Article

Not peer-reviewed version

CYGNSS Soil Moisture Performance in Guinea Savanna Region: Extended and Quadruple Collocation Evidence from Benue State, Nigeria

Ajoniloju Samuel Olatunde * , Sheikh Tawhidul Islam , Caleb Kelly , Alhassan Maltiti Abdul-Sobbur

Posted Date: 12 May 2026

doi: 10.20944/preprints202605.0778.v1

Keywords: CYGNSS; soil moisture; Extended Triple Collocation; Quadruple Collocation; drought monitoring; Guinea savanna

 doi:10.20944/preprints202605.0778.v1

Article

CYGNSS Soil Moisture Performance in Guinea Savanna Region: Extended and Quadruple Collocation Evidence from Benue State, Nigeria

Ajoniloju Samuel Olatunde *, Sheikh Tawhidul Islam, Caleb Kelly and Alhassan Maltiti Abdul-Sobbur

Hangzhou International Innovation Institute of Beihang University, Hangzhou, China
* Correspondence: olatunde.ajoniloju@outlook.com

Highlights

What are the main findings?

- CYGNSS Level 3 soil moisture shows measurable but limited standalone skill in Guinea savanna agriculture.
- Quadruple Collocation reveals CYGNSS-SMAP error dependence and gives a lower SMAP-independent skill estimate.

What are the implications of the main finding?

- CYGNSS performs worst under dry soils and Harmattan conditions, when drought monitoring is most critical.
- Bias correction should account for vegetation, soil moisture regime, precipitation, land cover, and seasonality.

Abstract

Reliable soil moisture information is essential for agricultural drought warning, but tropical smallholder regions often lack ground networks for validating satellite products. This study evaluates CYGNSS Level 3 soil moisture in Guinea savanna agriculture over Benue State, Nigeria, from 2021 to 2023. Extended Triple Collocation was applied to CYGNSS, SMAP Enhanced Level 3, and ERA5-Land anomalies. Quadruple Collocation then used ESA CCI ACTIVE as a fourth product to quantify CYGNSS-SMAP error dependence. The standard SMAP-inclusive configuration gives CYGNSS a correlation with unknown true soil moisture of $r = 0.425$, an error standard deviation of $0.036\,\mathrm{m^3\,m^{-3}}$, and a signal-to-noise ratio of -6.56 dB. Quadruple Collocation identifies a CYGNSS-SMAP cross-error correlation of 0.325 and reduces the SMAP-independent CYGNSS estimate to $r = 0.386$, indicating that SMAP-inclusive validation overstates retrieval skill. Performance is weakest under dry soils ($r = 0.331$), where drought detection is most important, and location-level ETC convergence fails during Harmattan conditions as anomaly variance collapses. Skill is higher over cropland ($r = 0.447$), shrubland or grassland ($r = 0.455$), and moderate precipitation conditions ($r = 0.630$), but lower over tree cover ($r = 0.342$). These findings show that uncorrected CYGNSS Level 3 soil moisture is not sufficient for standalone year-round drought monitoring in Guinea savanna agriculture. Its value is strongest in bias-corrected, multi-sensor systems that account for vegetation, soil moisture state, precipitation history, land cover, and seasonality.

Keywords: CYGNSS; soil moisture; Extended Triple Collocation; Quadruple Collocation; drought monitoring; Guinea savanna

 doi:10.20944/preprints202605.0778.v1

1. Introduction

1.1. Soil Moisture and Drought Monitoring in Rainfed Agriculture

Soil moisture regulates the exchange of water and energy between the land surface and atmosphere. It controls infiltration, runoff, evapotranspiration, plant water uptake, and the partitioning of rainfall into productive and nonproductive pathways [1,2]. In agricultural systems, soil moisture is often more informative than rainfall alone. Rainfall describes water input, while soil moisture shows whether water remains available to crops after infiltration, drainage, evaporation, and uptake.

This distinction is important in rainfed farming systems. Crop water stress can develop even after rainfall if the water is poorly distributed or the soil dries rapidly during sensitive growth stages [3,4]. In sub-Saharan Africa, smallholder production is highly exposed to rainfall variability. Soil moisture monitoring can support drought early warning, planting decisions, water management, and food-security assessment in these systems [5,6].

Benue State, Nigeria, is a strong example of this need. The state is widely recognized as a major food-producing region and supports rainfed cultivation of yam, cassava, rice, sorghum, soybean, and other staple crops [7–10]. Agricultural production depends on the timing, duration, and persistence of seasonal soil water. Reliable soil moisture information is therefore essential for monitoring agricultural drought risk in the region.

1.2. The Tropical Soil Moisture Observation Gap

Many tropical agricultural regions lack dense ground-based soil moisture networks. This weakens direct validation of satellite products and limits confidence in their operational use. The International Soil Moisture Network has improved access to ground observations, but large parts of sub-Saharan Africa remain poorly represented [5,11]. Benue State has no adequate in-situ soil moisture network for conventional satellite validation.

This creates a difficult validation problem. Satellite products are needed because ground observations are sparse, but their quality cannot be assessed through ordinary ground comparison. Correlation with rainfall or land surface models can provide useful context, but it cannot estimate retrieval error variance, correlation with the unknown true signal, or signal-to-noise ratio. Reference-free methods are therefore needed for rigorous assessment in this environment [12–14].

1.3. CYGNSS as a High-Revisit Tropical Soil Moisture Opportunity

The Cyclone Global Navigation Satellite System offers a promising opportunity for tropical soil moisture monitoring. CYGNSS consists of eight microsatellites that receive reflected Global Positioning System signals within the tropical and subtropical belt between approximately 38°N and 38°S [15]. These reflected L-band signals are sensitive to surface dielectric properties and therefore to near-surface soil water content [16,17].

The main advantage of CYGNSS for agriculture is temporal sampling. Conventional L-band missions such as SMAP and SMOS provide accurate soil moisture estimates, but their revisit intervals are commonly measured in days [18–20]. Tropical rainfall is often convective and spatially variable. Soil wetting and drying can occur faster than the revisit interval of passive microwave missions. CYGNSS can sample tropical land more frequently, which makes it attractive for monitoring short-term wetting and drying cycles [15,21].

The CYGNSS Level 3 soil moisture product converts calibrated surface reflectivity into volumetric soil moisture through regression relationships trained against SMAP observations [21]. This design transfers information from a mature L-band mission to a high-revisit GNSS-R constellation. It also creates a validation concern. If CYGNSS is trained against SMAP, then CYGNSS and SMAP retrieval errors may be correlated. This matters because standard collocation methods assume independent errors.

 doi:10.20944/preprints202605.0778.v1

1.4. Why Existing CYGNSS Validation Is Insufficient

Published CYGNSS validation studies provide important evidence, but they do not fully resolve the West African Guinea savanna case. Several studies evaluated CYGNSS retrievals in temperate regions with in-situ networks or used algorithms calibrated against SMAP and ground observations [21–25]. Other studies provided global or quasi-global assessments that identify broad performance patterns [26,27]. These studies do not directly answer whether uncorrected CYGNSS Level 3 soil moisture can support drought monitoring in tropical smallholder agriculture.

Guinea savanna agriculture presents linked retrieval challenges. Crop canopies develop seasonally, cropland and woodland are mixed within satellite footprints, rainfall is strongly seasonal, and Harmattan conditions can suppress soil moisture variability. These conditions differ from many temperate validation sites. They also interact with the SMAP-based training history of the CYGNSS Level 3 product. A regional validation study is therefore needed to separate actual tropical retrieval skill from performance that may be inflated by shared training dependence.

1.5. Study Novelty and Objectives

This study provides a reference-free assessment of CYGNSS Level 3 soil moisture performance in Guinea savanna agriculture. It uses Benue State, Nigeria, as a case study for West African rainfed smallholder systems. Its novelty lies in combining Extended Triple Collocation, Quadruple Collocation, environmental stratification, and Harmattan-period diagnostics within one tropical agricultural validation framework.

The objectives are to: (1) quantify CYGNSS Level 3 soil moisture skill in Guinea savanna agriculture using Extended Triple Collocation; (2) estimate CYGNSS-SMAP error dependence using Quadruple Collocation; (3) diagnose retrieval sensitivity to vegetation density, soil moisture regime, precipitation regime, and land cover; (4) assess wet-season, dry-season, and Harmattan-period limitations for drought monitoring; and (5) identify the environmental covariates that should guide tropical CYGNSS bias correction.

2. Materials and Methods

2.1. Study Area

Benue State is located in the North Central geopolitical zone of Nigeria between approximately 6.44°N and 8.14°N, and 7.55°E and 9.94°E. It covers about 34,059 km^2 across 23 Local Government Areas. Elevation ranges from low-lying floodplains near the Benue and Katsina-Ala Rivers to upland areas approaching 450 m above sea level [28]. The state lies fully within the CYGNSS observation domain.

The climate is tropical savanna under the Koppen Aw classification [29]. Rainfall is strongly seasonal. The wet season generally extends from April to October, while the dry season extends from November to March under northeasterly Harmattan winds [30,31]. Annual rainfall commonly ranges from about 1200 to 1500 mm, with peak rainfall between July and September [7,31]. This seasonal cycle produces large soil moisture changes and provides a useful dynamic range for satellite evaluation.

The land surface is an agricultural mosaic. ESA WorldCover 2021 indicates that cropland accounts for 46.6% of the study domain, tree cover for 29.2%, and shrubland or grassland for 24.2%. This mixture is important because GNSS-R retrievals respond differently over seasonal crop canopies, open savanna, and persistent woodland. The absence of a dense in-situ soil moisture network makes Benue State an appropriate setting for collocation-based validation. Figure 1 shows the study area and the intended analysis-grid context.

 doi:10.20944/preprints202605.0778.v1

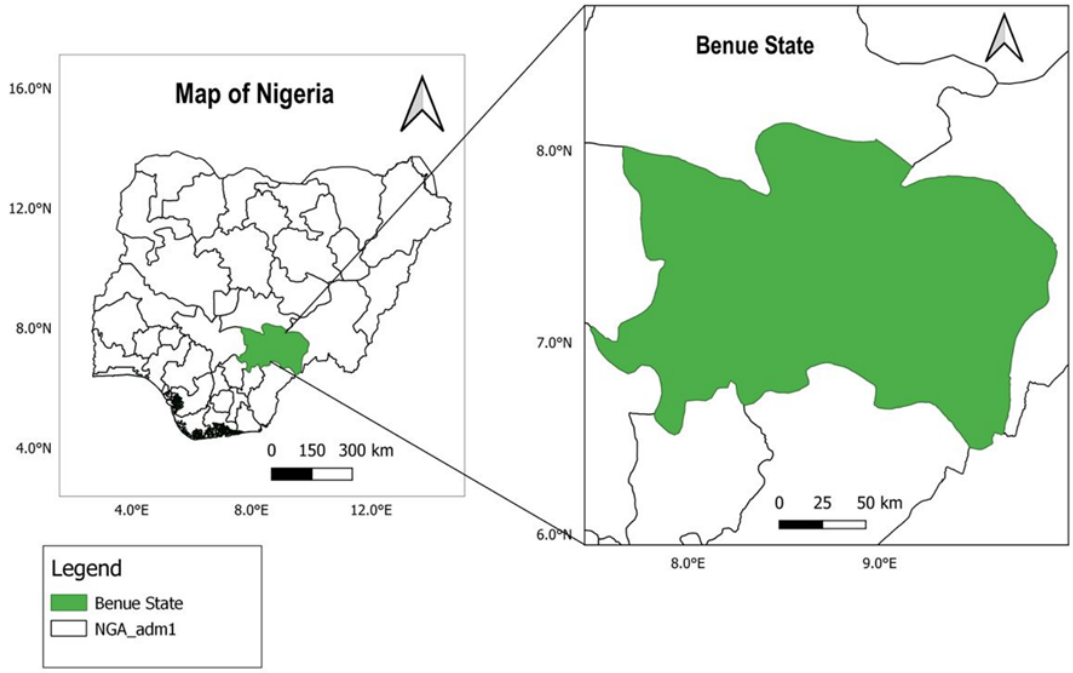

Figure 1. Study area of Benue State, Nigeria

2.2. Datasets

Six datasets were used for collocation analysis, environmental stratification, and robustness testing. They include one GNSS-R soil moisture product, one passive microwave product, one land surface reanalysis, one vegetation index product, one active microwave soil moisture product, and one land cover product (Table 1).

Table 1. Datasets used in the study.

