Multi-Sensor Earth Observation and Machine Learning for Monitoring and Forecasting Reservoir Surface Water Dynamics: A Case Study of Gurara Reservoir, Nigeria

Alamin Musa Magaga[1*], Daberechi Okorie[2], Yazid S Mikail[2], Daniel Akanbi[2], Meron Abate[2], Abubakar Isa-Abubakar [2], Nimatallahi Masuud[2]

[1,2]Nest Africa AI Innovation Lab, Kaduna, Nigeria

***Correspondence:**

Alamin Musa Magaga

magaga@nestafrica.ai

Abstract

This study developed and evaluated a reproducible satellite-to-forecast workflow for Gurara Reservoir, Nigeria, using monthly Landsat 5/7/8/9 observations from January 2001 through April 2026, Sentinel-2 cross-sensor comparison, CHIRPS rainfall, TerraClimate variables, statistical trend tests, baseline and machine-learning forecasts, and bias-corrected CMIP6 forcing. Of 304 calendar months, 196 yielded valid Landsat estimates (64.5%); the longest uninterrupted valid sequence was 23 months. MNDWI gave the most defensible cross-sensor result. Over 73 paired months, agreement was moderate ($r = .540$; MAE = 3.463 km^2; RMSE = 6.838 km^2; Landsat-minus-Sentinel-2 bias = −0.967 km^2). Surface area increased across the full record (Mann–Kendall $\tau = .431$, $p = 3.19 \times 10^{-19}$; Sen slope = 1.817 km^2 $year^{-1}$), but the trend spans reservoir establishment and cannot be interpreted as climate-only change. Mean seasonal area ranged from 27.285 km^2 in April to 43.908 km^2 in August. PDSI had the strongest contemporaneous association ($r = -.454$), whereas rainfall correlations were weak. On the held-out 2022–2026 period, SARIMA was best at one and three months (RMSE = 8.014 and 9.648 km^2); persistence methods were best at six and twelve months. The May 2026 forecast was 41.890 km^2 (95% interval: 29.449–54.331), with rapidly widening uncertainty. CMIP6-informed scenarios through 2030 separated only modestly relative to uncertainty. The workflow supports transparent monitoring and conditional planning, but not storage accounting, release optimization, or deterministic long-range prediction without in situ data.

Keywords: Gurara Reservoir, Landsat, Sentinel-2, surface-water area, SARIMA, machine learning, CMIP6, uncertainty

Introduction

Reservoirs are strategic infrastructure for water supply, irrigation, hydropower, drought buffering, and flood regulation, yet their operational value depends on observations that resolve both seasonal dynamics and abnormal departures. In many parts of sub-Saharan Africa, these observations remain fragmented across institutions or unavailable at the temporal resolution needed for anticipatory management. Satellite remote sensing cannot replace reservoir gauges, but it can provide an auditable record of visible water extent and can anchor a monitoring system where conventional data are sparse.

The scientific challenge is not merely to map water. A usable reservoir forecasting system must harmonize sensors, control cloud contamination, preserve missingness honestly, separate hindcast validation from fitting, compare complex algorithms against simple baselines, and communicate uncertainty at each lead time. These requirements are especially important for an impoundment such as Gurara, whose satellite record contains structural change associated with reservoir formation and whose visible surface area is an imperfect proxy for stored volume.

Recent Earth-observation research has made long, repeatable surface-water records possible through global archives, analysis-ready products, and cloud computing (Gorelick et al., 2017; Pekel et al., 2016; Pickens et al., 2020; Abatzoglou et al., 2018; Muñoz-Sabater et al., 2021). At the same time, hydrological forecasting scholarship increasingly emphasizes benchmark-aware evaluation, chronological validation, and calibrated uncertainty rather than model novelty alone (Kratzert et al., 2018; Gauch et al., 2021; Nearing et al., 2021; Shen, 2018; Tyralis et al., 2019; Chen & Guestrin, 2016; Lim et al., 2021; Makridakis et al., 2018; Bergmeir et al., 2018; Knoben et al., 2019). These advances motivate an integrated case study that treats monitoring and forecasting as one evidence chain.

This study addresses four questions:

(a) What are the long-term and seasonal patterns of Gurara Reservoir surface-water area from 2001 to 2026?

(b) Which water index provides the most defensible cross-sensor agreement between Landsat and Sentinel-2?

(c) How do persistence, seasonal persistence, SARIMA, random forest, and extreme gradient boosting compare at 1-, 3-, 6-, and 12-month horizons under chronological validation?

(d) what conditional surface-area trajectories emerge through 2030 when the fitted statistical baseline is perturbed with bias-corrected CMIP6 climate information?

The contribution is methodological and decision-oriented. It delivers a traceable monthly dataset, explicit quality controls, independent sensor comparison, horizon-specific forecast evaluation, and scenario uncertainty. It also states what the evidence does not support: the results are not a bathymetric storage estimate, a causal attribution analysis, or a substitute for operating records.

Literature Review

Satellite Observation of Inland Water

The Landsat archive provides the temporal depth needed to study multi-decadal water dynamics, while Sentinel-2 supplies finer spatial resolution and denser recent observations. Combining these families is attractive but nontrivial because bandpasses, radiometry, geolocation, atmospheric correction, and cloud masks differ. Cross-sensor consistency must therefore be measured empirically rather than assumed.

Recent studies demonstrate that optical and radar constellations can be fused to improve the continuity of surface-water monitoring, particularly where clouds interrupt optical retrievals (Declaro & Kanae, 2024; Wang et al., 2022; Flood, 2017; Perez & Vitale, 2023; Cho & Ryu, 2024).

Comparisons among Landsat and Sentinel products show that harmonization, co-registration, and sensor-specific quality controls materially affect mapped water boundaries (Page et al., 2019; Yang et al., 2020; Stumpf et al., 2018; Jiang et al., 2022; Wu et al., 2023).

Lake and reservoir applications increasingly use long image archives to derive water-area trajectories, but their reported accuracy remains sensitive to thresholds, shore-zone mixing, and image availability (Frantz, 2019; Schwatke et al., 2019; Pan, 2024; Yao et al., 2019; Hafeez et al., 2022).

