Journal of Spatial Information Sciences

... JSIS

A RELIABILITY FRAMEWORK FOR SENTINEL-2 WATER-QUALITY RETRIEVAL IN OPTICALLY COMPLEX, DATA-SCARCE TROPICAL INLAND WATER: EVIDENCE FROM OGUTA LAKE (NIGERIA)

C. I. C. ACHIONYE, J. I. IGBOKWE, M. O. IGBAFEN

JOURNAL OF SPATIAL INFORMATION SCIENCES ISSN: 2354-3361
VOL. 3, ISSUE 1, PP 225–242, 2026 DOI: https://doi.org/10.5281/zenodo.20709231
PUBLISHED 15-05-2026

www.journals.unizik.edu.ng/jsis

A RELIABILITY FRAMEWORK FOR SENTINEL-2 WATER-QUALITY RETRIEVAL IN OPTICALLY COMPLEX, DATA-SCARCE TROPICAL INLAND WATER: EVIDENCE FROM OGUTA LAKE (NIGERIA)

[1]C. I. C. ACHIONYE, [2]J. I. IGBOKWE, [3]M. O. IGBAFEN

[1,2,3]Department of Surveying and Geoinformatics, Faculty of Environmental Sciences, Nnamdi Azikiwe University, Awka, Nigeria

[1]Corresponding Author: iykeachionye@gmail.com

DOI: https://doi.org/10.5281/zenodo.20709231

Abstract

A reliability-oriented Sentinel-2 workflow was implemented for Oguta Lake, Nigeria, to retrieve Coloured Dissolved Organic Matter (CDOM), turbidity, chlorophyll a (Chl-a), Secchi disk depth (SDD), and total suspended solids (TSS). Sentinel-2 MSI Level-2A imagery spanning 2020 to 2025 was processed in Google Earth Engine using cloud and shadow masking, NDWI water masking, and radiometric normalisation. Field and laboratory measurements were collected at six sites on 6 February 2025; one site (GT6) was excluded after NDWI water-domain screening, so model calibration and LOOCV used N = 5 match-ups. Given the extreme calibration scarcity, high capacity regressors were avoided and final models were restricted to low capacity, positivity by design log linear forms, with fixed pipeline leave one out cross validation used for internal stability assessment (MAE and MdAPE). Ground-truth representativeness across optical regimes was evaluated by sampling a K-Means optical stratification (k = 8) at the sampling sites, and mapping-domain coherence was examined using Pearson and Spearman plausibility diagnostics across a stratified auxiliary time series. Results were parameter dependent. CDOM and turbidity supported cautious semi-quantitative/relative interpretation, while SDD was best treated as relative-only; TSS and Chl-a remained exploratory under N = 5.

Keywords: Cloud and Shadow Masking; Google Earth Engine; Sentinel-2; Case-2 Waters; Atmospheric Correction

JOURNAL OF SPATIAL INFORMATION SCIENCES ISSN: 2354-3361
VOL. 3, ISSUE 1, PP 225–242, 2026 DOI: https://doi.org/10.5281/zenodo.20709231
PUBLISHED 15-05-2026
www.journals.unizik.edu.ng/jsis

1.0 INTRODUCTION

Sustained inland water-quality monitoring is difficult to maintain where field programmes are under-resourced, logistics are challenging, and laboratory capacity is limited. In many sub-Saharan African settings, routine monitoring is reported as costly and operationally difficult to sustain, so measurements are often collected intermittently rather than as continuous, long-term observations [1][2]. Documented cost drivers include staff time, transport, and consumables, such that even basic monitoring designs can become prohibitive when scaled spatially or sustained over long periods [3]. These constraints produce a familiar scientific dilemma: the water bodies of highest management interest are often those with the least consistent in situ data support, limiting evidence-based decision-making and the ability to diagnose change.

Satellite remote sensing is increasingly used as a complementary modality in such contexts because it can provide repeated, synoptic coverage that is not feasible through field sampling alone. Remote sensing can support screening, spatiotemporal mapping, hotspot identification, and relative change detection—functions that are valuable even when absolute concentration retrieval is uncertain. However, general reviews of inland water-quality remote sensing also emphasise that algorithm transferability remains a central limitation: performance can vary strongly by parameter, by lake, and by season, particularly in optically complex waters and when in situ calibration is sparse [4][5][6]. The implication for data-scarce environments is that the key scientific question is often not merely "Can we fit a model?", but "What is the most defensible level of interpretation the model can support, given known constraints?"

1.1 Case-2 optical complexity

Many inland waters are optically complex (often described as Case-2), meaning that surface reflectance is jointly influenced by phytoplankton pigments, Coloured Dissolved Organic Matter (CDOM), and suspended particulate matter. This conceptual framing highlights that spectral signals rarely respond to a single constituent in isolation; rather, mixtures and covariation can confound simple indices and reduce transferability across sites and seasons [7][5]. In practice, this means that retrieval performance is often parameter-dependent and regime-dependent. Multispectral retrievals of turbidity and suspended matter are frequently reported as more robust than chlorophyll-a retrievals in Case-2 conditions, because pigment absorption features can be masked by sediment backscatter and CDOM absorption, especially when atmospheric correction uncertainty and adjacency effects are present [4][6]. Consequently, inland-water remote sensing workflows must be designed to manage

confounding, minimise avoidable radiometric artefacts, and communicate uncertainty in a way that matches the strength of evidence.