Dataset	Source	Resolution	Period	Role in analysis
CYGNSS L3 Soil Moisture v3.2	NASA PO.DAAC	0.1°	2021–2023	Target GNSS-R soil moisture product evaluated in all ETC and stratified performance analyses.
SMAP Enhanced L3 v6	NASA NSIDC	9 km	2021–2023	Passive microwave soil moisture product used in the primary ETC triplet and as the SMAP-inclusive benchmark.
ERA5-Land	ECMWF CDS	0.1°	2021–2023	Structurally independent land-surface model product and source of precipitation data.
MODIS MOD13Q1 NDVI v61	NASA LP DAAC	250 m	2021–2023	Vegetation-density covariate used to stratify CYGNSS retrieval performance.
ESA CCI ACTIVE v09.2	CEDA and H-SAF	0.25°	2021–2023	Active microwave soil moisture product used as the fourth product in Quadruple Collocation.
ESA WorldCover 2021 v200	ESA and Copernicus	10 m	Static	Land-cover product used to assign dominant cropland, shrubland or grassland, and tree-cover classes.

CYGNSS Level 3 Soil Moisture v3.2 was obtained from NASA PO.DAAC. The product provides volumetric soil moisture estimates derived from calibrated GNSS-R surface reflectivity through SMAP-trained regression relationships [21]. Only retrievals with uncertainty below 0.1 $m^3\ m^{-3}$ were retained.

SMAP Enhanced Level 3 v6 was obtained from NASA NSIDC. SMAP measures L-band brightness temperature at 1.41 GHz and provides surface soil moisture estimates for the upper few centimeters of the soil column [18,20,32]. Descending morning overpasses were used because they are generally less affected by thermal disequilibrium than afternoon retrievals.

 doi:10.20944/preprints202605.0778.v1

ERA5-Land was obtained from the ECMWF Climate Data Store. It provides hourly land surface variables at 0.1° resolution from an offline land surface model driven by ERA5 atmospheric forcing [33]. The 0 to 7 cm soil layer was used for surface soil moisture, and seven-day accumulated precipitation was used for stratification.

MODIS MOD13Q1 NDVI was used to characterize vegetation density [34]. ESA CCI ACTIVE v09.2 was used as the fourth product for Quadruple Collocation. It is derived from ASCAT C-band scatterometer observations and has no shared sensor type, retrieval algorithm, or training data with CYGNSS or SMAP [14,35]. ESA WorldCover 2021 was used to assign the dominant land cover class in each grid cell [36].

2.3. Spatial and Temporal Harmonisation

All datasets were harmonised to a common 0.1° grid with 264 spatial locations over Benue State. This grid matches the native resolution of ERA5-Land and the finer CYGNSS Level 3 gridding option. ERA5-Land fields were interpolated to the target grid. SMAP observations were averaged within 0.05° of each grid point. CYGNSS observations were assigned by nearest-neighbour matching within the same radius. MODIS NDVI was spatially averaged within each grid cell.

Temporal harmonisation used daily matching. CYGNSS and SMAP observations were paired with ERA5-Land daily mean surface soil moisture. A valid ETC triplet required simultaneous CYGNSS, SMAP, and ERA5-Land observations after all quality filters. A valid quadruplet additionally required ESA CCI ACTIVE. The analysis period covered 2021 to 2023, equivalent to 1095 daily time steps.

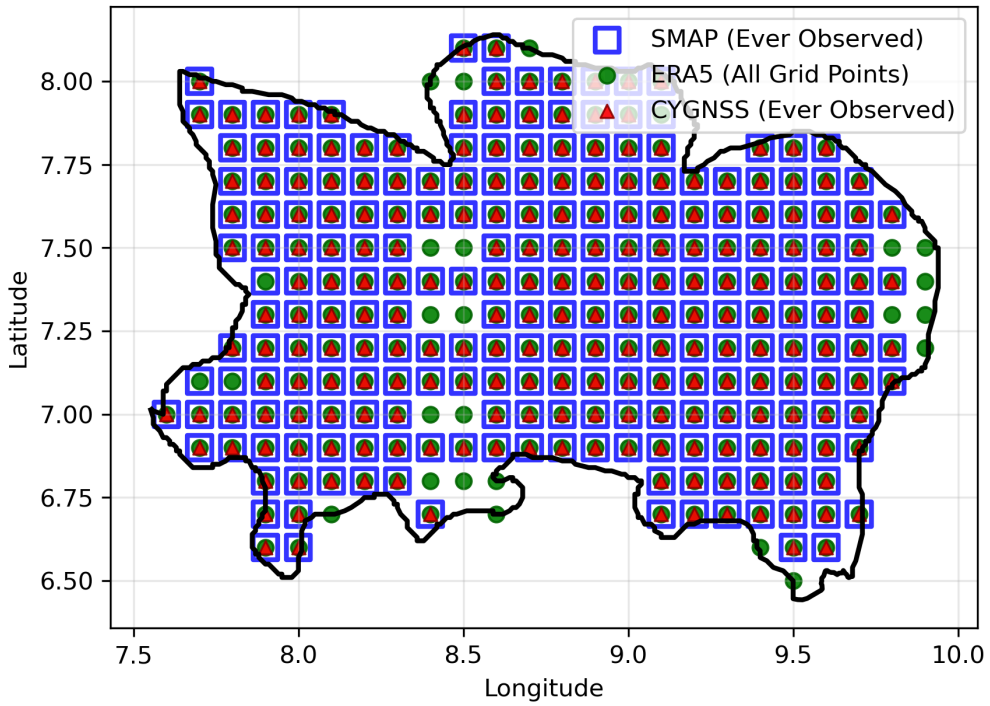

Figure 2. Spatial distribution of SMAP, ERA5-Land, and CYGNSS observation coverage across the Benue State analysis grid. Blue squares indicate SMAP-observed grid cells, green circles show ERA5-Land grid points, and red triangles mark CYGNSS-observed locations.

2.4. Soil Moisture Anomaly Calculation

Collocation analysis was applied to soil moisture anomalies rather than raw values. This reduces systematic offsets among products and focuses the analysis on temporal variability [14]. It also improves compliance with the linearity assumption in triple collocation.

 doi:10.20944/preprints202605.0778.v1

For each product, grid cell, and calendar year, the anomaly was calculated using a 30-day moving window centred as closely as possible on the target day. Let $x_{i,t}$ denote soil moisture from product i at time t. The moving-window mean and anomaly are:

$$\bar{x}_{i,t}^{(30)} = \frac{1}{n_{i,t}} \sum_{k=-14}^{15} x_{i,t+k}, \qquad x'_{i,t} = x_{i,t} - \bar{x}_{i,t}^{(30)}, \tag{1}$$

where $n_{i,t}$ is the number of valid observations for product i within the 30-day window from day $t-14$ to day $t+15$. This window contains 30 possible daily observations and follows the common soil moisture anomaly convention used in triple-collocation studies. The window was applied independently within each calendar year to avoid cross-year contamination near seasonal transitions. Sensitivity tests using 15-day and 60-day windows are reported in Section 3.6 and Appendix A.1.

2.5. Extended Triple Collocation

Extended Triple Collocation estimates error variance and correlation with the unknown true signal without requiring ground truth [12,13]. Each product is represented as an affine function of the unknown true soil moisture anomaly:

$$x_i = \alpha_i + \beta_i \theta + \epsilon_i, \tag{2}$$

where x_i is the observation from product i, θ is the unknown true anomaly, α_i is an intercept, β_i is a scaling coefficient, and ϵ_i is a zero-mean random error. The standard assumptions are linearity, stationarity, error orthogonality with the truth, and zero error cross-correlation among products [13,14].

For a triplet (X, Y, Z), the error variance of product X is estimated as:

$$\sigma_{\epsilon_X}^2 = \sigma_X^2 - \frac{\sigma_{XY}\sigma_{XZ}}{\sigma_{YZ}}, \tag{3}$$

with analogous expressions for Y and Z. Correlation with the unknown truth is estimated as:

$$\rho_{X,\theta} = \sqrt{1 - \frac{\sigma_{\epsilon_X}^2}{\sigma_X^2}}. \tag{4}$$

The positive root is used because soil moisture products should increase with true soil moisture. Signal-to-noise ratio was expressed in decibels as:

$$\mathrm{SNR}_X = -10 \log_{10} \left(\frac{\sigma_{\epsilon_X}^2}{\sigma_X^2 - \sigma_{\epsilon_X}^2} \right). \tag{5}$$

Positive SNR indicates that signal variance exceeds noise variance. Negative SNR indicates that retrieval noise exceeds the useful signal.

2.6. Quadruple Collocation

Quadruple Collocation was used because the standard CYGNSS–SMAP–ERA5-Land triplet may violate the ETC assumption of mutually independent random errors. The concern is structural. CYGNSS Level 3 soil moisture is generated using regression relationships trained against SMAP observations [21]. Shared training can therefore introduce correlated retrieval errors between CYGNSS and SMAP.

In the standard ETC model, the random errors of the three products are assumed to be mutually uncorrelated. For the CYGNSS and SMAP pair, this assumption can be written as:

$$\mathrm{cov}(\varepsilon_{\mathrm{CYGNSS}}, \varepsilon_{\mathrm{SMAP}}) = 0, \tag{6}$$

 doi:10.20944/preprints202605.0778.v1

where $\varepsilon_{\mathrm{CYGNSS}}$ and $\varepsilon_{\mathrm{SMAP}}$ are the random errors of CYGNSS and SMAP, respectively. Because CYGNSS Level 3 is trained against SMAP, Equation (6) may not hold in practice.

Quadruple Collocation addresses this limitation by introducing a fourth structurally independent product and estimating the error dependence between the potentially correlated pair. After the error covariance and error standard deviations are obtained from the Quadruple Collocation solution, the CYGNSS–SMAP cross-error correlation is defined as:

$$r_{\varepsilon,\mathrm{CYGNSS,SMAP}} = \frac{\mathrm{cov}(\varepsilon_{\mathrm{CYGNSS}}, \varepsilon_{\mathrm{SMAP}})}{\sigma_{\varepsilon,\mathrm{CYGNSS}}\sigma_{\varepsilon,\mathrm{SMAP}}}, \tag{7}$$

where $r_{\varepsilon,\mathrm{CYGNSS,SMAP}}$ is the cross-error correlation, $\mathrm{cov}(\varepsilon_{\mathrm{CYGNSS}}, \varepsilon_{\mathrm{SMAP}})$ is the CYGNSS–SMAP error covariance, and $\sigma_{\varepsilon,\mathrm{CYGNSS}}$ and $\sigma_{\varepsilon,\mathrm{SMAP}}$ are the corresponding error standard deviations.

ESA CCI ACTIVE was introduced as the fourth product because it is derived from ASCAT C-band scatterometer observations and is independent of the CYGNSS and SMAP retrieval chains [14,35,41]. The Quadruple Collocation analysis produced three outputs: the CYGNSS–SMAP cross-error correlation, the SMAP-inclusive CYGNSS performance estimate, and the SMAP-independent CYGNSS performance estimate. The SMAP-independent estimate is used as the conservative baseline for interpreting CYGNSS skill in this study.

2.7. Environmental Stratification

CYGNSS performance was stratified by four environmental controls that influence GNSS-R soil moisture retrievals: vegetation density, soil moisture regime, precipitation regime, and land cover class (Table 2). Vegetation density was represented by MODIS NDVI. Soil moisture regime was defined using ERA5-Land surface soil moisture. Precipitation regime was represented by seven-day accumulated ERA5-Land precipitation. Land cover was assigned from ESA WorldCover 2021.

Table 2. Environmental stratification classes used in the analysis.

Variable	Class	Range	Physical rationale
NDVI	Low	0.0 to 0.3	Sparse canopy and limited vegetation attenuation.
NDVI	Medium	0.3 to 0.6	Active crop or savanna growth with moderate attenuation.
NDVI	High	0.6 to 1.0	Dense canopy and stronger L-band attenuation.
Soil moisture	Dry	0.00 to 0.15 $m^3\ m^{-3}$	Drought-relevant conditions and reduced dielectric sensitivity.
Soil moisture	Medium	0.15 to 0.30 $m^3\ m^{-3}$	Intermediate moisture and stronger dielectric sensitivity.
Soil moisture	Wet	0.30 to 0.60 $m^3\ m^{-3}$	Possible saturation, ponding, or reduced contrast.
Precipitation	Low	0 to 10 mm $week^{-1}$	Limited recent wetting.
Precipitation	Moderate	10 to 30 mm $week^{-1}$	Active wetting and drying with limited ponding risk.
Precipitation	High	$>$ 30 mm $week^{-1}$	Possible surface water and wet-condition interference.
Land cover	Cropland	Majority class	Seasonal agricultural canopy.
Land cover	Shrubland/grassland	Majority class	Open savanna surface.
Land cover	Tree cover	Majority class	Persistent woody canopy.