Cloud platforms have expanded reproducible surface-water mapping from individual scenes to regional and multi-decadal time series, making explicit workflow documentation increasingly important (Li & Roy, 2017; Du et al., 2023; Zhao & Gao, 2019; Zhao & Gao, 2018; Keys & Scott, 2017).

Recent studies demonstrate that optical and radar constellations can be fused to improve the continuity of surface-water monitoring, particularly where clouds interrupt optical retrievals (Kolarik et al., 2023; Tsolakidis & Vafiadis, 2019; Sorkhabi et al., 2022; Yao et al., 2022; Das et al., 2022).

Comparisons among Landsat and Sentinel products show that harmonization, co-registration, and sensor-specific quality controls materially affect mapped water boundaries (Chipman, 2019; Cai et al., 2024; Yue et al., 2023; Zhao & Wang, 2024; Inman & Lyons, 2020).

Lake and reservoir applications increasingly use long image archives to derive water-area trajectories, but their reported accuracy remains sensitive to thresholds, shore-zone mixing, and image availability (Wang et al., 2020; Wang et al., 2018; Lu & Sun, 2023; Owusu et al., 2022; Hu et al., 2022).

Cloud platforms have expanded reproducible surface-water mapping from individual scenes to regional and multi-decadal time series, making explicit workflow documentation increasingly important (Taheri Dehkordi et al., 2022).

The broader evidence base includes global surface-water histories, harmonized Landsat–Sentinel processing, cloud-screening evaluations, analysis-ready Landsat products, satellite altimetry, and forthcoming swath-altimetry capabilities (Gorelick et al., 2017; Pekel et al., 2016; Pickens et al., 2020; Wulder et al., 2022; Foga et al., 2017; Fisher et al., 2016; Hollstein et al., 2016; Vermote et al., 2016; Busker et al., 2019; Biancamaria et al., 2016). Together, this literature supports the use of Landsat as the long record and Sentinel-2 as a comparison source rather than treating either sensor as field truth.

Water Indices, Thresholds, and Validation

Index-based mapping remains attractive because it is interpretable and computationally efficient. NDWI contrasts green and near-infrared reflectance, MNDWI replaces the near-infrared term with shortwave infrared to reduce built-up and soil confusion, and AWEI was designed to improve discrimination in shadowed or dark backgrounds. These methods are foundational, yet none guarantees a transferable threshold: turbidity, aquatic vegetation, shallow margins, seasonal drawdown, and sensor resolution can all alter the spectral mixture.

Contemporary evaluations reinforce the need to select indices against local evidence and to report omission, commission, and boundary uncertainty rather than only map appearance (Fisher et al., 2016; Hollstein et al., 2016; Vermote et al., 2016; McFeeters, 1996; Xu, 2006; Feyisa et al., 2014). For Gurara, monthly agreement with Sentinel-2 was used as a cross-sensor robustness test. That test strengthens internal consistency but is not equivalent to validation against contemporaneous shoreline surveys or gauge-derived area–elevation curves.

Climate Data and Reservoir Response

Reservoir area reflects the integrated effects of inflow, direct rainfall, evaporation, antecedent basin wetness, operating rules, abstraction, seepage, and the reservoir's area–elevation geometry. Contemporaneous bivariate correlations therefore provide descriptive association, not process attribution. Lagged rainfall can be important even when same-month association is weak, and an aridity index can summarize water-balance persistence that a single rainfall total does not capture.

Validation studies show that gridded precipitation products have region-, season-, elevation-, and timescale-dependent skill; local evaluation remains necessary before hydrological interpretation (Prakash, 2019; Nawaz et al., 2021; Rivera et al., 2018; Shen et al., 2020; Katsanos et al., 2016).

CHIRPS has been assessed across diverse hydroclimates, with studies reporting useful spatial coverage alongside biases in event magnitude and extremes (Alsilibe et al., 2023; Fessehaye et al., 2022; Marra et al., 2022; Poméon et al., 2017; Satgé et al., 2020).

West and Central African comparisons emphasize that no precipitation product dominates for every basin or metric, supporting ensemble or sensitivity-based use (Gyamfi et al., 2021; Mendoza Paz & Willems, 2022; Kouakou et al., 2023; Sacré Regis M. et al., 2020).

The climate inputs used here reflect widely adopted gridded data architectures: CHIRPS for rainfall, TerraClimate for monthly water-balance variables, and ERA5-Land as an example of modern land reanalysis (Abatzoglou et al., 2018; Muñoz-Sabater et al., 2021; Funk et al., 2015; Thrasher et al., 2022). Their spatial completeness is valuable, but the reservoir analysis still inherits forcing uncertainty and cannot recover unobserved releases or withdrawals.

Forecasting, Baselines, and Evaluation

Hydrological forecasting now includes statistical, tree-based, recurrent neural, transformer, and hybrid models. Yet data volume and validation design often matter more than architectural sophistication. A monthly series with missing observations and a short uninterrupted training sequence may not contain enough independent events to estimate high-capacity models reliably. Benchmarking against persistence and seasonal persistence is therefore indispensable.

Recent applications compare statistical and machine-learning approaches for water-level, inflow, drought, and flood forecasting, while repeatedly identifying lead-time-dependent degradation (Monteiro & Costa, 2018; Barbetta et al., 2017; Pham et al., 2021; Weng et al., 2023; Yu et al., 2023).

Probabilistic and robust forecasting studies show that uncertainty should widen with the forecast horizon and should be evaluated alongside point accuracy (Leffrang & Müller, 2024; Barbetta et al., 2023; Astagneau et al., 2024; Papacharalampous & Tyralis, 2020; Zou et al., 2023).

Multi-step forecasting research distinguishes recursive, direct, and sequence-based strategies; each imposes different data demands and error-propagation risks (Solgi et al., 2021; Jeon & Seong, 2022).