1.2 Atmospheric Correction and Adjacency Effects

Atmospheric correction is a dominant uncertainty source in aquatic remote sensing because the water-leaving signal is typically a small fraction of the radiance measured by the sensor. Teaching materials commonly emphasise that, over water, atmospheric path radiance can account for the large majority of the top-of-atmosphere signal, so modest residual errors in aerosol correction may propagate into substantial uncertainty in water-quality retrievals. Over inland waters, these challenges can be compounded by adjacency effects from surrounding land, spatially heterogeneous aerosols, and thin haze, which can bias surface reflectance over dark targets and degrade algorithm transferability [8]. These constraints are particularly important when calibration data are sparse because they limit the ability to separate model error from radiometric error empirically.

In this study, Sentinel-2 MSI Level-2A surface reflectance was used as the primary radiometric input because it provides a consistent multi-year archive that is directly accessible in Google Earth Engine (GEE), supporting reproducible time-series processing at scale. Level-2A reflectance is produced operationally through Sen2Cor processing, whose assumptions and limitations are documented in the Sentinel-2 Level-2A ATBD [9]. Although aquatic-focused atmospheric correction processors (e.g., ACOLITE) have been developed to improve correction over inland and coastal waters through approaches such as dark spectrum fitting [10], a Sen2Cor-based baseline was retained here to maintain internal consistency across the 2020-2025 archive and to avoid mixing processor-dependent reflectance baselines in an extreme sparse-calibration setting.

These atmospheric-correction limitations were not treated as incidental; instead, they were incorporated into the workflow as explicit reliability constraints. The pre-processing chain includes aggressive cloud and shadow masking (Scene Classification Layer combined with Sentinel-2 cloud probability masking), NDWI water masking to enforce a water-only domain, and radiometric normalisation (C-correction) to reduce scene-to-scene illumination variability prior to predictor computation. In addition, reflectance integrity is evaluated through a pixel-wise negative reflectance prevalence diagnostic over the NDWI-water mask for bands B2-B5, used as a transparent indicator of non-physical artefacts in the mapping domain rather than assuming radiometric correctness by default. Collectively, these safeguards aim to reduce cloud/haze contamination and make residual atmospheric and adjacency uncertainty visible in reported diagnostics, consistent with inland-water

syntheses that frame retrieval performance as jointly constrained by optical complexity, atmospheric correction, and transferability [5][6].

1.3 The Sparse-Calibration Problem: Internal Stability Versus Transfer Risk

A recurring weakness in water-quality remote sensing studies is the tendency to report predictive performance without clearly separating internal stability (how a calibrated relationship behaves under resampling of available match-ups) from transfer risk (how predictions may degrade across time, water types, and atmospheric conditions not represented in the calibration set). Under severe calibration scarcity, these can diverge sharply: a model may appear stable within a small set of match-ups but remain unreliable when transferred across seasons or optical regimes. This risk is amplified when model selection (predictor choice, transformations, hyperparameters) is conducted on the same limited dataset used for evaluation, producing optimistically biased performance estimates and unstable inference [11]. Therefore, this study adopts a reliability-oriented reporting stance in which internal error is quantified conservatively while external transfer limitations are communicated explicitly through match-up timing, optical representativeness, and mapping-domain diagnostics rather than through overstated claims of "independent validation" [12]. Accordingly, internal stability is evaluated using leave-one-out cross-validation (LOOCV) under extreme scarcity, and model capacity is restricted to reduce overfitting risk. Transfer risk is treated as a central interpretive axis: it is evidenced through (i) satellite-field temporal offsets, (ii) representativeness of ground-truth points across optical regimes, and (iii) mapping-domain coherence checks. Under such constraints, reliability assessment is best treated as a structured evidence package—combining error-based summaries, conservative model specification, representativeness diagnostics, and plausibility checks rather than as a single "accuracy" statistic [5][6].

The study addresses three research questions: 1. Which Sentinel-2 predictors derived from B2-B5 yield the most stable internal behaviour under LOOCV when model capacity is restricted and non-negativity is enforced by design? 2. How does operational interpretability differ among parameters when internal stability, representativeness constraints, and plausibility diagnostics are jointly considered? 3. How should transfer risk be communicated for a multi-year mapping domain where calibration support is sparse?

2.0 STUDY AREA

Oguta Lake is a major natural freshwater lake in Imo State, southeastern Nigeria, supporting fisheries, domestic use, irrigation and tourism, and exhibiting documented physicochemical variability linked

to human activity and strong wet-dry seasonality [13][14][15][16]. The lake is a Ramsar-designated wetland (Site 1757; designated 30 April 2008) with Ramsar coordinates 05°42′13″N, 06°48′03″E and a reported site area of 572 ha [17]. The ILEC World Lake Database places Oguta Lake within ~5°41′-5°44′N and 6°41′-6°50′E at low elevation (<50m a.s.l.) and characterises it as shallow (maximum depth ~8.0 m; mean depth ~5.5 m) with shoreline length ~10 km and seasonal surface area variability (dry: ~1.8 km²; wet: ~2.5 km²; Ramsar water area ~180-300 ha) [18][19][17]. Hydrologically, Ramsar reports perennial inflows from the Njaba, Utu and Awbuna rivers and outflow to the Orashi River; these connections, together with seasonal forcing, can alter suspended sediment inputs, water colour and clarity on short timescales, contributing to tropical Case-2 optical complexity [17][16][5]. To contextualise the magnitude of satellite-derived products (as scale-of-variation evidence, not validation), published ranges indicate turbidity of 0.40-20.60 NTU (dry) and 0.80-23.30 NTU (rainy) and TSS of 2.20-19.60 mg L^{-1} (dry) and 2.70-20.60 mg L^{-1} (rainy) [16]; transparency/SDD of 0.6-4.0 m (May 1982-April 1983) and 0.61-4.50 m (compilation) [18] [19]; and chlorophyll-a of 0.18-0.51 µg L^{-1} (surface waters, early 1983) [18]. Direct published CDOM/aCDOM measurements specific to Oguta Lake were not identified in accessible sources; a reported “water colour” of 90.67 Pt-Co is therefore provided only as qualitative context [20].