2.8. Uncertainty and Robustness Assessment

Uncertainty was quantified using bootstrap resampling, pairwise significance tests, sensitivity analysis, and spatial diagnostics. Bootstrap confidence intervals used 1000 iterations. Pairwise environmental contrasts were considered significant when the 95% confidence interval of the difference excluded zero.

 doi:10.20944/preprints202605.0778.v1

Robustness tests examined 15-day, 30-day, and 60-day anomaly windows; random subsampling at 75%, 50%, and 25% of valid triplets; wet-season and dry-season partitioning; monthly convergence diagnostics; bootstrap seed independence; and spatial autocorrelation correction. Moran's I and the Clifford effective sample size correction were used to assess spatial dependence among location-level ETC estimates [37]. Detailed diagnostic tables are placed in Appendix A. The complete processing and evaluation sequence is summarized in Figure 3.

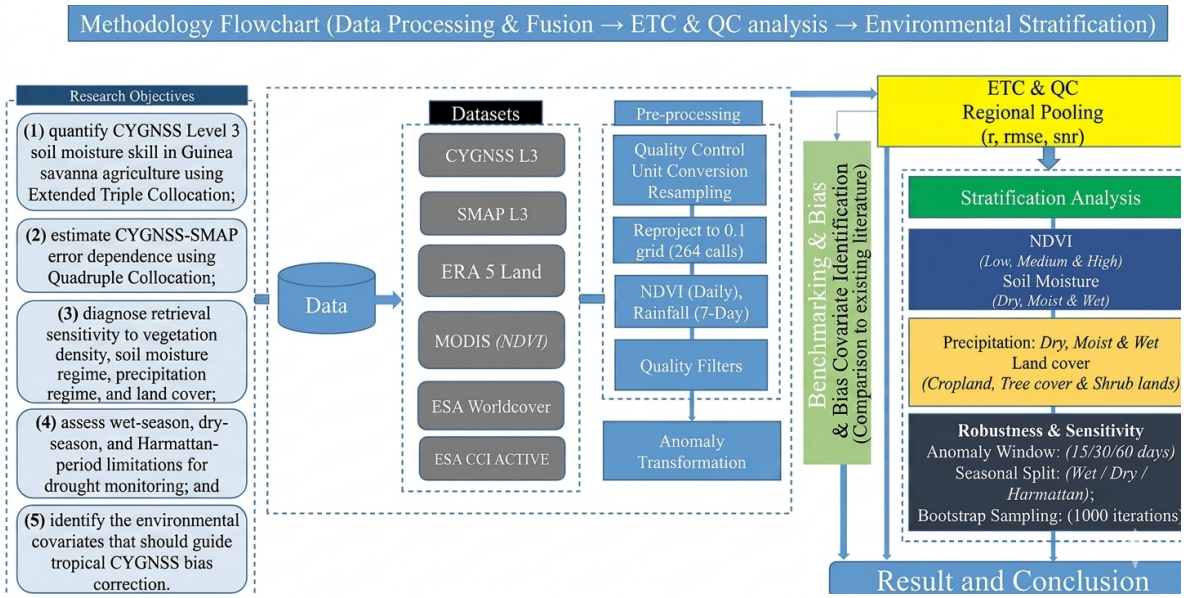

Figure 3. Methodology Flowchart

3. Results

3.1. Collocated Data Availability and Sampling Characteristics

The full space-time domain contained 289,080 possible daily grid-time slots. After quality screening, CYGNSS provided 56,947 valid observations, equal to 19.7% of the full domain. SMAP provided 189,413 valid observations, equal to 65.5%. ERA5-Land provided complete coverage by definition.

The final ETC dataset contained 40,003 valid CYGNSS-SMAP-ERA5-Land triplets. Quadruple Collocation used 39,906 valid four-product records (Table 3). Wet-season triplets accounted for 70.0% of the valid ETC sample. Dry-season triplets accounted for 30.0%.

Table 3. Sample flow and collocation summary for the 2021 to 2023 analysis period.

Stage	N	Description
Full space-time slots	289,080	264 grid locations multiplied by 1095 days.
Valid CYGNSS observations	56,947	19.7% of all slots after quality filtering.
Valid SMAP observations	189,413	65.5% of the full domain.
Valid ERA5-Land observations	289,080	100% continuous reanalysis coverage.
Valid triplets	40,003	Primary ETC dataset.
Wet-season triplets	28,015	April to October.
Dry-season triplets	11,988	November to March.
Valid quadruplets	39,906	Dataset used for Quadruple Collocation.

CYGNSS sampling was not environmentally uniform. Figure 4 shows that SMAP maintained much higher temporal coverage than CYGNSS throughout 2021 to 2023. Mean SMAP coverage was about 65.4%, while mean CYGNSS coverage was about 19.6%. The resulting valid CYGNSS–SMAP triplet coverage averaged only 13.8%, showing that the ETC sample was constrained mainly by CYGNSS availability.

 doi:10.20944/preprints202605.0778.v1

This temporal sparsity also varied seasonally. Valid CYGNSS coverage was higher during the wet season than during the dry season, with 24.2% coverage during April to October and 13.3% during November to March. Figure 5 further shows that CYGNSS coverage was lower under dry soil conditions and during Harmattan months. This means that the dry stratum is likely weighted toward transitional dry cases rather than the most severe Harmattan dryness.

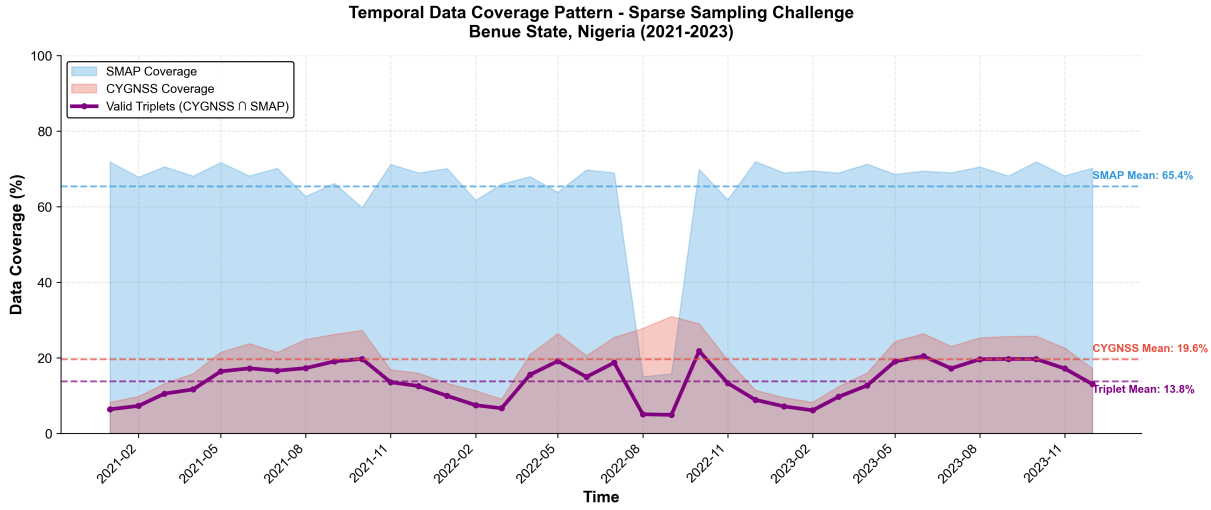

Figure 4. Temporal data coverage for SMAP, CYGNSS, and valid ETC triplets over Benue State from 2021 to 2023. The figure shows the sparse CYGNSS sampling challenge and the lower proportion of valid CYGNSS–SMAP triplets relative to SMAP coverage.

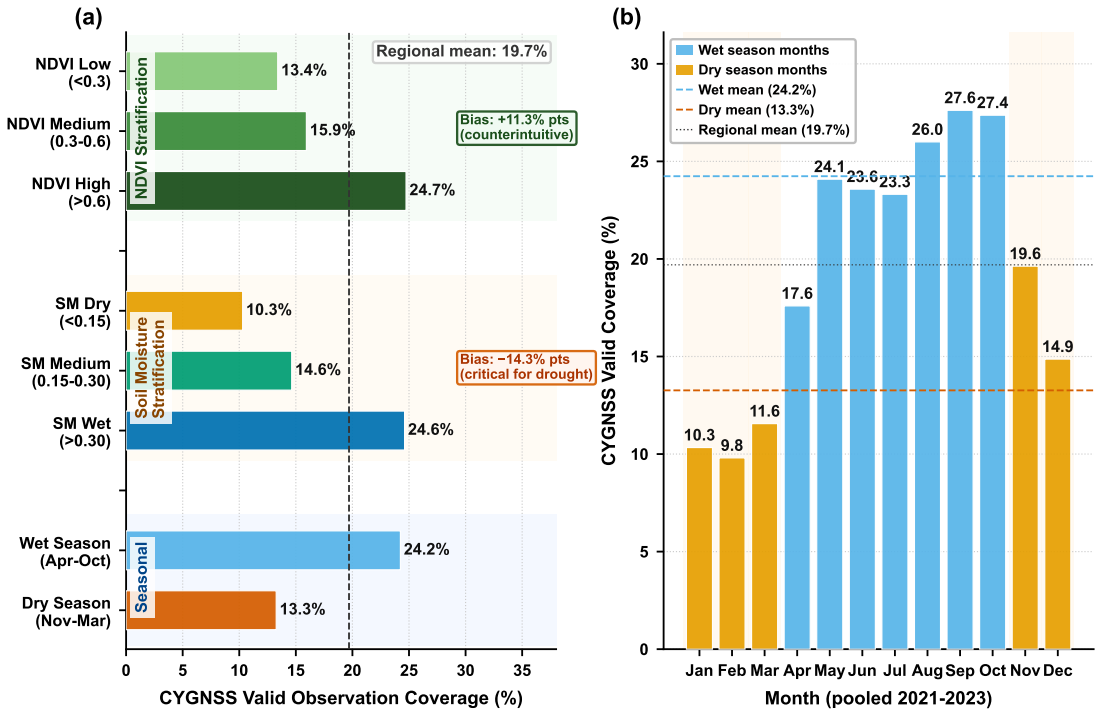

Figure 5. CYGNSS valid observation coverage stratified by environmental condition and calendar month over Benue State, Nigeria, from 2021 to 2023. Panel (a) shows coverage rates across NDVI, soil moisture, and seasonal strata relative to the regional mean of 19.7%. Panel (b) shows monthly coverage suppression during the dry season, especially from November to March, consistent with Harmattan-period signal collapse.

 doi:10.20944/preprints202605.0778.v1

3.2. Regional ETC Performance of CYGNSS, SMAP, and ERA5-Land

Regional ETC results show clear differences among the three primary products (Table 4). SMAP had the strongest performance, with $r = 0.760$, RMSE of $0.033\,m^3\,m^{-3}$, and SNR of $+1.37$ dB. It was the only signal-dominated product.

CYGNSS had intermediate performance. Its SMAP-inclusive ETC correlation was $r = 0.425$, with RMSE of $0.036\,m^3\,m^{-3}$ and SNR of -6.56 dB. ERA5-Land had the lowest RMSE ($0.024\,m^3\,m^{-3}$), but it had the weakest anomaly correlation ($r = 0.290$) and the most negative SNR (-10.37 dB). This pattern indicates that ERA5-Land captures the background moisture state but smooths short-term wetting and drying variability. Table 4 reports these metrics, while Figure 6 visualizes the regional ETC contrasts.

Table 4. Regional ETC performance for the primary CYGNSS-SMAP-ERA5-Land configuration.