Machine-learning reviews and benchmark studies likewise recommend chronological evaluation, strong simple comparators, and transparent metrics (Kratzert et al., 2018; Gauch et al., 2021; Nearing et al., 2021; Shen, 2018; Tyralis et al., 2019; Chen & Guestrin, 2016; Lim et al., 2021; Makridakis et al., 2018; Bergmeir et al., 2018; Knoben et al., 2019; Mosavi et al., 2018; Hyndman & Khandakar, 2008; American

Society of Agricultural and Biological Engineers (ASABE), 2015). Random forest and XGBoost offer nonlinear response functions and variable-importance summaries, whereas SARIMA supplies a parsimonious seasonal structure. Deep recurrent models can be effective in data-rich hydrology, but their success does not justify fitting them to an underpowered reservoir series (Gauch et al., 2021; Nearing et al., 2021; Shen, 2018; Tyralis et al., 2019; Chen & Guestrin, 2016; Breiman, 2001).

Climate Scenarios and Decision Use

CMIP6 expands the set of coupled models and socioeconomic forcing pathways available for climate-risk analysis. Reservoir applications typically require bias adjustment, multi-model ensembles, and scenario framing because raw model output does not reproduce local distributions exactly. Trend-preserving corrections and ensemble spread are particularly important when results are used beyond the range of observed operations.

Recent reservoir studies use CMIP6 and related ensembles to examine inflow, reliability, vulnerability, water quality, and adaptive operation under climate change (Onyutha, 2024; Zhang et al., 2022; Azadi et al., 2018; Adeniyi, 2020; Nguyen et al., 2020).

African studies report substantial spatial heterogeneity in projected precipitation and runoff, which cautions against transferring conclusions among basins (Chiarelli et al., 2016; Alemu et al., 2023; Schilling et al., 2020; Quan V. Dau & Kittiwet Kuntiyawichai, 2020; Golfam & Ashofteh, 2022).

Research on adaptive reservoir management increasingly treats projections as conditional stress tests rather than deterministic operating forecasts (Mateus & Tullos, 2017; Sang et al., 2023; Azadi et al., 2021; Op de Hipt et al., 2018; Faye & Akinsanola, 2021).

The scenario architecture used in this study is consistent with CMIP6 experiment design, scenario-based forcing, multivariate bias adjustment, and downscaled daily climate products (Thrasher et al., 2022; Eyring et al., 2016; O'Neill et al., 2016; Cannon, 2018). Because no future release schedule, demand series, or storage–area curve was available, the projections are reported as conditional surface-area scenarios rather than reservoir yield forecasts.

African Water Security and Small-Reservoir Evidence

Water-security research in sub-Saharan Africa emphasizes that physical scarcity, infrastructure, institutional coordination, informal access, and unequal exposure interact. Remote sensing can reduce an information gap, but actionable governance still requires local ownership, accessible indicators, and a path from anomaly detection to field verification.

Recent regional work links hydrological monitoring to questions of access, collective action, climate adaptation, and distributional risk (Bisung, 2021; Dery & Bisung, 2022; Cecchi et al., 2020; Mutanda & Nhamo, 2024; Banda et al., 2022).

