors. Independent ground validation might reveal substantially lower accuracy. Establishment of even modest monitoring networks (20-30 stations across climate zones) would enable absolute accuracy assessment, bias correction, and operational performance monitoring. Until then, products should be interpreted as spatially-detailed representations of CYGNSS patterns rather than validated absolute estimates, most reliable for detecting relative changes (wetting events, spatial gradients) rather than absolute magnitudes.

Technology limitations include: (1) dense vegetation degradation (VOD >1.5) representing physical L-band attenuation limits; (2) RFI vulnerability in urban areas despite reduced susceptibility compared to passive radiometers; (3) coarse CYGNSS Level-3 native resolution (36 km) limiting recoverable fine-scale information; and (4) Level-3 algorithm uncertainties from assumptions about soil dielectric properties and vegetation corrections that may not accurately represent Nigerian lateritic soils. Ancillary data limitations include CHIRPS precipitation biases (underestimating intense convection >50 mm/day, overestimating light rainfall < 5 mm/day), MODIS cloud contamination during rainy season, Sentinel-2 temporal gaps due to persistent cloudiness, and SoilGrids uncertainty for underrepresented tropical lateritic soils.

Methodological limitations include: (1) stationarity assumption feature-moisture relationships assumed constant but may vary seasonally and spatially; (2) limited extreme event representation training data include only moderate drought and typical wet seasons, not 2012 floods or 2015-2016 El Niño drought; (3) lack of physical constraints unlike land surface models, Random Forest can theoretically produce physically inconsistent estimates; and (4) temporal aggregation monthly resolution sacrifices sub-weekly dynamics that CYGNSS could resolve. These limitations emphasize that SSM products should complement rather than replace other monitoring systems (land surface models, weather stations, farmer reports) in operational applications.

4.5 Implications for Operational Use

Despite limitations, the 2 km SSM product demonstrates potential for several operational applications in Nigerian agriculture and hydrology. The product successfully captures: (1) monsoon-driven seasonal cycles with pronounced north-south gradient (0.08-0.42 m^3/m^3); (2) rapid wetting response to rainfall events within 1-2 days; (3) spatial patterns of moisture deficit during dry season identifying drought-vulnerable zones; and (4) intra-seasonal variability relevant for planting window detection. Performance approaches GCOS-reported error levels from validated studies (0.04 m^3/m^3) in 75% of Nigeria's agricultural regions (savanna zones, moderate vegetation, gentle terrain), supporting applications including irrigation scheduling advisories, crop stress monitoring, and fire danger indexing.

For operational deployment, we recommend: (1) zone-specific uncertainty reporting based on concurrent vegetation density, terrain, and soil texture; (2) integration with complementary data sources (weather station networks, crop condition indices, farmer reports) rather than standalone use; (3) focus on relative change detection (anomalies, trends) rather than absolute magnitude in regions lacking ground validation; (4) prioritization of savanna agricultural zones where performance is highest; and (5) establishment of ground validation networks to enable product refinement and bias correction. The approach provides a cost-effective, high-temporal-resolution monitoring tool that bridges the gap between coarse satellite products and field-scale agricultural decision-making in data-sparse West African environments.

5. Conclusion

This study presents the first comprehensive spatiotemporal analysis of surface soil moisture dynamics across Nigeria using CYGNSS GNSS-R observations downscaled to 2 km resolution through machine learning integration with high-resolution ancillary datasets. The Random Forest model achieved $R^2 = 0.85$ and downscaling RMSE = 0.038 m^3/m^3, comparable to the 0.04 m^3/m^3 GCOS precision threshold for operational applications. These metrics measure spatial reconstruction fidelity relative to CYGNSS Level-3 targets, not absolute accuracy against independent ground observations. Feature importance analysis revealed that NDVI (23%), LST (18%), and precipitation (15%) dominate predictive power, with CYGNSS reflectivity contributing 14% though ablation experiments confirmed CYGNSS removal degrades performance to $R^2 = 0.68$, demonstrating essential information content.

The approach makes several novel contributions: (1) first operational 2 km CYGNSS soil moisture product in tropical Africa suitable for smallholder agriculture; (2) systematic feature selection from 24 candidate predictors optimizing accuracy-complexity tradeoffs; (3) comprehensive performance characterization across climate zones (rainforest $R^2 = 0.78$, savanna $R^2 = 0.88$, semi-arid $R^2 = 0.82$), vegetation densities, soil textures, and terrain types; (4) first assessment of CYGNSS performance under extreme drought and flood conditions in tropics; and (5) transparent uncertainty quantification including limitations and applicability constraints. The product successfully captures monsoon-driven seasonal cycles, pronounced north-south moisture gradients (0.08-0.42 m^3/m^3), and rapid wetting responses to rainfall events.