Product	RMSE ($m^3\,m^{-3}$)	R	SNR (dB)	N
CYGNSS	0.036	0.425	-6.56	40,003
SMAP	0.033	0.760	+1.37	40,003
ERA5-Land	0.024	0.290	-10.37	40,003

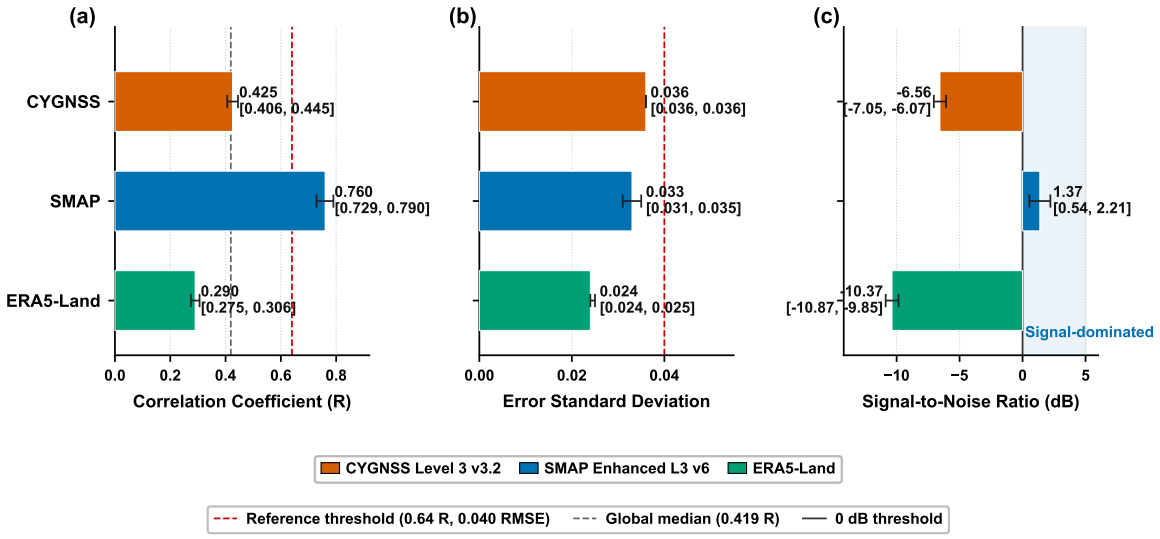

Figure 6. CYGNSS valid observation coverage stratified by environmental condition and calendar month over Benue State, Nigeria (2021 to 2023). (a) Coverage rates across NDVI, soil moisture, and seasonal strata relative to the regional mean (19.7%, dashed line). . (b) Monthly coverage pattern showing the pronounced dry season suppression from November to March, with peak coverage in October and the lowest values in February and December, consistent with the Harmattan moisture collapse documented in the seasonal ETC convergence analysis.

The pairwise anomaly structure underlying the ETC covariance decomposition is shown in Figure 7.

The regional result establishes two points. CYGNSS contains measurable soil moisture information in Guinea savanna agriculture. However, the negative SNR shows that the daily anomaly signal is noise-dominated. This matters because drought monitoring requires reliable tracking of temporal departures, not only low average error.

 doi:10.20944/preprints202605.0778.v1

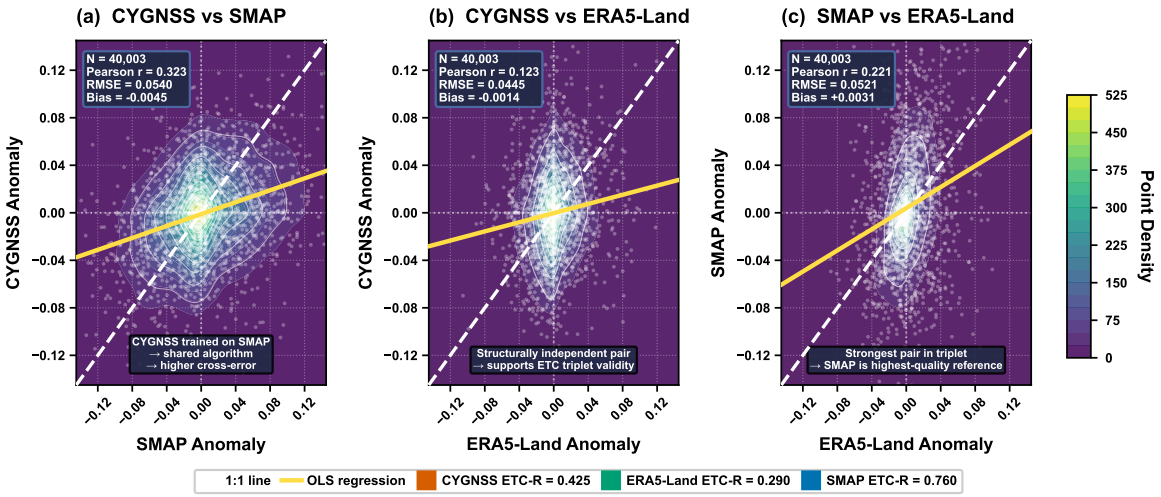

Figure 7. Pairwise kernel density scatter plots of 30-day soil moisture anomalies from 40,003 valid ETC triplets over Benue State from 2021 to 2023. Colour intensity shows local point density, the dashed white line marks the 1:1 reference, and the yellow line shows the OLS fit. Panels compare (a) CYGNSS-SMAP, (b) CYGNSS-ERA5-Land, and (c) SMAP-ERA5-Land, highlighting shared CYGNSS-SMAP dependence and the stronger SMAP-ERA5-Land agreement

3.3. Quadruple Collocation and SMAP-Dependence Correction

Quadruple Collocation directly quantified the CYGNSS-SMAP error dependence. The estimated cross-error correlation was 0.325, with a 95% bootstrap confidence interval of [0.307, 0.342]. This interval excludes zero and confirms significant error dependence. The result is consistent with the SMAP-trained structure of the CYGNSS Level 3 product.

The SMAP-inclusive CYGNSS estimate from Triplet 1 was $r = 0.425$. The SMAP-independent Triplet 3, which used CYGNSS, ERA5-Land, and ESA CCI ACTIVE, reduced the CYGNSS estimate to $r = 0.386$. The difference of $\Delta r = 0.039$ represents performance inflation associated with CYGNSS-SMAP dependence. The SMAP-independent estimate is therefore used as the conservative baseline for interpreting tropical CYGNSS skill.

Table 5. CYGNSS skill estimates from the ETC triplets and the Quadruple Collocation error-dependence test.

Analysis	Product configuration	Quantity reported	Estimate	Interpretation
T1	CYGNSS, SMAP, ERA5-Land	CYGNSS correlation with unknown truth	$R = 0.425$ [0.406, 0.445]	SMAP-inclusive estimate. This value is retained for comparison with earlier validation studies and treated as an upper bound.
T2	CYGNSS, SMAP, ESA CCI ACTIVE	CYGNSS correlation with unknown truth	$R = 0.387$ [0.376, 0.400]	Diagnostic configuration using an active microwave product as the third member of the triplet.
T3	CYGNSS, ERA5-Land, ESA CCI ACTIVE	CYGNSS correlation with unknown truth	$R = 0.386$ [0.366, 0.405]	SMAP-independent conservative baseline used for interpreting tropical CYGNSS skill.
QC	CYGNSS, SMAP, ERA5-Land, ESA CCI ACTIVE	CYGNSS-SMAP cross-error correlation	$r_\epsilon = 0.325$ [0.307, 0.342]	Confirms statistically significant CYGNSS-SMAP error dependence caused by shared SMAP training information.

This result is one of the central findings of the study. A standard SMAP-inclusive ETC triplet overstates CYGNSS performance. The SMAP-independent configuration gives a more defensible estimate of retrieval skill in this tropical agricultural environment.

 doi:10.20944/preprints202605.0778.v1

3.4. Environmental Controls on CYGNSS Performance

Environmental stratification shows that regional performance is not uniform. CYGNSS skill varies with vegetation density, soil moisture state, recent precipitation, and land cover (Table 6). These patterns explain why a single regional value cannot fully describe tropical retrieval performance. Pairwise bootstrap tests for the environmental contrasts are reported in Appendix A.2 and Table A2.

Table 6. CYGNSS ETC performance stratified by environmental controls.

Variable	Class	N	RMSE ($m^3\ m^{-3}$)	R	SNR (dB)
NDVI	Low	2,058	0.036	0.510	-4.55
NDVI	Medium	15,643	0.038	0.443	-6.12
NDVI	High	22,038	0.034	0.408	-7.00
Soil moisture	Dry	11,198	0.039	0.331	-9.11
Soil moisture	Medium	17,985	0.033	0.481	-5.21
Soil moisture	Wet	10,557	0.037	0.336	-8.95
Precipitation	Low	9,663	0.037	0.366	-8.10
Precipitation	Moderate	2,660	0.036	0.630	-1.83
Precipitation	High	15,037	0.037	0.427	-6.51
Land cover	Cropland	19,708	0.035	0.447	-6.02
Land cover	Shrubland/grassland	10,578	0.037	0.455	-5.83
Land cover	Tree cover	9,717	0.036	0.342	-8.79

3.4.1. Vegetation Density

CYGNSS performance declined with increasing NDVI. Correlation decreased from $r = 0.510$ under low NDVI to $r = 0.443$ under medium NDVI and $r = 0.408$ under high NDVI. SNR also declined from -4.55 dB to -7.00 dB. This monotonic pattern is physically consistent with increasing canopy attenuation and non-soil scattering contributions under dense vegetation [17,38].

3.4.2. Soil Moisture Regime

Soil moisture regime produced the clearest operational pattern. CYGNSS performed worst under dry soils, where $r = 0.331$ and SNR was -9.11 dB. Performance improved in the medium soil moisture range, where $r = 0.481$. It decreased again under wet conditions, where $r = 0.336$.

The dry-condition weakness is operationally important. It occurs when early detection of agricultural drought stress is most valuable. It may also be optimistic because severe dry-season conditions are under-sampled by valid CYGNSS retrievals.

3.4.3. Precipitation Regime

The precipitation stratification shows a distinct performance optimum under moderate recent rainfall. CYGNSS achieved $r = 0.630$ under 10 to 30 mm $week^{-1}$, compared with $r = 0.366$ under low precipitation and $r = 0.427$ under high precipitation. Moderate rainfall likely produces detectable wetting and drying dynamics without the surface water interference associated with heavy rainfall [39].

3.4.4. Land Cover Class

Land cover stratification separates agricultural and open savanna surfaces from persistent tree cover. Cropland yielded $r = 0.447$, and shrubland or grassland yielded $r = 0.455$. Tree cover performed substantially worse at $r = 0.342$. This result shows that CYGNSS is more promising over agricultural and open savanna surfaces than over persistent woodland. Table 6 reports the full stratified performance matrix, and Figure 8 summarizes the environmental contrasts visually.

 doi:10.20944/preprints202605.0778.v1

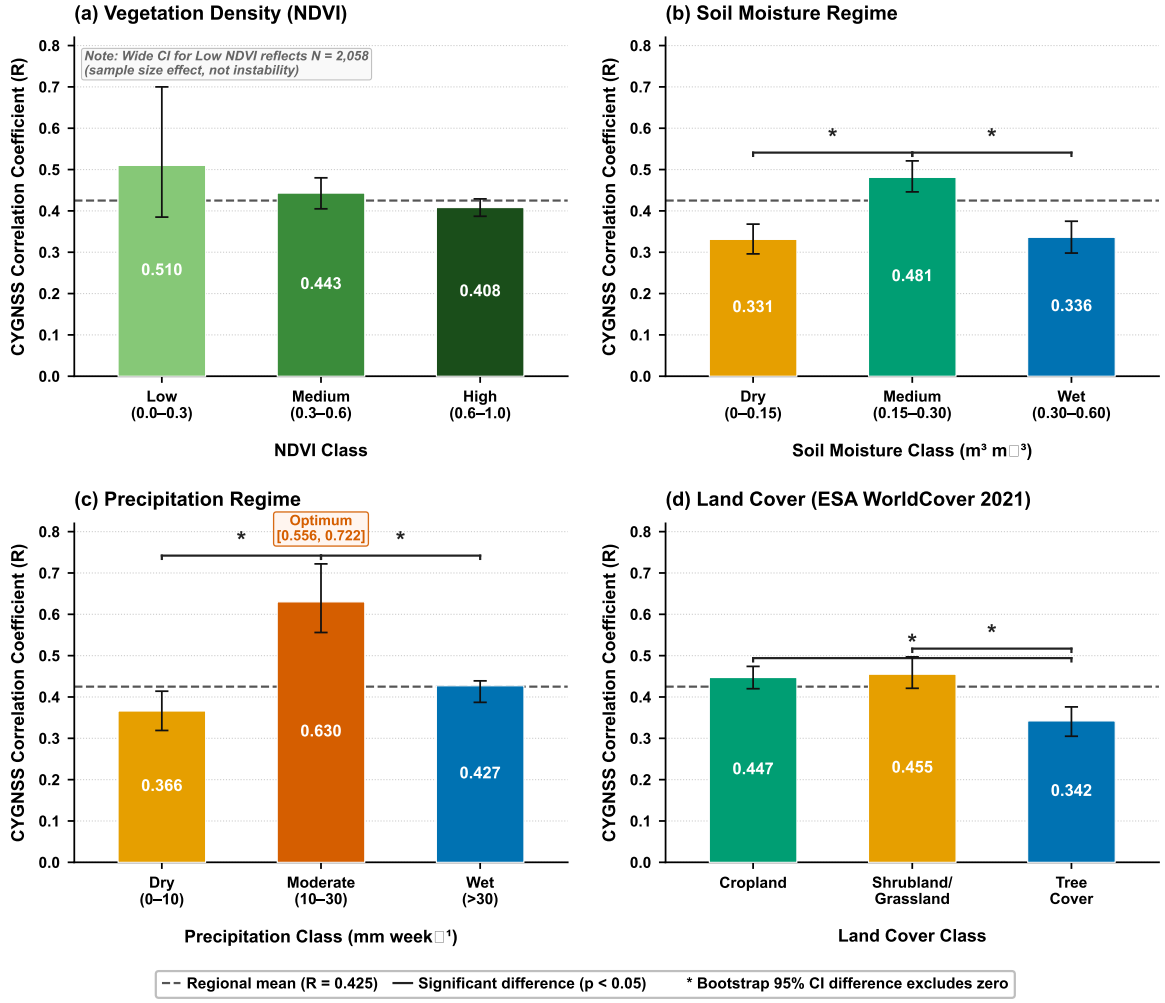

Figure 8. CYGNSS ETC-derived correlation stratified by vegetation density, soil moisture regime, seven-day precipitation, and land cover. Bars show point estimates with 95% bootstrap confidence intervals; the dashed line marks the regional mean (r=0.425), and asterisks indicate significant pairwise differences. Skill is highest under moderate precipitation and open agricultural surfaces, but weakest under dry soils and persistent tree cover.