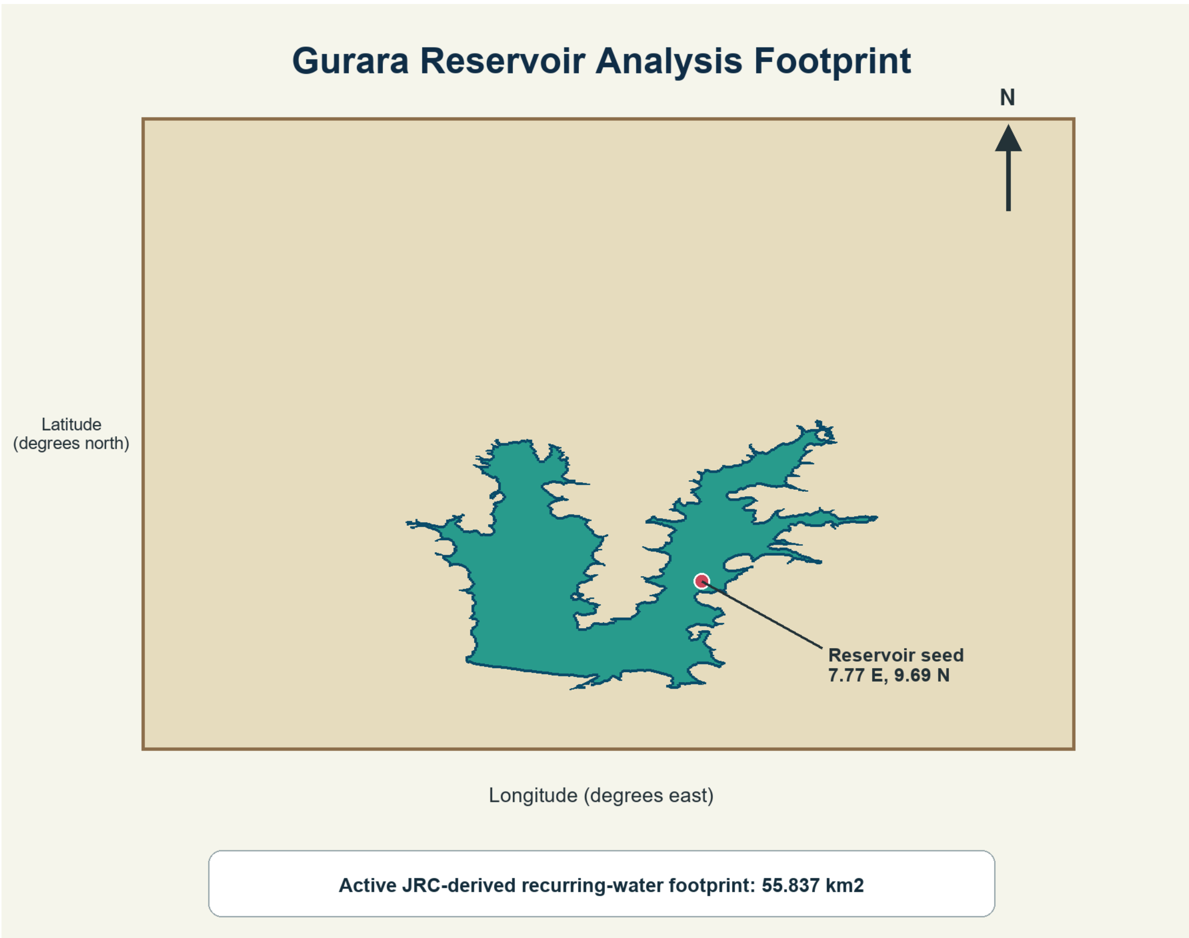
Materials and Methods

Study Area

Gurara Reservoir is located in north-central Nigeria within a seasonally wet tropical setting. The reservoir serves regional water-resource functions and is subject to pronounced wet- and dry-season hydroclimatic forcing. The analysis used a reproducible reservoir boundary and a recurring-water footprint of 55.837 km^2 derived from the Joint Research Centre surface-water occurrence product. The footprint was used as a physically interpretable ceiling for surface-area estimates and projection intervals.

Figure 1

Location and analysis extent of Gurara Reservoir, Nigeria



Note. The map was generated by the implemented project workflow from the analysis geometry and contextual basemap layers.

Research Design and Data

The design comprised six linked stages: Earth-observation extraction, quality control, cross-sensor index assessment, hydroclimatic analysis, horizon-specific forecast evaluation, and CMIP6-informed scenario simulation. Landsat 5, 7, 8, and 9 supplied the January 2001–April 2026 monthly record. Sentinel-2 observations supported recent cross-sensor comparison. CHIRPS supplied rainfall, and TerraClimate supplied precipitation, minimum and maximum temperature, actual and potential evapotranspiration, and PDSI. Five NEX-GDDP-CMIP6 models under SSP2-4.5 and SSP5-8.5 produced ten scenario members for 2026–2030.

Table 1

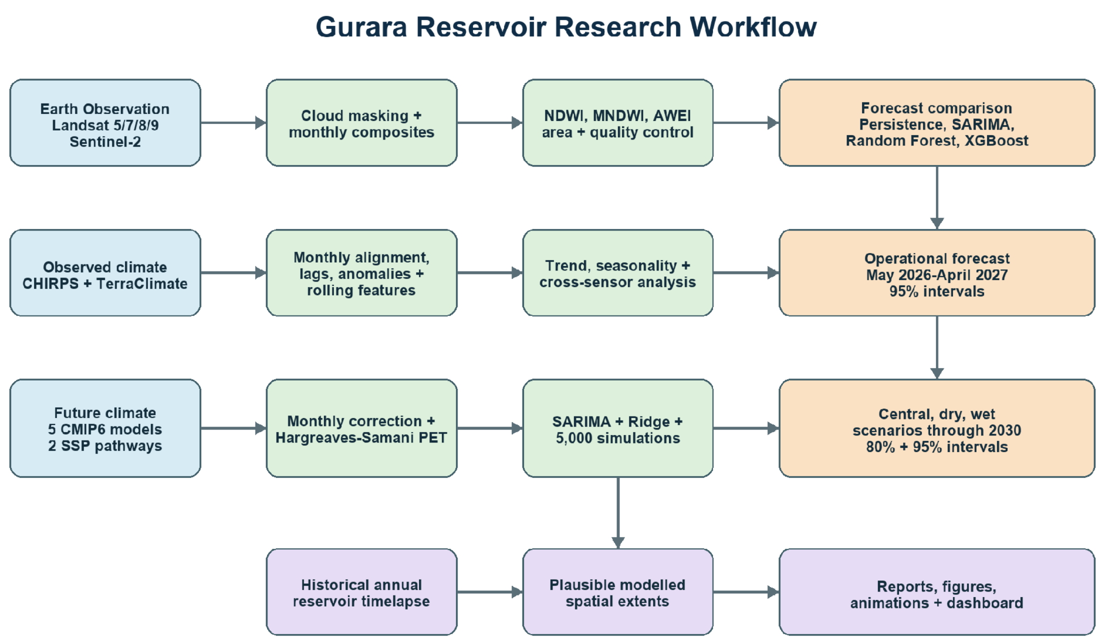
Data sources and analytical roles

Dataset	**Period / resolution**	**Role**	**Principal caution**
Landsat 5/7/8/9 surface reflectance	Monthly, Jan. 2001–Apr. 2026; 30 m	Primary water-area time series	Clouds, sensor transitions, Landsat 7 scan gaps
Sentinel-2 surface reflectance	Recent overlap; 10–20 m	Cross-sensor comparison	Not field ground truth
JRC Global Surface Water	Historical occurrence	Recurring-water mask and 55.837 km^2 ceiling	Historical occurrence is not bathymetry
CHIRPS	Monthly, 2001–2026	Rainfall forcing	Gridded-product bias and scale mismatch
TerraClimate	Monthly, 2001–2026	Temperature, AET, PET, PDSI	Modelled gridded variables
NEX-GDDP-CMIP6	Monthly scenarios, 2026–2030	Conditional climate perturbations	Scenario and model uncertainty

Note. AET = actual evapotranspiration; PET = potential evapotranspiration; PDSI = Palmer Drought Severity Index.

Figure 2

End-to-end monitoring, forecasting, and scenario workflow



Satellite Processing and Quality Control

Monthly surface reflectance composites were filtered using sensor quality-assurance information and an explicit clear-observation requirement. NDWI, MNDWI, and AWEI variants were calculated from sensor-appropriate bands. Candidate thresholds were evaluated for physically plausible water delineation and agreement with Sentinel-2 during overlapping valid months. Months that failed the valid-pixel criterion were retained as missing rather than converted to zero or automatically interpolated. This decision prevented cloud contamination from being interpreted as reservoir contraction.

MNDWI was selected for the final series because it provided the strongest practical balance of cross-sensor correlation, bias, and error. Agreement was summarized with Pearson correlation, mean bias, MAE, and RMSE over paired valid months. Because both sensors observe reflectance rather than surveyed shoreline position, the analysis uses the term cross-sensor agreement rather than accuracy validation.