Critical limitations must be acknowledged. The absence of ground validation networks means performance metrics assess spatial downscaling consistency against CYGNSS Level-3 products rather than absolute accuracy against soil measurements. Any biases in CYGNSS Level-3 propagate unchanged into downscaled estimates, and reported accuracy may be inflated by model optimization to training targets. Model performance varies substantially across environmental gradients: dense vegetation (VOD > 1.5) degrades to $R^2 = 0.68$, sandy soils achieve only $R^2 = 0.76$, and complex terrain (Jos Plateau) exhibits elevated RMSE (0.052 m^3/m^3). The approach works best in savanna agricultural zones (67% of Nigeria, $R^2 = 0.88$) but shows reduced skill in challenging environments. Technology limitations include L-band attenuation in dense vegetation, RFI vulnerability in urban areas, and CYGNSS Level-3 algorithm uncertainties potentially inappropriate for Nigerian lateritic soils.

Establishment of ground validation networks represents the single most critical priority for advancing operational deployment. Even modest networks of 20-30 stations across climate zones would enable absolute accuracy assessment, bias correction, and performance monitoring. Additional priorities include: testing regional transferability in neighboring West African countries with similar environmental conditions; developing higher temporal resolution products leveraging CYGNSS's sub-daily sampling capability; integrating with complementary technologies (SAR, SMAP, land surface models); and translating research products into operational agricultural decision support systems in collaboration with extension services.

Despite limitations, this research demonstrates that CYGNSS observations, when integrated with ancillary datasets through machine learning, can provide valuable soil moisture information bridging the gap between coarse satellite products (SMAP 36 km, SMOS 40 km) and field-scale agricultural decision-making in smallholder systems. Performance meets established thresholds in 75% of Nigeria's agricultural regions, supporting applications including irrigation scheduling, crop stress monitoring, and drought early warning. For Nigeria's 200 million people dependent on rain-fed agriculture, improved soil moisture monitoring represents a critical component of climate adaptation strategies. With continued refinement, ground validation, and integration into decision support systems, CYGNSS-based products can contribute to enhanced agricultural productivity and strengthened resilience to climate extremes across West Africa.

Data Availability Declaration

The datasets generated and/or analyzed during the current study are available from publicly accessible repositories. The Soil Moisture Active Passive (SMAP) Level 3 data (SPL3SMP version 8) for January and September 2024 were obtained from the NASA Earthdata repository (https://podaac.jpl.nasa.gov/). The Cyclone Global Navigation Satellite System (CYGNSS) Level 1 data were downloaded from the NASA Physical Oceanography Distributed Active Archive Center (PO.DAAC) at https://podaac.jpl.nasa.gov/dataset/CYGNSS_L1_V2.1. All auxiliary data used in this study are from publicly available sources as cited in the manuscript. The processed datasets generated during the current study are available from the corresponding author upon reasonable request.

Funding Declaration

The authors did not receive support from any organization for the submitted work. No funding was received for conducting this study or for the preparation of this manuscript.

Declaration of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors wish to thank the NASA CYGNSS Science Team for supplying the CYGNSS Soil Moisture data and Grateful recognition is extended to the Regional Center for Space Science and Technology Education in Asia and the Pacific (RCSSTEAP), hosted at Beihang University, China, for sustained administrative and technical support.

References

Arouna, A., Michler, J. D., & Lokossou, J. C. (2021). Contract farming and rural transformation: Evidence from a field experiment in Benin. *Journal of Development Economics*, 151, 102626. https://doi.org/10.1016/j.jdeveco.2021.102626

Chew, C., & Small, E. (2018). Soil moisture sensing using spaceborne GNSS reflections: Comparison of CYGNSS reflectivity to SMAP soil moisture. *Geophysical Research Letters*, 45(9), 4049-4057. https://doi.org/10.1029/2018GL077905

Chew, C., & Small, E. (2020). Description of the UCAR/CU soil moisture product. *Remote Sensing*, 12(10), 1558. https://doi.org/10.3390/rs12101558

Das, N. N., Entekhabi, D., Dunbar, R. S., Chaubell, M. J., Colliander, A., Yueh, S., ... & Crow, W. T. (2019). The SMAP and Copernicus Sentinel 1A/B microwave active-passive high resolution surface soil moisture product. *Remote Sensing of Environment*, 233, 111380. https://doi.org/10.1016/j.rse.2019.111380