3.5. Seasonal and Harmattan-Period Performance Failure

Seasonal diagnostics show that the dry season is the main operational weakness. Wet-season CYGNSS performance improved to $r = 0.497$ using 28,015 triplets. Dry-season location-level ETC failed to converge, despite 11,988 dry-season triplets at the pooled level. This failure reflects physical signal collapse rather than a simple numerical artifact.

Three factors explain the dry-season failure. First, CYGNSS valid coverage declines during the dry season. Second, SMAP soil moisture variability collapses under Harmattan conditions. Third, local monthly subsets often contain too few dynamic wetting and drying events to support covariance decomposition. In the dry season, 84.3% of SMAP observations fall below $0.15\,m^3\,m^{-3}$, and SMAP anomaly standard deviation declines by 36.2% compared with the wet season.

Monthly diagnostics show that the core wet-season months produced the most reliable ETC solutions. January, March, April, and December failed to converge, while February and November converged but produced very low correlations. These results show that CYGNSS is weakest during the period when early detection of agricultural drought stress is most important. Figure 9 illustrates the monthly convergence pattern and the Harmattan-period signal collapse.

 doi:10.20944/preprints202605.0778.v1

Table 7. Seasonal sampling and signal characteristics.

Metric	Wet season	Dry season
Season definition	April to October	November to March
Valid triplets	28,015	11,988
CYGNSS coverage	24.2%	13.3%
SMAP soil moisture mean	0.2591 $m^3\ m^{-3}$	0.0956 $m^3\ m^{-3}$
SMAP observations below 0.15 $m^3\ m^{-3}$	12.2%	84.3%
SMAP anomaly standard deviation	0.056576 $m^3\ m^{-3}$	0.036121 $m^3\ m^{-3}$
CYGNSS anomaly standard deviation	0.040889 $m^3\ m^{-3}$	0.037026 $m^3\ m^{-3}$
ETC result	$R = 0.497$	Failed at location level

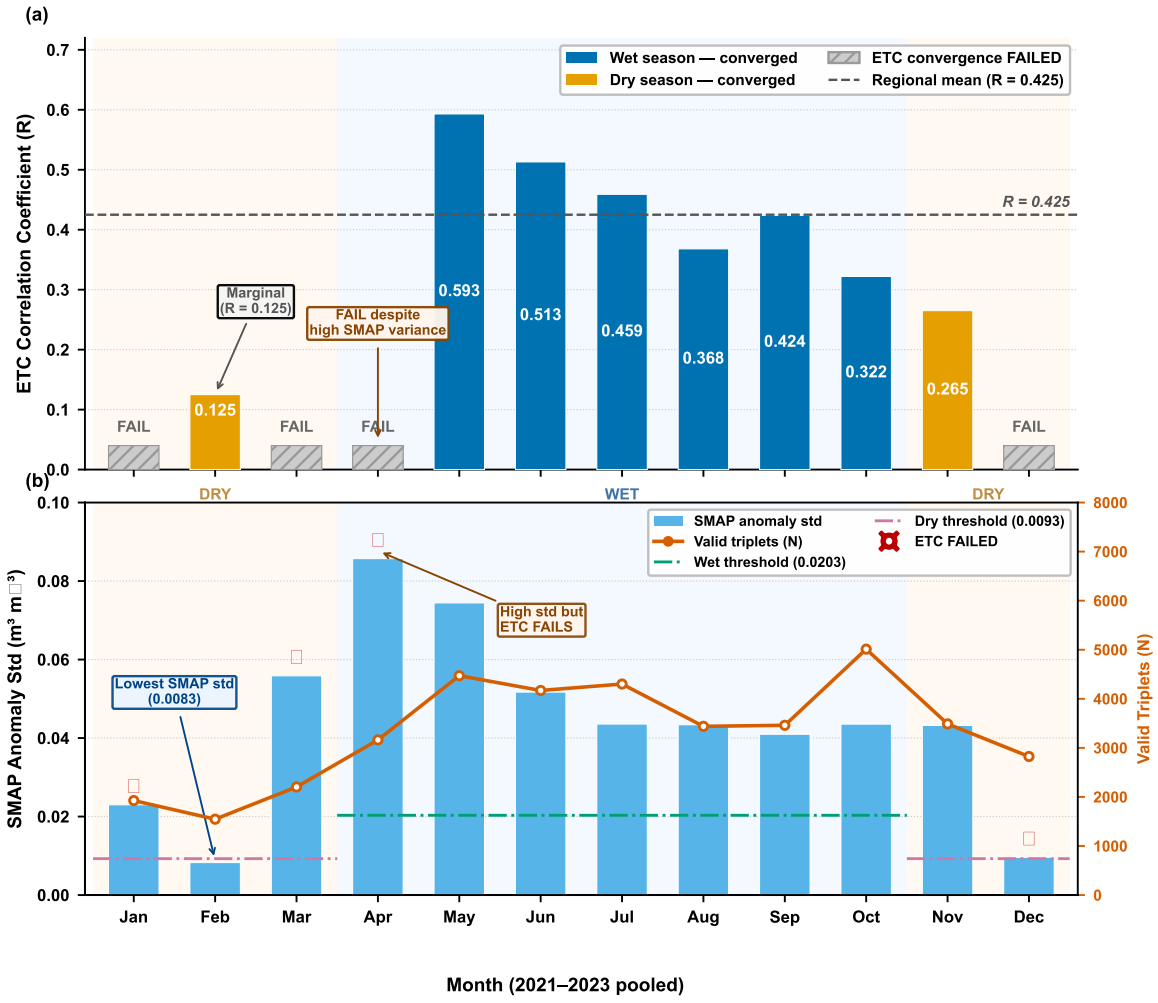

Figure 9. Monthly ETC convergence and signal variability over Benue State from 2021 to 2023. Panel (a) shows monthly CYGNSS ETC correlation, with hatched bars marking failed convergence and the dashed line showing the annual mean $r = 0.425$. Panel (b) shows monthly SMAP anomaly standard deviation and valid triplet counts. ETC performance is reliable mainly from May to October, while Harmattan and transition months show convergence failure or weak retrieval skill.

3.6. Sensitivity and Robustness of the ETC Results

Robustness tests confirm that the main conclusions are not artifacts of anomaly-window selection, sample size, bootstrap seed, or spatial dependence (Table 8). The 15-day anomaly window produced $r = 0.318$, the 30-day window produced $r = 0.425$, and the 60-day window produced $r = 0.446$. The 30-day window balances short-term event sensitivity with seasonal filtering.

Repeated subsampling from the full dataset to 25% of valid triplets produced a correlation range of only 0.419 to 0.432, with a standard deviation of 0.006. Bootstrap seed testing across 11 seeds

 doi:10.20944/preprints202605.0778.v1

produced a CYGNSS correlation range of 0.000744, with the full seed-level diagnostics provided in Appendix A.3 and Table A3. Spatial autocorrelation was statistically significant but modest. The spatially adjusted confidence interval remained below temperate validation benchmarks discussed in Section 4.1. Table 8 summarizes these robustness checks.

Table 8. Compact robustness summary for the primary CYGNSS ETC result. Full diagnostics are provided in Appendix A.

Test	Main result	Interpretation
Anomaly-window sensitivity	R changes from 0.318 at 15 days to 0.446 at 60 days.	The 30-day result is not arbitrary and preserves drought-relevant variability.
Sample-size stability	R remains within 0.419 to 0.432 from 100% to 25% subsampling.	The result is not a sample-size artifact.
Bootstrap seed independence	CYGNSS R range across 11 seeds is 0.000744.	Confidence intervals are stable across seeds.
Spatial autocorrelation	Moran's I is 0.041, and effective sample size is reduced.	Spatial adjustment does not change the interpretation.
Cross-validation	Random folds are stable, while temporal folds show seasonal variability.	The annual result is robust, and seasonal differences are physically meaningful.

Spatial robustness was examined both through the map and distribution of per-location ETC correlations in Figure 10 and through the spatial autocorrelogram in Figure 11

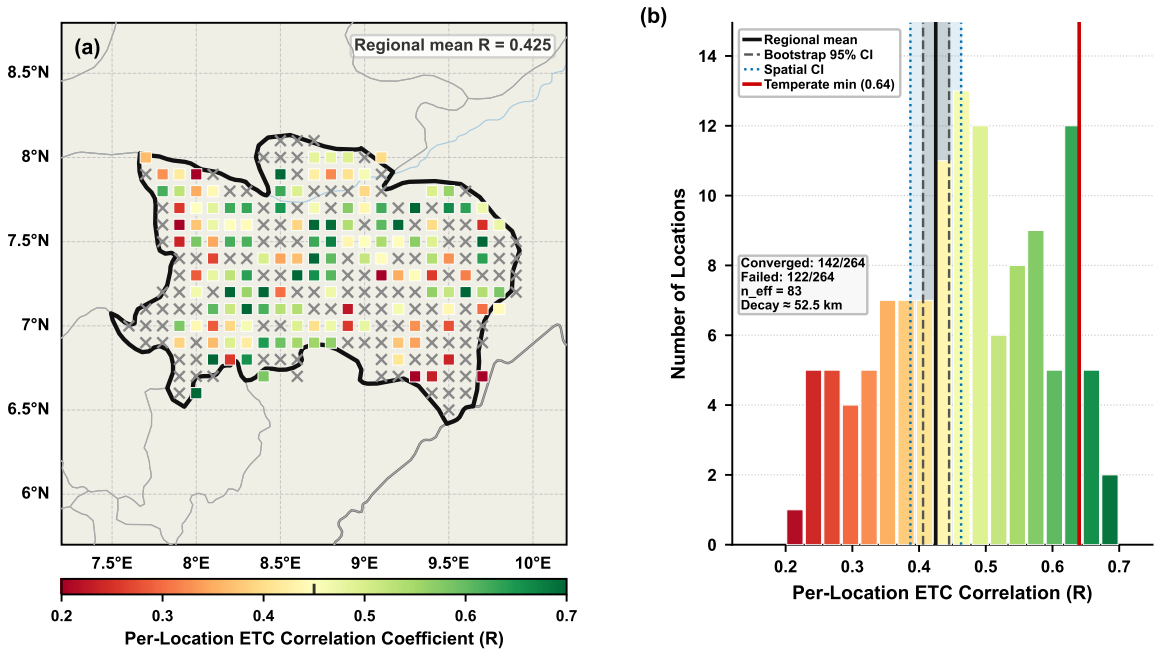

Figure 10. Monthly ETC convergence and signal variability over Benue State from 2021 to 2023. Panel (a) shows monthly CYGNSS ETC correlation, with hatched bars marking failed convergence and the dashed line showing the annual mean $r = 0.425$. Panel (b) shows monthly SMAP anomaly standard deviation and valid triplet counts. ETC performance is reliable mainly from May to October, while Harmattan and transition months show convergence failure or weak retrieval skill.