Trend, Seasonality, and Climate Association

Long-term monotonic change was assessed with the Mann–Kendall test and Sen's slope (Sen, 1968). Monthly climatology was computed from all valid observations grouped by calendar month. Climate association was assessed using Pearson correlation and Kendall's tau on concurrent valid pairs. These analyses were descriptive. No causal claim was made because operating decisions, abstraction, reservoir geometry, and lagged basin processes were unavailable.

Forecast Models and Chronological Evaluation

The modelling series preserved the project-defined chronology: training through December 2018, validation from January 2019 through December 2021, and testing from January 2022 onward. Persistence and seasonal persistence formed the benchmarks. The statistical candidate was SARIMA(1,1,1)(1,1,1)$_{12}$. Random forest and XGBoost models used lagged water area, calendar-season terms, and available hydroclimatic covariates. Tree models were fitted directly by horizon to avoid recursive error accumulation.

Performance was evaluated separately at 1, 3, 6, and 12 months using RMSE, MAE, coefficient of determination, mean absolute percentage error, Nash–Sutcliffe efficiency (NSE), and Kling–Gupta efficiency (KGE). Model selection prioritized the held-out test period and retained simple models when they outperformed more complex alternatives. The longest continuous valid run contained only 23 months; therefore LSTM, GRU, and transformer results were not presented as defensible evidence.

Operational Forecast and CMIP6-Informed Scenarios

The selected SARIMA specification was refitted through April 2026 to generate an operational May 2026–April 2027 forecast. Ninety-five-percent prediction intervals were clipped to the physical range from zero to 55.837 km^2. For 2026–2030, monthly CMIP6 series were bias-adjusted against 2001–2014 observations using calendar-month ratios for rainfall and PET and additive correction for temperature. A SARIMA baseline was combined with a ridge-regression climate-response term. Five thousand simulations propagated statistical and climate-member uncertainty.

Three planning scenarios were reported. The central case retained bias-corrected climate inputs; the dry stress test multiplied rainfall by 0.85 and PET by 1.10 and added 1 °C; the wet stress test multiplied rainfall by 1.15 and PET by 0.95 and subtracted 0.5 °C. Values after April 2027 are explicitly long-range conditional scenarios, not operational forecasts.

Results

Observation Availability and Cross-Sensor Agreement

The analysis covered 304 calendar months, of which 196 produced valid Landsat area estimates (64.5%). Missingness was therefore 35.5%, and the longest uninterrupted valid run was 23 months. This availability supported descriptive analysis and parsimonious forecasting but limited the effective sample size for high-capacity sequence models.

Seventy-three months had paired valid Landsat and Sentinel-2 estimates. For the selected MNDWI workflow, Pearson correlation was .540, MAE was 3.463 km^2, RMSE was 6.838 km^2, and the mean Landsat-minus-Sentinel-2 bias was −0.967 km^2. The near-zero average bias relative to the footprint was encouraging, but the error magnitude indicates material month-level boundary disagreement.

Table 2

Cross-sensor agreement and data completeness

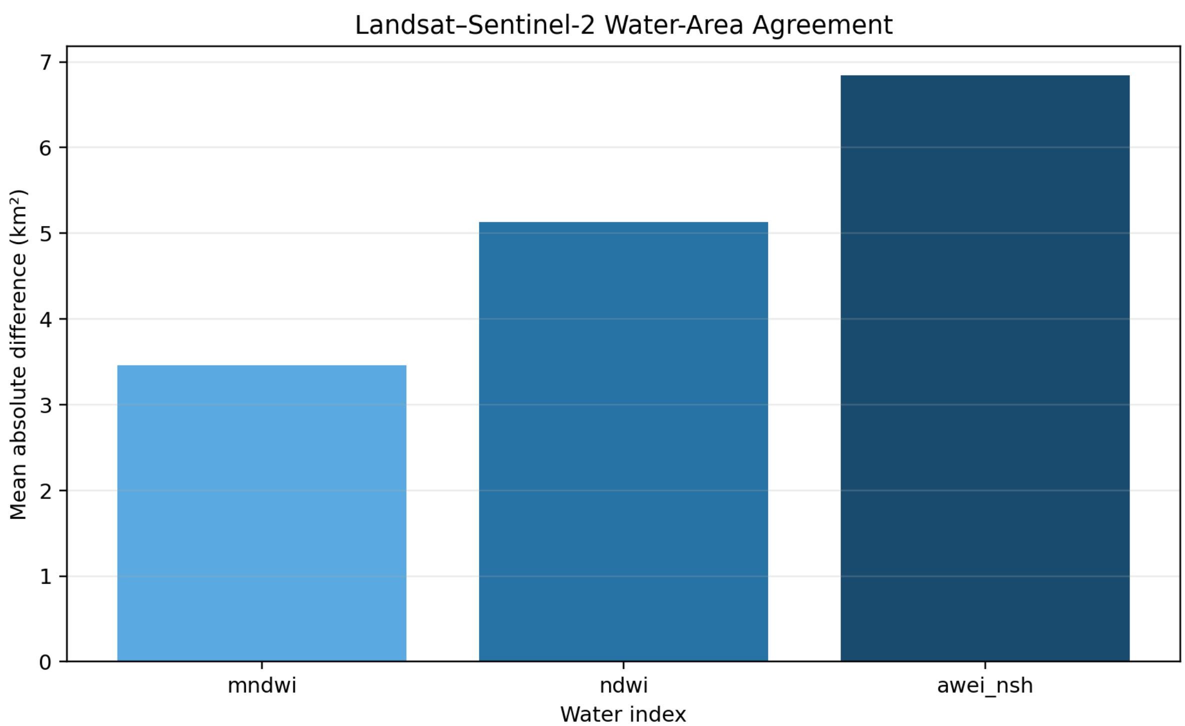
Indicator	Value	Interpretation
Calendar months	304	January 2001–April 2026
Valid Landsat months	196 (64.5%)	Missing months retained as missing
Longest continuous valid run	23 months	Constrains sequence-model complexity
Paired Landsat–Sentinel-2 months	73	Cross-sensor comparison sample
Pearson correlation	.540	Moderate same-month agreement

Indicator	Value	Interpretation
MAE / RMSE	3.463 / 6.838 km²	Material month-level disagreement remains
Mean bias	−0.967 km²	Landsat lower on average
MNDWI agreement score	.625	Best composite agreement; selected
NDWI: r / MAE / RMSE / bias	.462 / 5.128 / 9.594 / −3.843 km²	Composite agreement score = .384
AWEI(nsh): r / MAE / RMSE / bias	.760 / 6.841 / 13.022 / 3.588 km²	High correlation but larger errors; score = .318

Note. Agreement metrics compare two satellite products and do not constitute field validation.