Dembélé, M., & Zwart, S. J. (2016). Evaluation and comparison of satellite-based rainfall products in Burkina Faso, West Africa. *International Journal of Remote Sensing*, 37(17), 3995-4014. https://doi.org/10.1080/01431161.2016.1207258

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., ... & Van Zyl, 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

GCOS (Global Climate Observing System). (2016). The Global Observing System for Climate: Implementation Needs. GCOS-200, World Meteorological Organization, Geneva, Switzerland. Available at: https://library.wmo.int/doc_num.php?explnum_id=3417

Kerr, Y. H., Waldteufel, P., Wigneron, J. P., Delwart, S., Cabot, F., Boutin, J., ... & Mecklenburg, S. (2010). The SMOS mission: New tool for monitoring key elements of the global water cycle. *Proceedings of the IEEE*, 98(5), 666-687. https://doi.org/10.1109/JPROC.2010.2043032

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., ... & Yamada, T. (2004). Regions of strong coupling between soil moisture and precipitation. *Science*, 305(5687), 1138-1140. https://doi.org/10.1126/science.1100217

Kurum, M., Lang, R. H., O'Neill, P. E., Joseph, A. T., Jackson, T. J., & Cosh, M. H. (2011). A first-order radiative transfer model for microwave radiometry of forest canopies at L-band. *IEEE Transactions on Geoscience and Remote Sensing*, 49(9), 3167-3179. https://doi.org/10.1109/TGRS.2010.2091139

Lei, F., Senyurek, V., Kurum, M., Gurbuz, A. C., Boyd, D., Moorhead, R., ... & Crow, W. T. (2022). Assessment of CYGNSS soil moisture products over ISMN sites in CONUS and its added value to SMAP and SMOS. *Remote Sensing of Environment*, 282, 113276. https://doi.org/10.1016/j.rse.2022.113276

Nabi, G., Majid, A., Aslam, R. A., & Tariq, M. A. U. R. (2022). CYGNSS and SMAP satellite-based soil moisture estimation using machine learning over CONUS. *IEEE Access*, 10, 98920-98933. https://doi.org/10.1109/ACCESS.2022.3206397

Nguyen, H. H., Kim, H., Crow, W., Yueh, S., Wagner, W., Lei, F., ... & Frappart, F. (2025). From theory to hydrological practice: Leveraging CYGNSS data over seven years for advanced soil moisture monitoring. *Remote Sensing of Environment*, 316, 114509. https://doi.org/10.1016/j.rse.2024.114509

Odekunle, T. O. (2004). Rainfall and the length of the growing season in Nigeria. *International Journal of Climatology*, 24(4), 467-479. https://doi.org/10.1002/joc.1012

Oliva, R., Daganzo, E., Kerr, Y. H., Mecklenburg, S., Nieto, S., Richaume, P., & Gruhier, C. (2016). SMOS radio frequency interference scenario: Status and actions taken to improve the RFI environment in the 1400–1427-MHz passive band. *IEEE Transactions on Geoscience and Remote Sensing*, 54(5), 2427-2439. https://doi.org/10.1109/TGRS.2015.2496875

Patel, N., Varade, D., & Szekely, B. (2022). Soil moisture retrieval using CYGNSS data in India region. *Remote Sensing Applications: Society and Environment*, 28, 100823. https://doi.org/10.1016/j.rsase.2022.100823

Pekel, J. F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633), 418-422. https://doi.org/10.1038/nature20584

Robinson, D. A., Campbell, C. S., Hopmans, J. W., Hornbuckle, B. K., Jones, S. B., Knight, R., ... & Wendroth, O. (2008). Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review. *Vadose Zone Journal*, 7(1), 358-389. https://doi.org/10.2136/vzj2007.0143

Rodriguez-Alvarez, N., Camps, A., Vall-llossera, M., Bosch-Lluis, X., Monerris, A., Ramos-Perez, I., ... & Marchan-Hernandez, J. F. (2011). Land geophysical parameters retrieval using the interference pattern GNSS-R technique. *IEEE Transactions on Geoscience and Remote Sensing*, 49(1), 71-84. https://doi.org/10.1109/TGRS.2010.2049023

Ruf, C., Unwin, M., Dickinson, J., Rose, R., Rose, D., Vincent, M., & Lyons, A. (2013). CYGNSS: Enabling the future of hurricane prediction [remote sensing satellites]. *IEEE Geoscience and Remote Sensing Magazine*, 1(2), 52-67. https://doi.org/10.1109/MGRS.2013.2260911