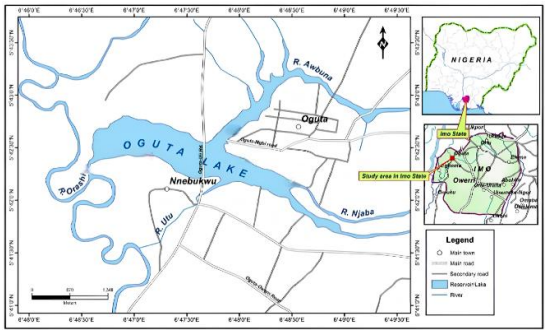

Figure 1a: Study area map showing lake extent. Adapted from [21]; original cartography credited therein

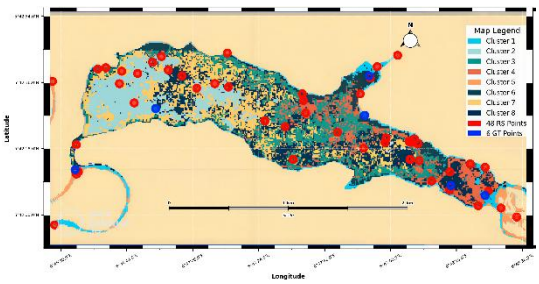

Figure 1b: Map of the Area of Interest with the Ground-truth sampling locations (GT1-GT6), and the 48 auxiliary spectrally stratified sampling points.

JOURNAL OF SPATIAL INFORMATION SCIENCES ISSN: 2354-3361
VOL. 3, ISSUE 1, PP 225–242, 2026 DOI: https://doi.org/10.5281/zenodo.20709231
PUBLISHED 15-05-2026
www.journals.unizik.edu.ng/jsis

The analysis domain was restricted to water pixels within a rectangular AOI (WGS84/EPSG:4326) bounded by 6.7737276-6.8107663°E and 5.6974802-5.7156194°N (≈4.1 km × 2.0 km; ~8.2 km²), with centroid at ~ (6.79225°E, 5.70655°N) (Fig. 1b).

3.0 MATERIAL AND METHODS

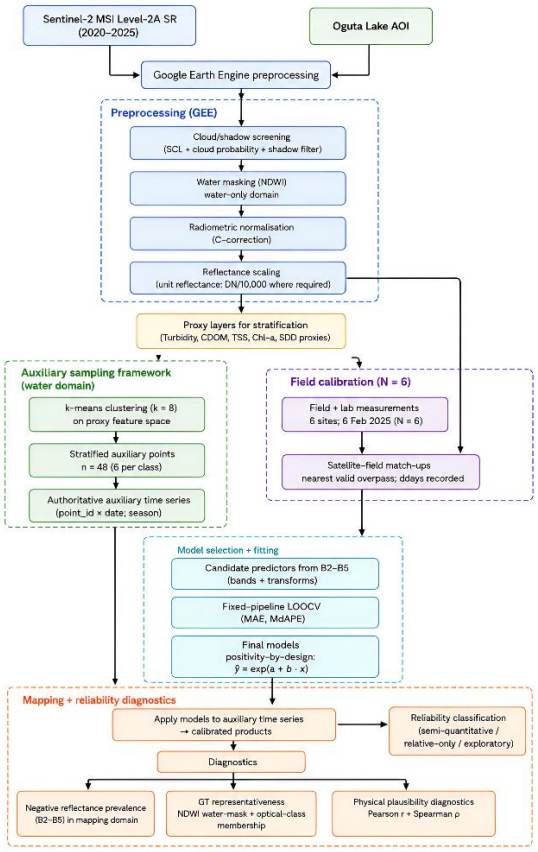

Figure 2: Reliability-oriented Sentinel-2 workflow for water-quality retrieval of the Study

3.1 Satellite data and spectral inputs

Sentinel-2 MSI Level-2A surface reflectance imagery was accessed in Google Earth Engine (GEE) from the Harmonised Sentinel-2 SR collection (COPERNICUS/S2_SR_HARMONIZED) for 2020-2025 [22]. Predictors were restricted to B2 (blue), B3 (green), B4 (red), and B5 (red-edge) to maintain consistent spatial support and avoid resampling artefacts. Where reflectance was stored as scaled integers, values were converted to unit reflectance using $\rho = DN / 10000$. Basic integrity screening masked clearly non-physical reflectance values prior to predictor computation (e.g., $\rho < 0$ or $\rho > 1$ under unit scaling). Negative-reflectance prevalence was then quantified separately as a

mapping-domain diagnostic over NDWI-water pixels (Section 3.8), rather than addressed through post-hoc clipping.

3.2 Field and laboratory data (N = 6) and match-ups

Field sampling and laboratory measurements were collected at six sites (N = 6) on 6 February 2025. One site (GT6) fell outside the NDWI-defined water domain used for water-only stratification and was therefore excluded from calibration. Modelling and LOOCV were performed on N = 5 match-ups. Laboratory Parameter ranges were: 1.29 - 2.42 NTU for Turbidity, 0.786 - 1.28 mg L^{-1} for TSS, 1.5-3.0m for SDD, Chl-a 0.00221 - 0.0506794 mg ml^{-1}. CDOM was used as the laboratory-reported concentration (g m^{-3}) with range 1.306 - 1.803, so absorbance measurements were not used as the calibration target.