Figure 3

Landsat and Sentinel-2 cross-sensor comparison



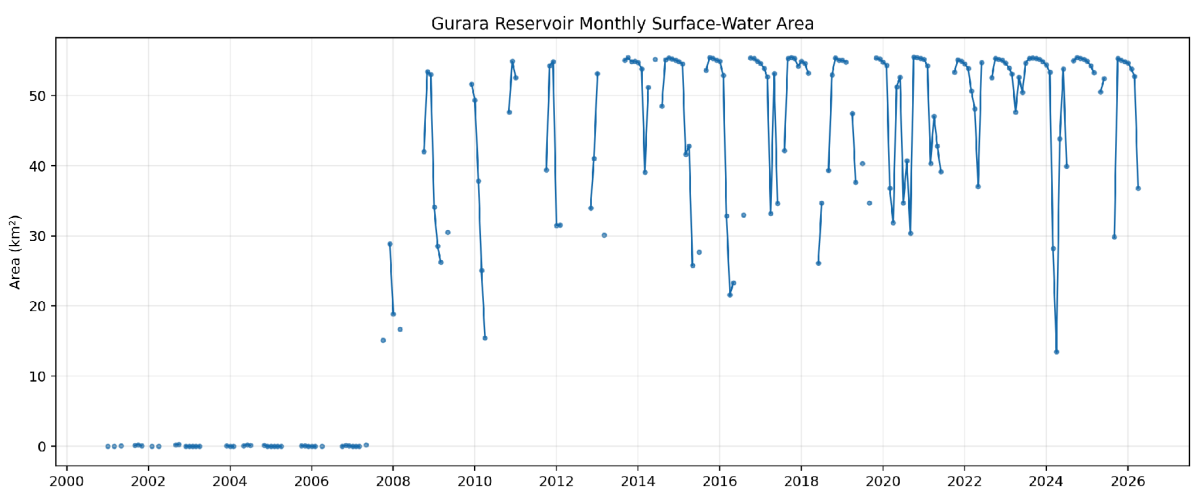
Long-Term Change and Seasonality

The valid Landsat observations show a strong positive monotonic pattern across the full record (Mann–Kendall $\tau = .431$, $p = 3.19 \times 10^{-19}$). Sen's slope was 1.817 km^2 year^{-1}. However, the record begins before the present impoundment reached its later extent; the statistic therefore combines construction, filling, hydrological variability, and subsequent operations. It must not be interpreted as an externally forced climate trend.

The seasonal cycle was pronounced. Mean surface area was lowest in April (27.285 km^2) and highest in August (43.908 km^2), a mean seasonal amplitude of 16.623 km^2. The rise into the wet season and contraction toward the late dry season provide a physically plausible pattern, although the timing of releases and abstractions could modify individual years.

Figure 4

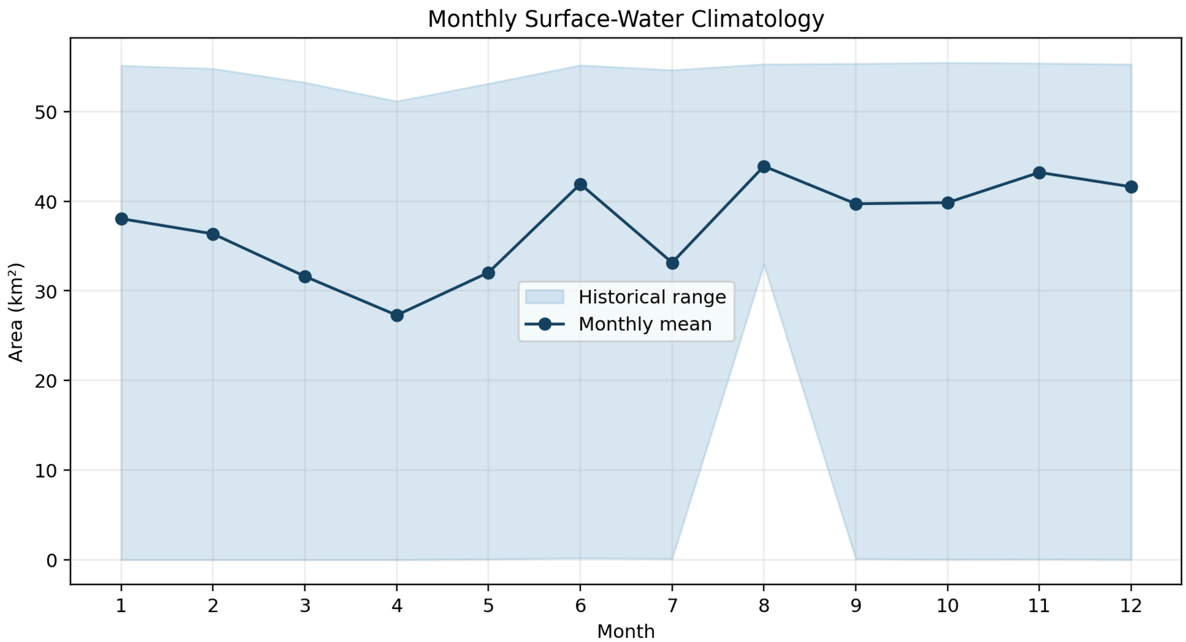
Monthly Gurara Reservoir surface-water area, January 2001–April 2026



Note. Gaps indicate months that did not satisfy the valid-observation criteria; they are not zero-area observations.