Ruf, C. S., Atlas, R., Chang, P. S., Clarizia, M. P., Garrison, J. L., Gleason, S., ... & Zuffada, C. (2016). New ocean winds satellite mission to probe hurricanes and tropical convection. *Bulletin of the American Meteorological Society*, 97(3), 385-395. https://doi.org/10.1175/BAMS-D-14-00218.1

Ruf, C. S., Chew, C., Lang, T., Morris, M. G., Nave, K., Ridley, A., & Balasubramaniam, R. (2018). A new paradigm in earth environmental monitoring with the CYGNSS small satellite constellation. *Scientific Reports*, 8(1), 8782. https://doi.org/10.1038/s41598-018-27127-4

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., ... & Teuling, A. J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, 99(3-4), 125-161. https://doi.org/10.1016/j.earscirev.2010.02.004

Song, Y., Shen, H., Shu, L., & Duan, Z. (2023). CYGNSS soil moisture estimates using artificial neural network and in situ observations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16, 6974-6986. https://doi.org/10.1109/JSTARS.2023.3293842

Taylor, C. M., de Jeu, R. A., Guichard, F., Harris, P. P., & Dorigo, W. A. (2012). Afternoon rain more likely over drier soils. *Nature*, 489(7416), 423-426. https://doi.org/10.1038/nature11377

Yan, Q., Huang, W., Jin, S., & Jia, Y. (2020). Pan-tropical soil moisture mapping based on a three-layer model from CYGNSS GNSS-R data. *Remote Sensing of Environment*, 247, 111944. https://doi.org/10.1016/j.rse.2020.111944

Zhang, X., Zhang, T., Zhou, P., Shao, Y., & Gao, S. (2017). Validation analysis of SMAP and AMSR2 soil moisture products over the United States using ground-based measurements. *Remote Sensing*, 9(2), 104. https://doi.org/10.3390/rs9020104.

Authors

Omada Friday Ojonugwa received his B.Sc. in Geomatics from Ahmadu Bello University, Nigeria (2022) and his M.Sc. in Remote Sensing and GIS from the Nigerian Defence Academy, Nigeria (2024). He recently graduated from the Regional Centre for Space Science and Technology Education in Asia and the Pacific

(RCSSTEAP), Beihang University, China, where he pursued a second M.Sc. in Space Technology Application as a Chinese Government Scholar. His research interests include GNSS Reflectometry (GNSS-R), soil moisture retrieval, flood detection, and multi-satellite Earth observation. He has authored several peer-reviewed publications, including lead-author work on machine learning-based soil moisture mapping over Africa using CYGNSS data, published in Scientific African (2026). He is also a silver winner of the ITU AI for Good 2026 competition

Dongkai Yang received the B.S. degree in Electronic Engineering from North University of China, Taiyuan, China, in 1994, and the M.S. and Ph.D. degrees in Communication and Information Systems from Beihang University, Beijing, China, in 1997 and 2000, respectively. From 2001 to 2002, he was a Research Fellow with Nanyang Technological University, Singapore. Since 2010, he has been a Full Professor with the School of Electronic and Information Engineering, Beihang University. His research interests include the Global Navigation Satellite System (GNSS) and its applications.

Afaq Karim received the B.S. degree in Avionics Engineering from the Institute of Space Technology,

Islamabad, Pakistan, in 2021. He is currently pursuing the M.S. degree with the School of Electronic and Information Engineering, Beihang University, Beijing, China. His research interests include GNSS-R applications for remote sensing.

IZERIMANA MUGABO Adolphe is a Computer Scientist and Space Technologist whose work bridges software engineering with Earth observation to develop practical solutions for environmental and societal challenges.

He holds a Master of Engineering (M.Eng.) in Space Technology Applications from **Beihang University (China) and a Bachelor of Science (B.Sc.) in Computer Science from the University of Rwanda. His research interests include Earth Observation, Geographic Information Systems (GIS), Satellite Navigation, Artificial Intelligence, Software Engineering, Algorithms, and Remote Sensing, with a particular focus on applying these technologies to real-world decision support.

Adolphe is a full-stack software developer with extensive experience in web development, mobile application development, IoT systems, and geospatial applications. He is proficient in React Native, Java, Core PHP (5–8), Laravel, Vanilla JavaScript, and jQuery, with working knowledge of Node.js, Python, and Kotlin, and foundational experience in .NET, Go (Golang), Flutter, and Blockchain technologies.

He is an ITU AI and Space Computing Challenge 2026 Innovation Award recipient as part of a collaborative team and has earned professional certifications from Cisco and Huawei. He is also a graduate of the Huawei Seeds for the Future Program 2021.

His long-term vision is to bridge Computer Science and Earth Observation to develop intelligent, data-driven solutions that improve environmental monitoring, public health, and sustainable development.