The temporal offset between field sampling and satellite acquisition (ddays) was recorded and used as a transfer-risk indicator.

3.3 Pre-Processing in Google Earth Engine

Cloud and shadow screening combined the Sentinel-2 Scene Classification Layer (SCL) with the Sentinel-2 cloud probability product (COPERNICUS/S2_CLOUD_PROBABILITY), following established s2cloudless-style GEE masking guidance [23][24]; this was necessary because residual haze/thin cloud can produce large artefacts over dark water targets [8]. Water pixels were delineated using the Normalised Difference Water Index (NDWI) [25][26], computed as NDWI = ((ρB3 - ρB8) / (ρB3 + ρB8)), and a threshold mask (NDWI > -0.01) was applied to enforce a water-only domain for retrieval and clustering. Radiometric normalisation used a C-correction approach to reduce inter-scene reflectance variability associated with illumination (solar/terrain geometry) differences [27]; although local relief is low, this step was retained as a precaution in the multi-year workflow.

3.4 Optical Stratification and Auxiliary Sampling (K = 8; N = 48)

To characterise optical variability independently of the sparse ground-truth set, an unsupervised stratification was performed over the NDWI-defined water domain. A proxy feature stack was built from the corrected Sentinel-2 composite using five proxy layers (turbidity, CDOM, TSS, Chl-a, and SDD proxies) and clustered with a Weka K-Means clusterer (k = 8). A stratified auxiliary sample of 48 points (six per class) was drawn from the clustered domain and used only for mapping-domain characterisation, plausibility diagnostics, and time-series extraction; these points were excluded from calibration to reduce circular evaluation. A dense auxiliary time series (2020-2025) was then extracted

at auxiliary points for corrected B2-B5 reflectance and proxy layers, with date metadata (year, month, optical class, and a normalised wet/dry seasonal label) retained for interpretation.

3.5 Ground-Truth Representativeness Diagnostics

As clustering was performed only over NDWI-water pixels, NDWI water-mask membership was sampled at each ground-truth site. Optical-class membership was obtained by sampling the final cluster raster at each site in GEE; to ensure masked points were retained in exports, unmask(-1) was applied such that opt_class = -1 indicated "outside classified water domain/no data." Results were reported in Table 2.

3.6 Predictor Engineering and Restricted Model Specification

Candidate predictors were generated from B2-B5 reflectance using low-order algebraic transforms (ratios, differences, inverse terms, normalised differences, and logarithmic transforms). To enforce non-negativity without post-hoc clipping and to reduce overfitting risk under very small N, final models were restricted to a positivity-by-design log-linear form:

y_{pred} = exp(a+bx). High-capacity nonlinear models were avoided because model selection on extremely small samples can produce unstable inference and optimistically biased performance estimates [28][11]. Predictor-model pipelines were ranked using a fixed-pipeline LOOCV procedure, prioritising MAE and supported by MdAPE.

3.7 Internal Stability Assessment

Internal stability was assessed with leave-one-out cross-validation (LOOCV) across the five match-ups: in each fold, one site was withheld, the model was fitted on the remaining sites, and the withheld site error was computed. Under N = 5, LOOCV was used as an internal stability indicator rather than a claim of external validity. Performance was reported using error-based metrics that remain interpretable under small samples: MAE (in parameter units) and MdAPE (median absolute percentage error). Correlation-style metrics (e.g., R^2) were not treated as headline evidence due to instability under very small N and sensitivity to leverage points. LOOCV results were summarised in Table 4 with an operational reliability interpretation.

3.8 Negative Reflectance Prevalence (pixel-wise; B2-B5)

A mapping-scale integrity diagnostic quantified negative surface reflectance prevalence over the NDWI-water domain. For each Sentinel-2 SR image passing cloud/shadow masking, the percentage of valid water pixels with reflectance < 0 was computed for B2-B5 as 100×(negCount/validCount),

where validCount is the number of unmasked water pixels. Images with validCount = 0 were flagged as not computable and excluded from summaries. No post-hoc clipping of reflectance values was applied in this diagnostic.

3.9 Physical plausibility diagnostics

Coherence among retrieved clarity/particulate indicators were computed using products generated from the final N=5 model coefficients applied across the auxiliary time series (n = 3,832) using Pearson r and Spearman ρ. Three relationships were assessed: SDD-turbidity, TSS-turbidity, and TSS-SDD. These correlations were interpreted as plausibility/coherence checks rather than accuracy measures, with relationships involving exploratory products treated conservatively; results were visualised in Figure 3.

3.10 Reliability Classification Scheme

Operational reliability was classified into three categories: (1) semi-quantitative (cautious), (2) relative-only, and (3) exploratory. Classification was assigned parameter-by-parameter using evidence from the final model forms (Table 3), LOOCV error behaviour (Table 4), ground-truth representativeness diagnostics (Table 2), negative-reflectance diagnostics, and plausibility coherence (Figure 3).

4.0 RESULTS AND DISCUSSION

4.1 Archive support and mapping-domain evidence (2020-2025)

A dense Sentinel-2 record supported multi-year mapping and diagnostics across 2020–2025. To characterise mapping-domain behaviour independently of calibration, an auxiliary framework was retained comprising 3,832 space–time observations extracted at stratified auxiliary points with month/season metadata. This separation reduces circular evaluation risk and supports domain-level plausibility assessment under known dependence pitfalls in remote-sensing accuracy reporting. As a pixel-wise reflectance-integrity check over the NDWI-water domain, negative reflectance prevalence was evaluated for B2–B5 across the retained archive (101 observations). Negative reflectance was not observed over valid water pixels (archive-wide prevalence 0.00% per band); two images yielded no valid water pixels after masking and were treated as not computable rather than as evidence of radiometric artefact. This supports numerical stability of predictor computation, while not precluding positive-valued adjacency/aerosol residual biases.