Figure 5

Monthly climatology of Gurara Reservoir surface-water area



Climate Associations

Concurrent rainfall association was weak: CHIRPS rainfall had Pearson r = .037 and Kendall τ = −.052 (p = .285), while TerraClimate precipitation had r = −.036 and τ = −.063 (p = .226). Mean temperature had r = −.144 and Kendall τ = −.146 (p = .003). AET had a small positive association (r = .094; τ = .149, p = .003), and PET was weakly negative (r = −.045; τ = −.092, p = .066). PDSI had the largest absolute association (r = −.454; τ = −.271, p = 5.41×10^{-8}). The sign of PDSI should be interpreted within the TerraClimate convention and the mixed structural history of the reservoir, not as a standalone causal effect.

Table 3

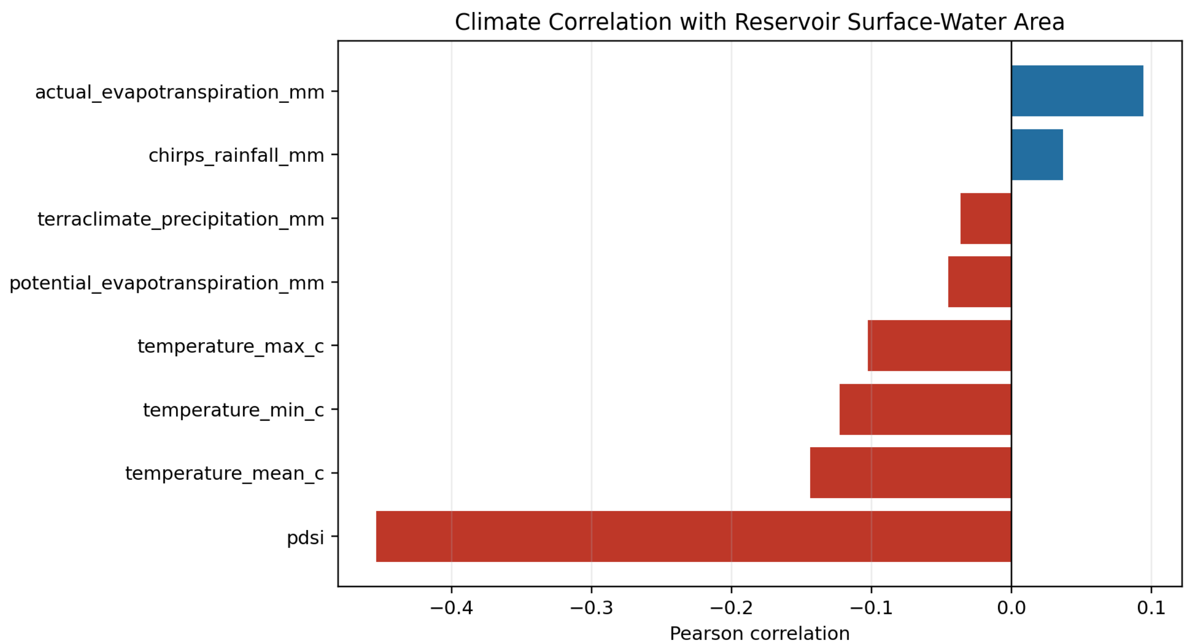
Concurrent associations between surface-water area and climate variables

Variable	n	Pearson r	Kendall τ	p
CHIRPS rainfall	196	0.037	-0.052	0.285
TerraClimate precipitation	183	-0.036	-0.063	0.226
Minimum temperature	183	-0.123	-0.161	0.00121
Maximum temperature	183	-0.102	-0.140	0.00484
Mean temperature	183	-0.144	-0.146	0.00348
AET	183	0.094	0.149	0.00282
PET	183	-0.045	-0.092	0.0662
PDSI	183	-0.454	-0.271	5.41e-08

Note. p values refer to Kendall's tau. These are descriptive same-month associations, not causal estimates.

Figure 6

Correlation of monthly surface-water area with hydroclimatic variables



Forecast Performance

Forecast skill was horizon-dependent. On the held-out test period, SARIMA ranked first at one month (RMSE = 8.014 km²; MAE = 5.393 km²; NSE = .075; KGE = .516) and three months (RMSE = 9.648 km²; MAE = 7.011 km²; NSE = −.341; KGE = .317). Seasonal persistence ranked first at six months, while persistence ranked first at twelve months; both had RMSE = 9.836 km², MAE = 5.029 km², NSE = −.292, and KGE = .088. Negative NSE at longer horizons indicates that predictive value was limited despite competitive ranking among candidates.

Table 4

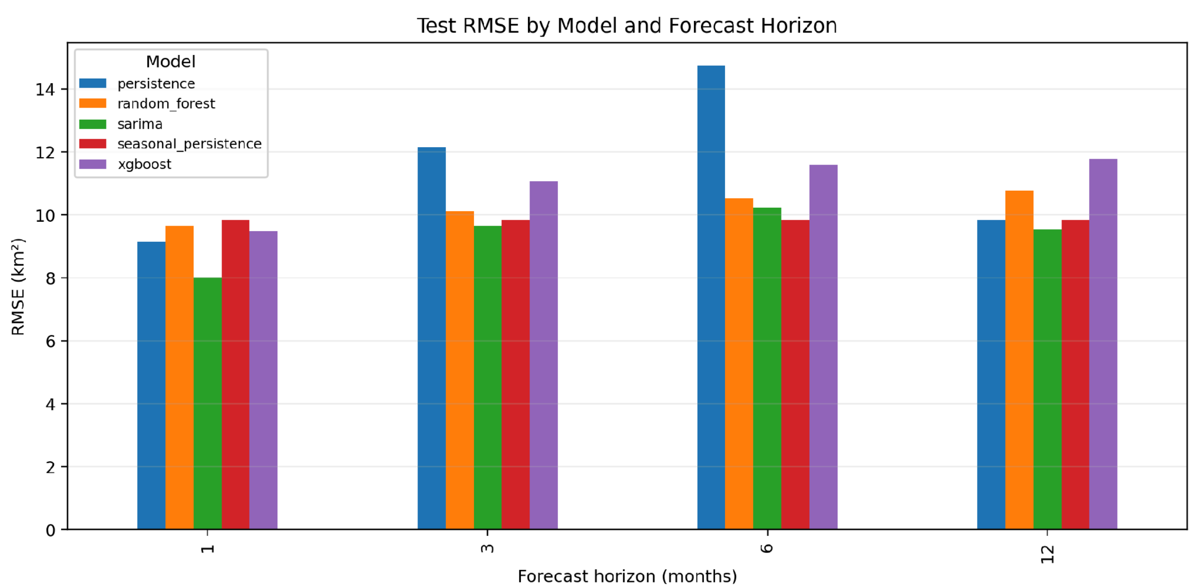
Best held-out test model by forecast horizon