 doi:10.20944/preprints202605.0778.v1

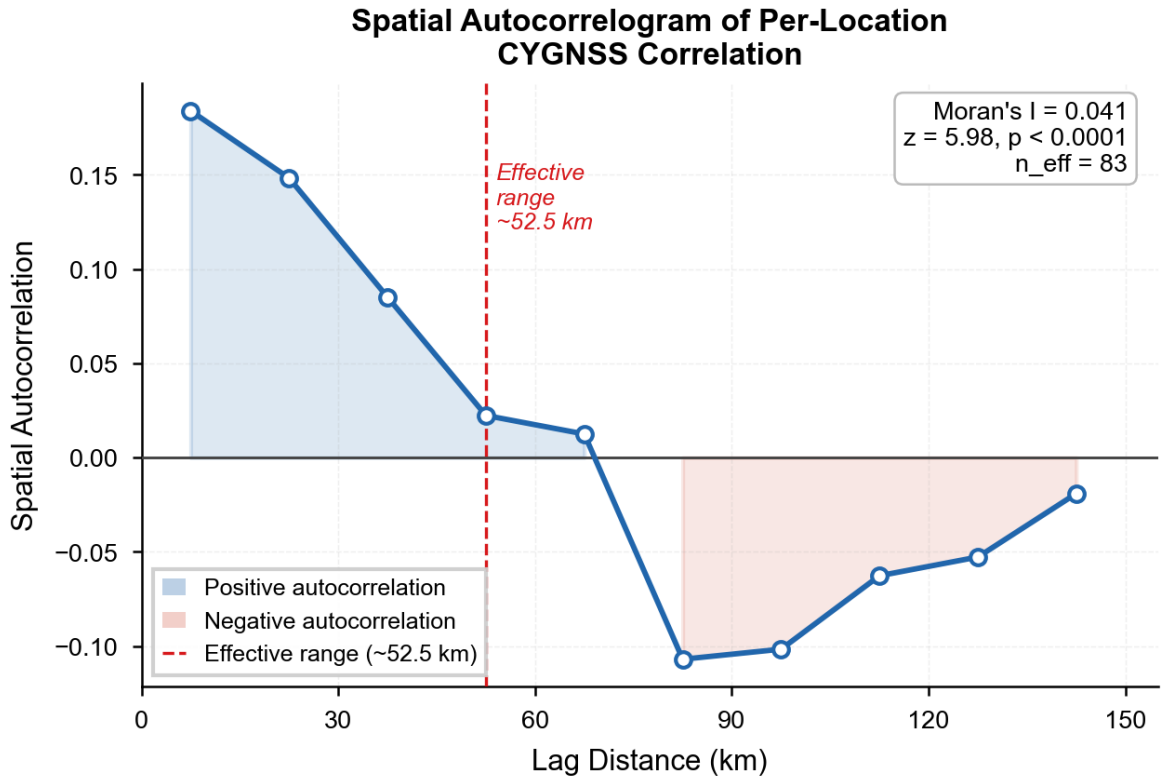

Figure 11. Spatial autocorrelogram of per-location CYGNSS ETC correlations across 159 converged grid locations in Benue State. Positive short-distance autocorrelation declines toward zero at an effective range of about 52.5 km. Moran's $I = 0.041$ confirms significant positive spatial autocorrelation, and the effective sample size is reduced to $n_{\text{eff}} = 83$ for spatially adjusted confidence intervals.

3.7. Cross-Validation Robustness

Three complementary cross-validation approaches were applied to test whether the primary ETC result generalizes beyond the full pooled dataset. The full-dataset CYGNSS correlation was $r = 0.4253$ for 40,003 valid triplets. Table 9 summarizes the random, spatial, and temporal cross-validation results.

Random 10-fold cross-validation partitioned the 40,003 valid triplets into ten approximately equal subsets. ETC was applied independently to each held-out fold. CYGNSS r ranged from 0.3932 to 0.4678, with a mean of 0.4261, standard deviation of 0.0253, and range of 0.0746. RMSE remained stable at $0.0360 \pm 0.0007\ \mathrm{m^3\ m^{-3}}$ across the random folds. The close agreement between the random-fold mean and the full-dataset estimate shows that the primary result is not an artifact of the specific sample composition.

Spatial 4-fold cross-validation divided Benue State into four geographic quadrants. Each quadrant was evaluated as a held-out test set while the remaining three quadrants provided the training pool. Test-fold CYGNSS R ranged from 0.4612 in the north-west quadrant to 0.5305 in the south-west quadrant, with a mean of 0.4851 and standard deviation of 0.0321. This consistency indicates that the primary result is not controlled by a single geographic subregion. The modest positive offset relative to the pooled result reflects the compositional effect of spatial partitioning, because geographically concentrated dry-season observations suppress the aggregate correlation when pooled.

Temporal cross-validation showed wider variability, as expected from the seasonal diagnostics. Leave-one-year-out validation produced R values of 0.4344 for 2021, 0.5430 for 2022, and 0.3606 for 2023. Temporal 5-fold validation produced values from 0.2680 to 0.6731. The higher temporal spread is physically meaningful. Folds dominated by active wet-season months perform better, while folds spanning dry-season or transition periods perform worse. This pattern independently confirms that seasonal soil moisture dynamics are the dominant driver of CYGNSS retrieval variation across Benue State.

 doi:10.20944/preprints202605.0778.v1

Table 9. Cross-validation robustness assessment for the primary CYGNSS ETC correlation. The full-dataset reference is $r = 0.4253$, with $N = 40{,}003$. Random, spatial, and temporal partitions evaluate generalization across sample composition, geography, and time.

CV approach	Folds	Held-out fold R values	Mean r	Std.	Range
Random 10-fold	10	0.3932, 0.4079, 0.4205, 0.4678, 0.4531, 0.4176, 0.3934, 0.4198, 0.4258, 0.4624	0.4261	0.0253	0.0746
Spatial 4-fold	4	NW: 0.4612; NE: 0.4635; SE: 0.4852; SW: 0.5305	0.4851	0.0321	0.0693
Temporal leave-one-year-out	3	2021: 0.4344; 2022: 0.5430; 2023: 0.3606	0.4460	0.0749	0.1824
Temporal 5-fold	5	0.4806, 0.3344, 0.6731, 0.4188, 0.2680	0.4350	0.1393	0.4051
Full dataset reference	–	Full pooled ETC sample	0.4253	–	–

Across all three approaches, random and spatial cross-validation confirm the statistical stability and geographic generalizability of the primary result. Temporal variability reflects genuine environmental seasonality and is consistent with the stratified analyses and convergence diagnostics. Together, the cross-validation tests show that the primary ETC result is robust, reproducible, and representative of CYGNSS retrieval behaviour in Guinea savanna tropical agriculture.

3.8. Summary of Key Results

CYGNSS Level 3 soil moisture has measurable but limited skill in Guinea savanna agriculture. SMAP-inclusive ETC inflates this skill because of shared CYGNSS-SMAP error dependence. Performance improves under moderate precipitation and open or agricultural land cover, but it weakens under dry soils, persistent tree cover, and Harmattan conditions. Robustness and cross-validation tests confirm that these conclusions are stable.

4. Discussion

4.1. Benchmarking Against Published CYGNSS and Soil Moisture Validation Studies

The SMAP-independent CYGNSS estimate of $r = 0.386$ is the most defensible measure of retrieval skill in this study. The SMAP-inclusive value of $r = 0.425$ is retained as an upper bound because it is more comparable to previous CYGNSS studies that include SMAP-related references. Both estimates are close to the global CYGNSS triple-collocation median reported by Deng et al. [26]. This suggests that the Guinea savanna result reflects a broader tropical limitation in the current CYGNSS soil moisture product.

Direct comparison with temperate in-situ validation studies requires caution. ETC-derived correlation with an unknown truth is not the same metric as Pearson correlation against in-situ sensors [13,14]. Still, the magnitude of the gap is informative. Temperate CYGNSS validation studies have reported correlations from about 0.64 to above 0.80 under some algorithms and validation settings [22,23,25]. Even the SMAP-inclusive upper bound in this study remains below those benchmarks.

SMAP's performance provides useful context. SMAP remained signal-dominated and achieved $r = 0.760$, which is consistent with its established role as a high-quality L-band soil moisture product [18,20]. The contrast between SMAP and CYGNSS does not remove the value of CYGNSS. It shows that CYGNSS needs regional correction before its high revisit frequency can be used reliably for agricultural drought monitoring.

4.2. Why CYGNSS Degrades in Guinea Savanna Agriculture

The degradation is not random. It is linked to physical and environmental controls. Vegetation affects the propagation and scattering of L-band reflected signals. Dense crop or woodland canopies attenuate the soil contribution and increase the influence of vegetation structure [17,38]. This explains the monotonic NDVI pattern and the weaker performance over tree cover.

Dry soils create a different limitation. GNSS-R retrieval depends on dielectric contrast between soil and air [16]. Under very dry conditions, soil moisture variability can become small, and the

 doi:10.20944/preprints202605.0778.v1

retrievable signal is reduced. This problem is strongest during Harmattan months, when moisture variability collapses and the collocation system fails at local scales.

Wet conditions can also reduce performance. High rainfall may introduce surface water, ponding, roughness changes, and mixed scattering responses [39]. The high precipitation stratum performed better than the dry stratum but worse than the moderate precipitation stratum. This supports the interpretation that CYGNSS performs best when soil moisture is changing dynamically but not dominated by extreme wetness or extreme dryness.

Mixed pixels further complicate retrieval. Cropland, shrubland, grassland, and tree cover can coexist within a single retrieval footprint. Such heterogeneity weakens the relationship between grid-cell average soil moisture and the scattering conditions sampled by individual CYGNSS specular points.

4.3. Why RMSE Alone Is Misleading

CYGNSS achieved an RMSE of $0.036\,m^3\,m^{-3}$, which is close to the commonly cited $0.04\,m^3\,m^{-3}$ accuracy target for L-band soil moisture missions [18,20]. This appears encouraging, but RMSE alone does not establish operational suitability for drought monitoring.

The negative SNR of -6.56 dB shows that noise exceeds the retrievable daily anomaly signal. This is the critical finding. A product can have acceptable average error while still failing to track short-term anomalies reliably. Drought early warning requires detection of onset, persistence, and recovery. These tasks depend on signal reliability, not only on low average deviation from an estimated truth.

ERA5-Land illustrates the same issue in another way. It had the lowest RMSE but the weakest anomaly correlation. This suggests that it captures the background seasonal state but not the daily wetting and drying variability needed for short-term drought applications.

4.4. Implications of CYGNSS-SMAP Cross-Error Dependence

The CYGNSS-SMAP cross-error correlation of 0.325 is one of the main methodological findings. It shows that including SMAP in the same ETC triplet as CYGNSS can overstate CYGNSS performance. This dependence is expected because CYGNSS Level 3 is trained against SMAP [21]. The result confirms that the issue is measurable, not only theoretical.

This has implications beyond this case study. Future validation of trained satellite products should consider whether validation references also appear in the training chain. When overlap exists, standard triplet assumptions can be weakened. Quadruple Collocation provides a practical way to quantify this issue when a fourth structurally independent product is available [14,41].

4.5. Operational Implications for Drought Monitoring

The results do not imply that CYGNSS has no value for drought monitoring. Its high revisit frequency remains attractive, especially in the tropics where rainfall events are frequent and localized [15]. The results instead show that uncorrected CYGNSS Level 3 soil moisture should not be used alone for year-round drought monitoring in Guinea savanna agriculture.

Operational use should be most cautious during dry and Harmattan periods. The dry-condition correlation of $r = 0.331$ is low, and severe dry cases are under-sampled. The true performance during the most extreme dryness may therefore be worse than the dry-stratum estimate suggests. Early-warning systems that rely directly on uncorrected CYGNSS anomalies could miss or misrepresent drought onset.

CYGNSS may be more valuable in multi-sensor systems. Its revisit frequency can complement SMAP's accuracy, ERA5-Land's continuity, precipitation observations, and vegetation information. In such systems, CYGNSS should be used with quality weighting, seasonal flags, and environmental bias correction.

 doi:10.20944/preprints202605.0778.v1

4.6. What Bias Correction Must Target

Bias correction should target the controls identified by the stratification analysis. Vegetation information is essential because retrieval skill declines with NDVI and tree cover. NDVI can be used, but vegetation optical depth or canopy water content may provide a more physically direct correction where available. Land cover should also be included because cropland and woodland can share similar NDVI during parts of the growing season.