4.2 Field calibration match-ups and constraints (six sites sampled; N = 5 modelled)

Field and laboratory observations were acquired at six sites on 6 February 2025 and matched to the nearest valid Sentinel-2 acquisition meeting masking rules. Match-up timing offsets were heterogeneous (e.g., −1 day, +4 days, +9 days), implying non-uniform transfer risk because optical mixtures and atmospheric residuals can shift over short intervals in tropical inland waters [5][6]. One site (GT6) was excluded prior to modelling because it was not classified as water by the NDWI mask used to define the water-only domain for clustering and retrieval support, yielding N = 5 match-ups for calibration and LOOCV. Given this scarcity and timing structure, independent external validation is not claimed; LOOCV is treated as an internal stability indicator and transfer limitations are communicated through representativeness and mapping-domain coherence diagnostics [28][11].

4.3 Optical stratification and representativeness of ground truth

Table 2: Optical class membership of each groundtruth point

SAMPLE_ID	IS_WATER	OPT_CLASS	LONGITUDE	LATITUDE
GT1	1	7	6.776152	5.703515
GT2	1	1	6.782249	5.708135
GT3	1	7	6.798308	5.710561
GT4	1	4	6.807126	5.701492
GT5	1	2	6.80456	5.70228
GT6	0	-1	6.798028	5.707595

Unsupervised stratification was implemented using K-Means (k = 8) over the NDWI-defined water domain. A balanced auxiliary sample (48 points) was drawn across classes for time-series extraction and plausibility diagnostics and was excluded from calibration to preserve a non-circular evidence stream. Ground-truth representativeness was assessed by sampling NDWI water membership and the final cluster raster at each ground-truth location (Table 2). Five sites were classified as water (is_water = 1) and assigned optical classes; four of the eight classes were represented (1, 2, 4, and 7), with two sites sharing class 7. This indicates that calibration spans multiple optical regimes but remains sparse at the class level, so transfer to unrepresented classes is weakly supported under N = 5 and mapped products require conservative interpretation outside represented regimes [7][5][6].

4.4 Final Retrieval Suite and Predictor Structure

Table 3: Final retrieval Models fit on N=5 match-ups

Parameter	Predictor x	Model	Coefficients (a, b)
CDOM	x=1/ρB3	log-y linear	a = 1.032144; b = -0.119324

Turbidity	x=1/ρB3	log-y linear	a = -1.261022; b = 0.342351
TSS	x=ρB4	log-y linear	a = -0.724147; b = 3.944924
SDD	x=ρB5/ρB3	log-y linear	a = 3.737193; b = -3.152079
Chl-a	x=ρB2-ρB4	log-y linear	a = -4.379754; b = 22.884158

Final retrievals were fitted on the N = 5 match-ups using a restricted positivity-by-design log-linear form, $y_{pred} = \exp(a+bx)$, to limit model capacity and enforce non-negativity without post-hoc clipping (Table 3). Predictors were parameter-specific: CDOM and turbidity both used an inverse green term (1/ρB3), SDD used (ρB5/ρB3), TSS used ρB4 and chlorophyll-a used (ρB2-ρB4) (Table 3).

Since multiple products share spectral dependence, especially through B3, inter-product correlations may be mechanically induced; mapping-domain correlations among products are interpreted as plausibility signals rather than as evidence of independent limnological coupling.

4.5 Internal Stability Assessment

Table 4: LOOCV performance (N=5) and reliability

Param.	Selected predictor	MAE	MdAPE (%)	Reliability interpretation
CDOM	1/ρB3	0.114667	2.8656	Relative / semi-quantitative (cautious)
Turbidity	1/ρB3	0.190223	12.5464	Relative / semi-quantitative (cautious)
SDD	ρB5/ρB3	0.425835	15.9790	Relative only (cautious)
TSS	ρB4	0.230336	23.0264	Exploratory / Relative only (cautious)
Chl-a	ρB2-ρB4	0.013025	75.6010	Exploratory / Qualitative only

Internal stability was assessed using LOOCV across the five match-ups retained for modelling and summarised using error-based metrics (MAE; MdAPE), which are more interpretable than correlation-style metrics under very small N and leverage sensitivity [28][11]. Results and operational reliability tiers are summarised in Table 4. Overall, CDOM and turbidity exhibited the most defensible internal error behaviour under N = 5, while SDD and TSS were weaker and chlorophyll-a remained exploratory (Table 4). Interpretations below are conditional on representativeness limitations (Table 2) and residual atmospheric/adjacency uncertainty. CDOM showed the lowest LOOCV error (MAE = 0.114667; MdAPE = 2.8656%; Table 4), supporting relative to cautious semi-quantitative interpretation within represented optical regimes. Turbidity exhibited comparatively stable LOOCV behaviour (MAE = 0.190223; MdAPE = 12.5464%; Table 4) and was classified as relative to cautious semi-quantitative, consistent with broader evidence that turbidity proxies can be robust in some settings while remaining sensitive to atmospheric correction residuals and regime differences in inland Case-2 waters [29][8]. SDD showed moderate internal support

(MAE = 0.425835; MdAPE = 15.9790%; Table 4) and was interpreted conservatively as relative-only, noting that SDD integrates absorption and scattering and can shift with seasonal mixture changes, limiting transfer beyond the calibration regime [5][6]. TSS exhibited weaker internal support (MAE = 0.230336; MdAPE = 23.0264%; Table 4) and was treated as exploratory/relative-only because transfer can be strongly modulated by particle composition and atmospheric correction assumptions under sparse calibration [29][6]. Chlorophyll-a was the least stable in proportional terms (MdAPE = 75.6%; Table 4) and was retained as exploratory/qualitative-only, consistent with known challenges of Chl-a retrieval in optically complex inland waters [30][6][31].

4.6 Mapping Domain Diagnostics and Plausibility Evidence

Reflectance integrity diagnostics indicated no negative surface reflectance values over valid NDWI-water pixels in bands B2-B5 across the retained 2020-2025 archive (101 observations; Σvalid water pixels = 1,619,317 per band; Σnegative pixels = 0; two dates yielded no valid water pixels after masking and were treated as not computable). This supports numerical stability of predictor computation, but it does not preclude positive-valued adjacency or aerosol-correction residual biases that can still distort aquatic retrievals over dark targets [9][8]. Across the calibrated auxiliary 2020-2025 time series (n = 3,832), retrieved SDD and turbidity did not exhibit a robust inverse association at mapping-domain scale (Pearson r=-0.0234; Spearman ρ=0.2359; Fig. 3). This implies that higher turbidity retrievals were not consistently accompanied by lower clarity retrievals across time and optical regimes, and that the expected turbidity-clarity relationship cannot be assumed to transfer uniformly at domain scale. Plausible contributors include shared dependence on ρB3 (turbidity uses 1/ρB3; SDD uses ρB5/ρB3; Table 3), differential band-wise bias from adjacency/aerosol residuals, and regime heterogeneity that is only weakly constrained by the N = 5 calibration set. Relationships involving TSS were directionally plausible but interpreted conservatively because TSS remained weakly supported under LOOCV (Table 4). TSS-turbidity showed a strong inverse monotonic association (Spearman ρ = -0.8994) with weak linear association (Pearson r = -0.0618; Fig. 3). TSS-SDD was inversely associated (Pearson r = -0.3284; Spearman ρ = - 0.5623; Fig. 3). Within the reliability framework, these results are treated as coherence signals rather than evidence of quantitative accuracy. Overall, Fig. 3 demonstrates that mapping-domain coherence may diverge from small-sample internal stability, reinforcing the need to integrate LOOCV behaviour, representativeness, and plausibility diagnostics rather than substituting one for another.

JOURNAL OF SPATIAL INFORMATION SCIENCES ISSN: 2354-3361
VOL. 3, ISSUE 1, PP 225–242, 2026 DOI: https://doi.org/10.5281/zenodo.20709231
PUBLISHED 15-05-2026
www.journals.unizik.edu.ng/jsis

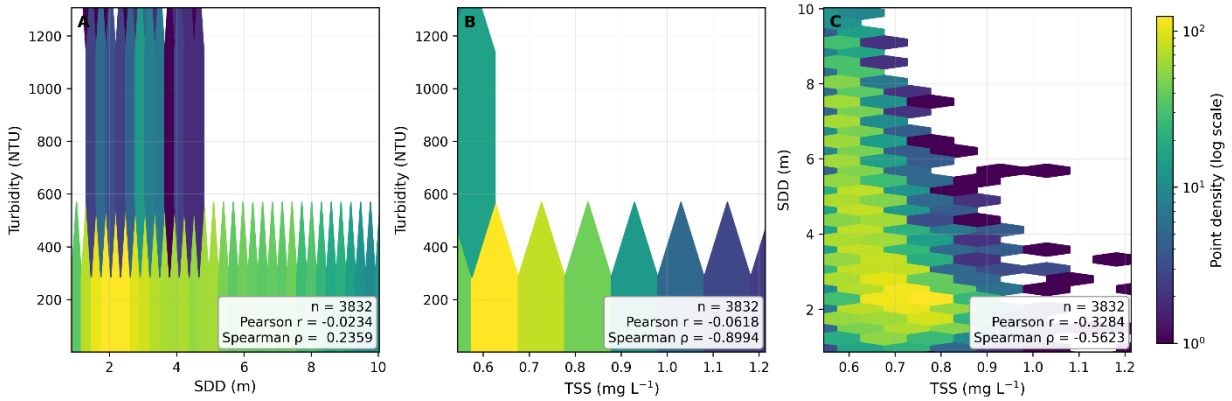

Figure 3: Physical plausibility diagnostics for the calibrated retrieval products

4.7 Reliability interpretation and operational implications

Taken together, model structure (Table 3), LOOCV behaviour (N = 5, Table 4), ground-truth representativeness (Table 2), and mapping-domain coherence (Fig. 3) support tiered operational guidance. CDOM shows the strongest internal stability and supports cautious semi-quantitative/relative interpretation within represented regimes. Turbidity is most defensible for relative screening and change detection and may support cautious magnitude interpretation under explicit constraints. SDD shows moderate internal support but is best treated conservatively as relative-only, particularly because domain-scale SDD-turbidity coherence is weak (Fig. 3). TSS and chlorophyll-a remain weakly supported under the present calibration scarcity and are therefore most appropriate for relative/qualitative mapping and hypothesis generation rather than quantitative concentration reporting. Overall, the divergence between internal stability and mapping-domain plausibility underscores why reliability tiering is necessary in data-scarce environments: internal LOOCV behaviour can appear acceptable for some products while transfer risk remains substantial due to temporal offsets in match-ups, limited optical-class representativeness, and residual atmospheric/adjacency effects.