Soil moisture regime must be included because performance changes strongly between dry, medium, and wet states. A single global correction function may not work across the full moisture range. The dry end requires special attention because it is operationally most important and most poorly sampled. Precipitation history should also be included because moderate rainfall produced the best performance, while low and high rainfall regimes had weaker skill.

Seasonal or Harmattan indicators should be considered for West African applications. The dry season is not only a low-moisture period. It is also a period of reduced signal variance, lower CYGNSS coverage, and local convergence failure. Regional recalibration should therefore include variables that identify Harmattan conditions, recent rainfall history, vegetation state, and land cover.

4.7. Limitations and Future Work

Several limitations should guide interpretation. First, there are no in-situ soil moisture stations in the study area that can provide direct ground validation. This is the reason ETC was used, but it also means that the study estimates correlation with an unknown statistical truth rather than direct agreement with field sensors.

Second, ETC relies on assumptions about linearity, stationarity, and error independence [13,14]. Quadruple Collocation directly addresses the most important independence concern, the CYGNSS-SMAP relationship, but residual dependence among other products cannot be ruled out completely. Third, land cover classes are assigned at 0.1° resolution using majority vote, which simplifies a heterogeneous agricultural woodland mosaic.

Fourth, CYGNSS sampling is uneven across season and soil moisture state. The dry stratum is under-sampled relative to wet conditions, and this likely makes the dry-condition performance estimate optimistic. Future work should combine field campaigns, low-cost in-situ sensors, regionally calibrated retrieval models, and independent validation sites across West African cropland, savanna, Sahelian drylands, and humid forest margins.

5. Conclusions

This study evaluated CYGNSS Level 3 soil moisture performance in Guinea savanna agriculture using Extended Triple Collocation and Quadruple Collocation. The results show that CYGNSS has measurable but limited skill in Benue State, Nigeria.

The SMAP-inclusive ETC estimate gives $r = 0.425$, but Quadruple Collocation reveals a CYGNSS-SMAP cross-error correlation of 0.325. The SMAP-independent estimate is lower at $r = 0.386$ and is the more defensible measure of tropical CYGNSS skill.

The main operational weakness is dry-season performance. CYGNSS performs poorly under dry soils ($r = 0.331$), and location-level ETC convergence fails during Harmattan conditions as soil moisture anomaly variance collapses. This limitation occurs when drought monitoring is most critical for rainfed smallholder agriculture.

Environmental stratification shows that performance is controlled by vegetation, soil moisture regime, precipitation, and land cover. CYGNSS performs better over cropland and shrubland or grassland than over tree cover, and it performs best under moderate precipitation. These patterns show that the product has agricultural value, but only after correction for environmental conditions.

Uncorrected CYGNSS Level 3 soil moisture is therefore not sufficient for standalone year-round drought monitoring in Guinea savanna agriculture. Its high revisit frequency is useful, but operational

 doi:10.20944/preprints202605.0778.v1

application should rely on bias-corrected multi-sensor systems that include vegetation state, soil moisture regime, precipitation history, land cover, and seasonal Harmattan indicators.

Author Contributions: Conceptualization, A.S.O. and S.T.I.; methodology, A.S.O. and S.T.I.; software, A.S.O.; validation, A.S.O., S.T.I., C.K. and A.M.A.-S.; formal analysis, A.S.O.; investigation, A.S.O.; resources, S.T.I.; data curation, A.S.O.; writing-original draft preparation, A.S.O.; writing-review and editing, S.T.I., C.K. and A.M.A.-S.; visualization, A.S.O.; supervision, S.T.I.; project administration, S.T.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets analyzed in this study are publicly available from NASA PO.DAAC, NASA NSIDC, ECMWF Climate Data Store, NASA LP DAAC, CEDA, and ESA WorldCover. Processed collocation tables and analysis scripts can be made available by the corresponding author upon reasonable request.

Acknowledgments: The authors acknowledge NASA PO.DAAC, NASA NSIDC, ECMWF, NASA LP DAAC, CEDA, EUMETSAT H-SAF, ESA, and Copernicus for providing the open datasets used in this study.

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

Abbreviations

The following abbreviations are used in this manuscript:

ASCAT	Advanced Scatterometer
CCI	Climate Change Initiative
CYGNSS	Cyclone Global Navigation Satellite System
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA5-Land	ECMWF land surface reanalysis
ETC	Extended Triple Collocation
GNSS-R	Global Navigation Satellite System Reflectometry
NDVI	Normalized Difference Vegetation Index
RMSE	Root mean square error
SMAP	Soil Moisture Active Passive
SMOS	Soil Moisture and Ocean Salinity
SNR	Signal-to-noise ratio

Appendix A. Supplementary Robustness Results

Appendix A.1. Full Anomaly-Window Sensitivity

Table A1. Sensitivity of CYGNSS ETC metrics to anomaly-window length.

Window (days)	RMSE ($m^3\ m^{-3}$)	R	SNR (dB)	Interpretation
15	0.031	0.318	-9.48	Shorter event window with lower temporal stability.
30	0.036	0.425	-6.56	Main configuration, balancing event response and seasonal filtering.
60	0.040	0.446	-6.04	Smoother anomaly definition with greater climatological stability.

 doi:10.20944/preprints202605.0778.v1

Appendix A.2. Pairwise Bootstrap Significance Tests Across Environmental Strata

Table A2. Pairwise bootstrap significance tests across environmental strata. Significance is declared when the 95% bootstrap confidence interval of the pairwise difference, $R_A - R_B$, excludes zero. The confidence intervals report bootstrapped difference intervals, not individual stratum correlation intervals.

Stratification	Comparison	R_A	R_B	CI lower	CI upper	Significant
NDVI	Low vs. medium	0.510	0.443	-0.064	+0.259	No
NDVI	Low vs. high	0.510	0.408	-0.022	+0.297	No
NDVI	Medium vs. high	0.443	0.408	-0.009	+0.077	No
Soil moisture	Dry vs. medium	0.331	0.481	-0.204	-0.097	Yes
Soil moisture	Dry vs. wet	0.331	0.336	-0.055	+0.054	No
Soil moisture	Medium vs. wet	0.481	0.336	+0.092	+0.197	Yes
Precipitation	Low vs. moderate	0.366	0.630	-0.358	-0.173	Yes
Precipitation	Low vs. high	0.366	0.427	-0.098	+0.005	No
Precipitation	Moderate vs. high	0.630	0.427	+0.141	+0.307	Yes
Land cover	Cropland vs. tree cover	0.447	0.342	+0.061	+0.153	Yes
Land cover	Cropland vs. shrubland or grassland	0.447	0.455	-0.057	+0.040	No
Land cover	Shrubland or grassland vs. tree cover	0.455	0.342	+0.059	+0.167	Yes

Non-significant NDVI comparisons reflect the limited sample size in the low-vegetation stratum. Significant land-cover comparisons confirm a statistically distinct retrieval penalty under persistent tree cover.

Appendix A.3. Full Bootstrap Seed-Independence Diagnostics

Table A3. Bootstrap seed-independence assessment for CYGNSS and SMAP ETC correlations. Each seed used 1000 bootstrap iterations on 40,003 valid triplets with a 30-day anomaly window. The CYGNSS R range across all 11 seeds is 0.000744, and the SMAP R range is 0.0018. Both are below the conventional stability threshold of 0.005.

Seed	CYG. r	CYG. CI	CYG. RMSE ($m^3\,m^{-3}$)	CYG. SNR (dB)	SMAP r	SMAP CI	SMAP RMSE ($m^3\,m^{-3}$)	Fail. iterations
0	0.4250	[0.4064, 0.4451]	0.0360	-6.568	0.7613	[0.7312, 0.7930]	0.0333	0
1	0.4256	[0.4068, 0.4447]	0.0360	-6.553	0.7603	[0.7318, 0.7922]	0.0333	0
5	0.4251	[0.4055, 0.4435]	0.0360	-6.567	0.7617	[0.7312, 0.7950]	0.0332	0
10	0.4256	[0.4073, 0.4450]	0.0360	-6.553	0.7610	[0.7284, 0.7928]	0.0333	0
24	0.4250	[0.4053, 0.4437]	0.0360	-6.570	0.7610	[0.7319, 0.7946]	0.0333	0
28	0.4255	[0.4058, 0.4451]	0.0360	-6.556	0.7600	[0.7277, 0.7943]	0.0333	0
33	0.4257	[0.4076, 0.4446]	0.0360	-6.551	0.7604	[0.7314, 0.7913]	0.0333	0
39	0.4253	[0.4050, 0.4446]	0.0360	-6.561	0.7605	[0.7307, 0.7909]	0.0333	0
40	0.4257	[0.4063, 0.4445]	0.0360	-6.551	0.7600	[0.7291, 0.7909]	0.0333	0
41	0.4254	[0.4068, 0.4439]	0.0360	-6.558	0.7606	[0.7318, 0.7908]	0.0333	0
42	**0.4253**	**[0.4061, 0.4451]**	**0.0360**	**-6.560**	**0.7599**	**[0.7289, 0.7902]**	**0.0333**	**0**
Mean	0.4254	N/A	0.0360	N/A	0.7606	N/A	0.0333	N/A
Std.	0.000276	N/A	N/A	N/A	0.000553	N/A	N/A	N/A
Range	0.000744	N/A	N/A	N/A	0.0018	N/A	N/A	N/A

Zero failed iterations across all seeds confirm numerical stability throughout. Seed 42 is the primary bootstrap seed used for the reported confidence intervals.

 doi:10.20944/preprints202605.0778.v1

Appendix A.4. Sample-Size Stability and Repeated Subsampling

Table A4. Sample-size stability summary.

Test	CYGNSS R behavior	Conclusion
Full dataset	0.425	Reference configuration.
75%, 50%, and 25% repeated subsampling	Range of 0.419 to 0.432	The primary result is stable under substantial sample reduction.
Random 25% subset	Approximately 9935 records per draw	The result is not driven by full-sample size alone.

Appendix A.5. Spatial Autocorrelation and Effective Sample Size

Table A5. Spatial autocorrelation diagnostics for location-level CYGNSS ETC correlations.

Metric	Value	Interpretation
Converged grid locations	159 of 264	Location-level ETC solutions were available for most grid cells.
Moran's I	0.041	Modest positive spatial autocorrelation.
Effective sample size	83	Clifford correction reduces the nominal number of converged locations.
Spatially adjusted 95% CI	[0.387, 0.463]	Wider than bootstrap CI but still below temperate benchmarks.

Appendix A.6. Monthly ETC Convergence Diagnostics

Table A6. Monthly ETC convergence status and SMAP anomaly variability.

Month	Season	Triplets	SMAP anomaly std. ($m^3\ m^{-3}$)	ETC result
January	Dry	1,926	0.022987	FAIL
February	Dry	1,546	0.008267	R = 0.125
March	Dry	2,203	0.055850	FAIL
April	Wet	3,162	0.085705	FAIL
May	Wet	4,471	0.074422	R = 0.593
June	Wet	4,170	0.051704	R = 0.513
July	Wet	4,301	0.043560	R = 0.459
August	Wet	3,439	0.043380	R = 0.368
September	Wet	3,460	0.040963	R = 0.424
October	Wet	5,012	0.043543	R = 0.322
November	Dry	3,489	0.043211	R = 0.265
December	Dry	2,824	0.009615	FAIL

Appendix A.7. Additional Triplet and Quadruple-Collocation Configurations

Table A7. Additional product-level results for the main triplet configurations.

Triplet	Product	R	RMSE ($m^3\ m^{-3}$)	SNR (dB)
T1	CYGNSS	0.425	0.0360	-6.62
T1	SMAP	0.758	0.0334	+1.32
T1	ERA5-Land	0.287	0.0244	-10.47
T2	CYGNSS	0.387	0.0366	-7.54
T2	SMAP	0.828	0.0287	+3.40
T2	ESA CCI ACTIVE	0.683	0.0810	-0.58
T3	CYGNSS	0.386	0.0367	-7.56
T3	ERA5-Land	0.294	0.0244	-10.20
T3	ESA CCI ACTIVE	0.597	0.0895	-2.11

Note: T1 = CYGNSS, SMAP, ERA5-Land; T2 = CYGNSS, SMAP, ESA CCI ACTIVE; T3 = CYGNSS, ERA5-Land, ESA CCI ACTIVE.

 doi:10.20944/preprints202605.0778.v1

References

1. Seneviratne, S.I.; Corti, T.; Davin, E.L.; Hirschi, M.; Jaeger, E.B.; Lehner, I.; Orlowsky, B.; Teuling, A.J. Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Sci. Rev.* **2010**, *99*, 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004.
2. Dorigo, W.; Wagner, W.; Albergel, C.; Albrecht, F.; Balsamo, G.; Brocca, L.; Chung, D.; Ertl, M.; Forkel, M.; Gruber, A.; et al. ESA CCI Soil Moisture for improved Earth system understanding: State-of-the-art and future directions. *Remote Sens. Environ.* **2017**, *203*, 185–215. https://doi.org/10.1016/j.rse.2017.07.001.
3. Bolten, J.D.; Crow, W.T.; Zhan, X.; Jackson, T.J.; Reynolds, C.A. Evaluating the utility of remotely sensed soil moisture retrievals for operational agricultural drought monitoring. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2010**, *3*, 57–66. https://doi.org/10.1109/JSTARS.2009.2037163.
4. Ford, T.W.; Harris, E.; Quiring, S.M. Estimating root zone soil moisture using near-surface observations from SMOS. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 139–154. https://doi.org/10.5194/hess-18-139-2014.
5. Dorigo, W.A.; Wagner, W.; Hohensinn, R.; Hahn, S.; Paulik, C.; Xaver, A.; Gruber, A.; Drusch, M.; Mecklenburg, S.; van Oevelen, P.; et al. The International Soil Moisture Network: A data hosting facility for global in situ soil moisture measurements. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 1675–1698. https://doi.org/10.5194/hess-15-1675-2011.
6. Yakubu, C.I.; Ayer, J.; Laari, P.B.; Amponsah, T.Y.; Hancock, C.M. A mutual assessment of the uncertainties of digital elevation models using the triple collocation technique. *Int. J. Remote Sens.* **2019**, *40*, 5076–5093. https://doi.org/10.1080/01431161.2019.1579388.
7. Tyubee, B.T. An analysis of food crop yields and climate relations in Benue State, Nigeria. *J. Niger. Meteorol. Soc.* **2006**, *6*, 13–22.
8. Terdoo, F.; Adekola, O.F. Assessing the role of climate in the dynamics of food crop production in Benue State, North-Central Nigeria. *Afr. Geogr. Rev.* **2014**, *33*, 142–158. https://doi.org/10.1080/19376812.2014.945674.
9. Ater, P.I.; Aye, R.N.; Vange, T.; Ikyume, J.T. Climate variability and smallholder farmers' adaptation in Benue State, Nigeria. *Afr. J. Agric. Res.* **2018**, *13*, 770–779.
10. Onyeneke, R.U.; Olayide, O.E.; Tasie, O.; Emenekwe, C.C.; Enyikwola, E.; Peter, H.H.; Deborah, M.H. *Benue State Climate-Smart Agriculture (CSA) Profile*; NAPA Research Paper No. 7; Feed the Future Nigeria Agricultural Policy Activity: Washington, DC, USA, 2023.
11. Dorigo, W.; Himmelbauer, I.; Aberer, D.; Schremmer, L.; Petrakovic, I.; Zappa, L.; Preimesberger, W.; Xaver, A.; Annor, F.; Ardö, J.; et al. The International Soil Moisture Network: Serving Earth system science for over a decade. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 5749–5804. https://doi.org/10.5194/hess-25-5749-2021.
12. Stoffelen, A. Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *J. Geophys. Res. Oceans* **1998**, *103*, 7755–7766. https://doi.org/10.1029/97JC03180.
13. McColl, K.A.; Vogelzang, J.; Konings, A.G.; Entekhabi, D.; Piles, M.; Stoffelen, A. Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target. *Geophys. Res. Lett.* **2014**, *41*, 6229–6236. https://doi.org/10.1002/2014GL061322.
14. Gruber, A.; Su, C.H.; Zwieback, S.; Crow, W.; Dorigo, W.; Wagner, W. Recent advances in soil moisture triple collocation analysis. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *45*, 200–211. https://doi.org/10.1016/j.jag.2015.09.002.
15. Ruf, C.S.; Chew, C.; Lang, T.; Morris, M.G.; Nave, K.; Ridley, A.; Balasubramaniam, R. A new paradigm in Earth environmental monitoring with the CYGNSS small satellite constellation. *Sci. Rep.* **2018**, *8*, 8782. https://doi.org/10.1038/s41598-018-27127-4.
16. Ulaby, F.T.; Moore, R.K.; Fung, A.K. *Microwave Remote Sensing: Active and Passive. Volume 3: From Theory to Applications*; Artech House: Norwood, MA, USA, 1986.
17. Chew, C.; Small, E. Soil moisture sensing using spaceborne GNSS reflections: Comparison of CYGNSS reflectivity to SMAP soil moisture. *Geophys. Res. Lett.* **2018**, *45*, 4049–4057. https://doi.org/10.1029/2018GL077905.
18. Entekhabi, D.; Njoku, E.G.; O'Neill, P.E.; Kellogg, K.H.; Crow, W.T.; Edelstein, W.N.; Entin, J.K.; Goodman, S.D.; Jackson, T.J.; Johnson, J.; et al. The Soil Moisture Active Passive (SMAP) mission. *Proc. IEEE* **2010**, *98*, 704–716. https://doi.org/10.1109/JPROC.2010.2043918.
19. Kerr, Y.H.; Waldteufel, P.; Richaume, P.; Wigneron, J.P.; Ferrazzoli, P.; Mahmoodi, A.; Al Bitar, A.; Cabot, F.; Gruhier, C.; Juglea, S.E.; et al. The SMOS soil moisture retrieval algorithm. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1384–1403. https://doi.org/10.1109/TGRS.2012.2184548.

 doi:10.20944/preprints202605.0778.v1

20. Colliander, A.; Jackson, T.J.; Bindlish, R.; Chan, S.; Das, N.; Kim, S.B.; Cosh, M.H.; Dunbar, R.S.; Dang, L.; Pashaian, L.; et al. Validation of SMAP surface soil moisture products with core validation sites. *Remote Sens. Environ.* **2017**, *191*, 215–231. https://doi.org/10.1016/j.rse.2017.01.021.
21. Chew, C.; Small, E. Description of the UCAR/CU soil moisture product. *Remote Sens.* **2020**, *12*, 1558. https://doi.org/10.3390/rs12101558.
22. Eroglu, O.; Kurum, M.; Boyd, D.; Gurbuz, A.C. High spatio-temporal resolution CYGNSS soil moisture estimates using artificial neural networks. *Remote Sens.* **2019**, *11*, 2272. https://doi.org/10.3390/rs11192272.
23. Al-Khaldi, M.M.; Johnson, J.T.; O'Brien, A.J.; Balenzano, A.; Mattia, F. Time-series retrieval of soil moisture using CYGNSS. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4322–4331. https://doi.org/10.1109/TGRS.2018.2890646.
24. Clarizia, M.P.; Pierdicca, N.; Costantini, F.; Floury, N. Analysis of CYGNSS data for soil moisture retrieval. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2227–2235. https://doi.org/10.1109/JSTARS.2019.2895510.
25. Senyurek, V.; Lei, F.; Boyd, D.; Kurum, M.; Gurbuz, A.C.; Moorhead, R. Machine learning-based CYGNSS soil moisture estimates over ISMN sites in CONUS. *Remote Sens.* **2020**, *12*, 1168. https://doi.org/10.3390/rs12071168.
26. Deng, X.; Zhu, L.; Wang, H.; Zhang, X.; Tong, C.; Li, S.; Wang, K. Triple collocation analysis and in situ validation of the CYGNSS soil moisture product. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 1883–1899. https://doi.org/10.1109/JSTARS.2023.3235111.
27. Lei, F.; Senyurek, V.; Kurum, M.; Gurbuz, A.C.; Boyd, D.; Moorhead, R.; Crow, W.T.; Eroglu, O. Quasi-global machine learning-based soil moisture estimates at high spatio-temporal scales using CYGNSS and SMAP observations. *Remote Sens. Environ.* **2022**, *276*, 113041. https://doi.org/10.1016/j.rse.2022.113041.
28. Udo, R.K. *Geographical Regions of Nigeria*; University of California Press: Berkeley, CA, USA, 1970.
29. Peel, M.C.; Finlayson, B.L.; McMahon, T.A. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 1633–1644. https://doi.org/10.5194/hess-11-1633-2007.
30. Oguntunde, P.G.; Friesen, J.; van de Giesen, N.; Savenije, H.H.G. Hydroclimatology of the Volta River Basin in West Africa: Trends and variability from 1901 to 2002. *Phys. Chem. Earth* **2006**, *31*, 1180–1188. https://doi.org/10.1016/j.pce.2006.02.062.
31. NiMet. *2023 Seasonal Climate Prediction (SCP)*; Nigerian Meteorological Agency: Abuja, Nigeria, 2023. Available online: http://ngfrepository.org.ng:8080/jspui/handle/123456789/5461 (accessed on 15 January 2026).
32. Chan, S.K.; Bindlish, R.; O'Neill, P.; Jackson, T.; Njoku, E.; Dunbar, S.; Chaubell, J.; Piepmeier, J.; Yueh, S.; Entekhabi, D.; et al. Development and assessment of the SMAP enhanced passive soil moisture product. *Remote Sens. Environ.* **2018**, *204*, 931–941. https://doi.org/10.1016/j.rse.2017.08.025.
33. Muñoz-Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H.; et al. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* **2021**, *13*, 4349–4383. https://doi.org/10.5194/essd-13-4349-2021.
34. Didan, K. *MODIS/Terra Vegetation Indices 16-Day L3 Global 250 m SIN Grid V061 [Data Set]*; NASA EOSDIS Land Processes Distributed Active Archive Center: Sioux Falls, SD, USA, 2021. https://doi.org/10.5067/MODIS/MOD13Q1.061.
35. Dorigo, W.; Gruber, A.; De Jeu, R.; Wagner, W.; Stacke, T.; Loew, A.; Albergel, C.; Brocca, L.; Chung, D.; Parinussa, R.; Kidd, R. Evaluation of the ESA CCI soil moisture product using ground-based observations. *Remote Sens. Environ.* **2015**, *162*, 380–395. https://doi.org/10.1016/j.rse.2014.07.023.
36. Zanaga, D.; Van De Kerchove, R.; Daems, D.; De Keersmaecker, W.; Brockmann, C.; Kirches, G.; Wevers, J.; Cartus, O.; Santoro, M.; Fritz, S.; et al. *ESA WorldCover 10 m 2021 v200 [Data Set]*; Zenodo: Geneva, Switzerland, 2022. https://doi.org/10.5281/zenodo.7254221.
37. Clifford, P.; Richardson, S.; Hémon, D. Assessing the significance of the correlation between two spatial processes. *Biometrics* **1989**, *45*, 123–134. https://doi.org/10.2307/2532059.
38. Camps, A.; Park, H.; Pablos, M.; Foti, G.; Gommenginger, C.P.; Liu, P.W.; Judge, J. Sensitivity of GNSS-R spaceborne observations to soil moisture and vegetation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 4730–4742. https://doi.org/10.1109/JSTARS.2016.2588467.
39. Panciera, R.; Walker, J.P.; Kalma, J.D.; Kim, E.J.; Hacker, J.M.; Merlin, O.; Berger, M.; Skou, N. The NAFE'05/CoSMOS data set: Toward SMOS soil moisture retrieval, downscaling, and assimilation. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 736–745. https://doi.org/10.1109/TGRS.2007.915403.

 doi:10.20944/preprints202605.0778.v1

40. Zwieback, S.; Scipal, K.; Dorigo, W.; Wagner, W. Structural and statistical properties of the collocation technique for error characterisation. *Nonlinear Process. Geophys.* **2012**, *19*, 69–80. https://doi.org/10.5194/npg-19-69-2012.
41. Pierdicca, N.; Fascetti, F.; Pulvirenti, L.; Crapolicchio, R.; Muñoz-Sabater, J. Quadruple collocation analysis for soil moisture product assessment. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1595–1599. https://doi.org/10.1109/LGRS.2015.2414654.
42. Odunze, A.C.; Kudi, M.T.; Daudu, C.; Adeosun, J.; Ayoola, G.; Amapu, I.Y.; Abu, S.T.; Mando, A.; Ezui, G.; Constance, D. Soil moisture stress mitigation for sustainable upland rice production in the Northern Guinea Savanna of Nigeria. *Adv. J. Microbiol. Res.* **2015**, *9*, 1–12.