4.8 Limitations

Several limitations constrain inference. Field/laboratory observations were collected at six sites on a single date (6 February 2025), but one site (GT6) fell outside the NDWI-defined water domain used for water-only stratification and was excluded from calibration; final model fitting and LOOCV therefore used N = 5 match-ups. Match-up timing offsets varied across sites, implying non-uniform transfer risk. Atmospheric correction residuals and adjacency effects were not decomposed into separate propagated uncertainty components, and Level-2A reflectance inputs inherit documented

algorithm assumptions. Model capacity was intentionally restricted (positivity-by-design log-linear form) to reduce overfitting risk under very small N; while this improves statistical defensibility, it does not represent a complete physical model of constituent interactions in Case-2 waters.

5.0 CONCLUSION

A reliability-oriented Sentinel-2 workflow was developed for inland water-quality retrieval in Oguta Lake, Nigeria, explicitly under extreme calibration scarcity. Sentinel-2 MSI Level-2A surface reflectance imagery (2020-2025) was processed in Google Earth Engine using cloud/shadow screening, NDWI water masking, and radiometric normalisation. Field/laboratory measurements were collected at six sites on 6 February 2025; one site (GT6) was excluded after NDWI water-domain screening, and final calibration and LOOCV were conducted on N = 5 match-ups. To limit overfitting and enforce non-negativity without post-hoc clipping, retrievals were restricted to low-capacity positivity-by-design log-linear models (Table 3) and internal stability was quantified using LOOCV with error-based metrics (Table 4). Findings were strongly parameter-dependent. CDOM exhibited the most defensible internal behaviour and supports cautious semi-quantitative (and relative) interpretation within represented optical regimes (Tables 2 and 4). Turbidity is most defensible for relative screening and change detection within a consistent processing workflow (Table 4). SDD showed moderate internal support but should be interpreted conservatively as relative-only, particularly because mapping-domain plausibility checks did not robustly reproduce the expected inverse turbidity–clarity coherence (Fig. 3). TSS and chlorophyll-a remain weakly supported under N = 5 and are best treated as exploratory products for qualitative mapping and hypothesis generation rather than concentration reporting (Table 4). Overall, the results support parameter-specific reliability tiering (semi-quantitative, relative-only, exploratory) based on internal error behaviour, optical representativeness, and mapping-domain plausibility diagnostics rather than implying uniform quantitative accuracy across all products. This work addresses a persistent challenge in aquatic remote sensing: in many tropical inland waters, in situ calibration is too sparse to support strong claims of validated concentrations, yet monitoring needs remain urgent. The paper contributes a reliability-oriented reporting framework that prioritises defensible interpretation tiers and transparent transfer-risk communication. The combination of conservative model capacity, LOOCV-based internal stability assessment, representativeness diagnostics, and mapping-domain plausibility checks provides a practical blueprint for responsible satellite product communication in data-scarce settings.

References

[1] Dube, T., Mutanga, O., Seutloali, K., Adelabu, S., & Shoko, C. (2015). Water quality monitoring in sub-Saharan African lakes: A review of remote sensing applications. *African Journal of Aquatic Science*, *40*(1), 1–7. https://doi.org/10.2989/16085914.2015.1014994

[2] Peletz, R., Kisiangani, J., Bonham, M., Ronoh, P., Delaire, C., Kumpel, E., Marks, S., & Khush, R. (2018). Why do water quality monitoring programs succeed or fail? A qualitative comparative analysis of regulated testing systems in sub-Saharan Africa. *International Journal of Hygiene and Environmental Health*, *221*(6), 907–920. https://doi.org/10.1016/j.ijheh.2018.05.010

[3] Peletz, R., Kumpel, E., Bonham, M., Rahman, Z., & Khush, R. (2016). How much will it cost to monitor microbial drinking water quality in sub-Saharan Africa? *Environmental Science & Technology*, *50*(20), 10869–10876. https://doi.org/10.1021/acs.est.6b06442

[4] Gholizadeh, M. H., Melesse, A. M., & Reddi, L. (2016). A comprehensive review on water quality parameters estimation using remote sensing techniques. *Sensors*, *16*(8), 1298. https://doi.org/10.3390/s16081298

[5] Palmer, S. C. J., Kutser, T., & Hunter, P. D. (2015). Remote sensing of inland waters: Challenges, progress and future directions. *Remote Sensing of Environment*, *157*, 1–8. https://doi.org/10.1016/j.rse.2014.09.021

[6] Sagan, V., Peterson, K. T., Maimaitijiang, M., Sidike, P., Sloan, J., Greeling, B. A., Maalouf, S., & Adams, C. (2020). Monitoring inland water quality using remote sensing: Potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. *Earth-Science Reviews*, *205*, 103187. https://doi.org/10.1016/j.earscirev.2020.103187

[7] Morel, A., & Prieur, L. (1977). Analysis of variations in ocean color. *Limnology and Oceanography*, *22*(4), 709–722. https://doi.org/10.4319/lo.1977.22.4.0709

[8] Pan, Y., Bélanger, S., & Huot, Y. (2022). Evaluation of atmospheric correction algorithms over lakes for high-resolution multispectral imagery: Implications of adjacency effect. *Remote Sensing*, *14*(13), 2979. https://doi.org/10.3390/rs14132979

[9] European Space Agency. (n.d.). *Sentinel-2 Level-2A algorithm theoretical basis document (ATBD) (Sen2Cor)* [Technical documentation]. Retrieved June 5, 2026, from https://sentinels.copernicus.eu/documents/247904/446933/Sentinel-2-Level-2A-Algorithm-Theoretical-Basis-Document-ATBD.pdf

[10] Vanhellemont, Q. (2019). Adaptation of the dark spectrum fitting atmospheric correction for aquatic applications of the Landsat and Sentinel-2 archives. *Remote Sensing of Environment*, *225*, 175–192. https://doi.org/10.1016/j.rse.2019.03.010

[11] Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, *7*, Article 91. https://doi.org/10.1186/1471-2105-7-91

[12] Andrade, C. (2018). Internal, external, and ecological validity in research design, conduct, and evaluation. *Indian Journal of Psychological Medicine*, *40*(5), 498–499. https://doi.org/10.4103/IJPSYM.IJPSYM_334_18

[13] Ahiarakwem, C. A. (2007). *Water quality assessment and some geochemical models of Oguta Lake, southeastern Nigeria* (Doctoral dissertation, Federal University of Technology, Owerri).

[14] Andong, F. A., Ezenwaji, N. E., Melefa, T. D., Hinmikaiye, F. F., Nnadi, O. V., & Oluwafemi, O. (2019). Assessment of the physico-chemical properties of Oguta Lake compared to the established values of the Federal Ministry of Environment, Nigeria. *Advances in Oceanography and Limnology*, *10*(2). https://doi.org/10.4081/aiol.2019.8522

[15] Verla, A. W., Enyoh, C. E., Verla, E. N., Okeke, P. N., & Pingale, S. S. (2019). Chemometric assessment of Orashi River after confluence with Oguta Lake. *Indonesian Journal of Fundamental and Applied Chemistry*, *4*(3), 91–102. https://doi.org/10.24845/ijfac.v4.i3.91

[16] Madu, F. M., Okoyeh, E. I., Okolo, C. M., Chibuzor, S. N., Onyebum, T. E., & Okpara, A. (2022). Physicochemical and microbial assessment of Oguta Lake, Southeastern Nigeria. *International Journal of Innovative Science and Research Technology*, *7*(11), 14–25.

[17] Ramsar Sites Information Service. (2008). *Oguta Lake (Ramsar Site No. 1757)*. Retrieved June 5, 2026, from https://rsis.ramsar.org/ris/1757?language=en

[18] International Lake Environment Committee Foundation. (n.d.). *Oguta Lake*. World Lake Database. Retrieved June 5, 2026, from https://wldb.ilec.or.jp/Display/html/3599

[19] Okorie, P. U. (2004). *Socio-economic appraisal of cage fish culture in Oguta Lake, Nigeria*. Retrieved June 5, 2026, from https://dlc.dlib.indiana.edu/dlc/bitstreams/23dc35f7-fc93-4e56-831e-4a4f965386d9/download

[20] Okorondu, S. I., & Anyadoh-Nwadike, S. O. (2015). Bacteriological and physicochemical analysis of Oguta Lake water, Imo State, Nigeria. *Science Journal of Public Health*, *3*(5-1), 14–19. https://doi.org/10.11648/j.sjph.s.2015030501.13

[21] Ogueri, C., Adebayo, E. T., & Nwaiwu, V. N. C. (2018). Assessment of heavy metals in fish species and some physico-chemical parameters of Oguta Lake, Oguta, Imo State, Nigeria. *International Journal of Hydrology*, *2*(2), 121–124. https://doi.org/10.15406/ijh.2018.02.00060

[22] Google Earth Engine. (n.d.). *Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-2A (SR) (COPERNICUS/S2_SR_HARMONIZED)* [Data set]. Earth Engine Data Catalog. Retrieved June 5, 2026, from https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED

[23] Google Earth Engine. (n.d.). *Sentinel-2: Cloud probability (COPERNICUS/S2_CLOUD_PROBABILITY)* [Data set]. Earth Engine Data Catalog. Retrieved June 5, 2026, from https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_CLOUD_PROBABILITY

[24] Google Earth Engine. (n.d.). *Sentinel-2 cloud masking with s2cloudless* [Tutorial]. Retrieved June 5, 2026, from https://developers.google.cn/earth-engine/tutorials/community/sentinel-2-s2cloudless

[25] McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, *17*(7), 1425–1432. https://doi.org/10.1080/01431169608948714

[26] Xu, H. (2006). Modification of Normalised Difference Water Index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, *27*(14), 3025–3033. https://doi.org/10.1080/01431160600589179

[27] Teillet, P. M., Guindon, B., & Goodenough, D. G. (1982). On the slope-aspect correction of multispectral scanner data. *Canadian Journal of Remote Sensing*, *8*(2), 84–106. https://doi.org/10.1080/07038992.1982.10855028

[28] Steyerberg, E. W., Harrell, F. E., Borsboom, G. J. J. M., Eijkemans, M. J. C., Vergouwe, Y., & Habbema, J. D. F. (2001). Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology*, *54*(8), 774–781. https://doi.org/10.1016/S0895-4356(01)00341-9

[29] Dogliotti, A. I., Ruddick, K. G., Nechad, B., Doxaran, D., & Knaeps, E. (2015). A single algorithm to retrieve turbidity from remotely-sensed data in all coastal and estuarine waters. *Remote Sensing of Environment*, *156*, 157–168. https://doi.org/10.1016/j.rse.2014.09.020

[30] Gitelson, A. A., Gurlin, D., Moses, W. J., & Barrow, T. (2009). A bio-optical algorithm for the remote estimation of the chlorophyll-a concentration in case 2 waters. *Environmental Research Letters*, *4*(4), 045003. https://doi.org/10.1088/1748-9326/4/4/045003

[31] Sòria-Perpinyà, X., Vicente, E., Urrego, P., Pereira-Sandoval, M., Tenjo, C., Ruíz-Verdú, A., Delegido, J., Soria, J. M., Peña, R., & Moreno, J. (2021). Validation of water quality monitoring algorithms for Sentinel-2 and Sentinel-3 in Mediterranean inland waters with in-situ reflectance data. *Water*, *13*(5), 686. https://doi.org/10.3390/w13050686