Horizon (months)	Model	n	RMSE	MAE	NSE	KGE
1	sarima	52	8.014	5.393	0.075	0.516
3	sarima	52	9.648	7.011	-0.341	0.317
6	seasonal_persistence	42	9.836	5.029	-0.292	0.088
12	persistence	42	9.836	5.029	-0.292	0.088

Note. Errors are in km². The model labelled best is the highest-ranked candidate under the implemented multi-metric comparison; negative NSE indicates limited predictive skill relative to the observed-mean benchmark.

Figure 7

RMSE comparison across forecasting models and horizons



Operational Forecast

The May 2026 point forecast was 41.890 km², with a 95% prediction interval from 29.449 to 54.331 km². The point forecast rose through the 2026 wet season, reaching 55.639 km² in December, before declining to 40.936 km² by April 2027. The April 2027 interval widened to 19.735–55.837 km² and reached the physical upper bound, illustrating how rapidly uncertainty dominates at longer leads.

Table 5

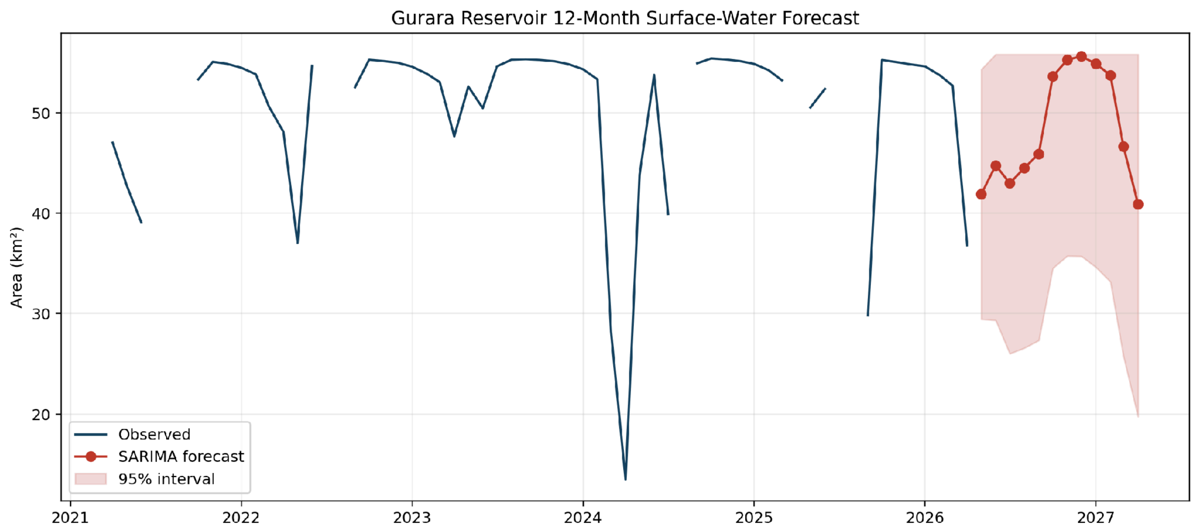
Operational SARIMA forecast from the April 2026 origin

Month	Forecast (km²)	Lower 95%	Upper 95%
May. 2026	41.890	29.449	54.331
Jun. 2026	44.776	29.334	55.837
Jul. 2026	42.992	26.009	55.837
Aug. 2026	44.504	26.574	55.837
Sep. 2026	45.930	27.335	55.837
Oct. 2026	53.645	34.536	55.837
Nov. 2026	55.292	35.754	55.837
Dec. 2026	55.639	35.723	55.837
Jan. 2027	54.890	34.627	55.837
Feb. 2027	53.765	33.177	55.837
Mar. 2027	46.640	25.740	55.837
Apr. 2027	40.936	19.735	55.837

Note. Intervals were constrained to the physical range 0–55.837 km².

Figure 8

Operational surface-area forecast, May 2026–April 2027



CMIP6-Informed Scenarios Through 2030

Scenario annual means remained relatively close compared with their uncertainty. In 2030, the central annual mean was 50.120 km², the dry stress-test mean was 47.936 km², and the wet stress-test mean was 50.899 km². Thus, the dry–wet separation in annual means was 2.963 km². By contrast, the annual-mean uncertainty bands were broad and the monthly upper tails frequently reached the 55.837 km² footprint ceiling.

Table 6

Projected 2030 annual mean surface-water area by scenario

Scenario	Mean	Lower 80%	Upper 80%	Lower 95%	Upper 95%
Central	50.120	26.190	55.837	13.224	55.837
Dry	47.936	23.554	55.837	10.702	55.837
Wet	50.899	27.562	55.837	14.903	55.837

Note. All values are km². These are conditional scenario outcomes, not operational forecasts.

At the final scenario month, December 2030, the central and wet point trajectories reached the 55.837 km² physical ceiling, while the dry trajectory was 54.797 km². The lower 95% outcomes ranged from 14.763 km² in the dry case to 20.639 km² in the wet case, confirming that end-of-horizon uncertainty is much larger than the separation among scenario point trajectories.

Table 7

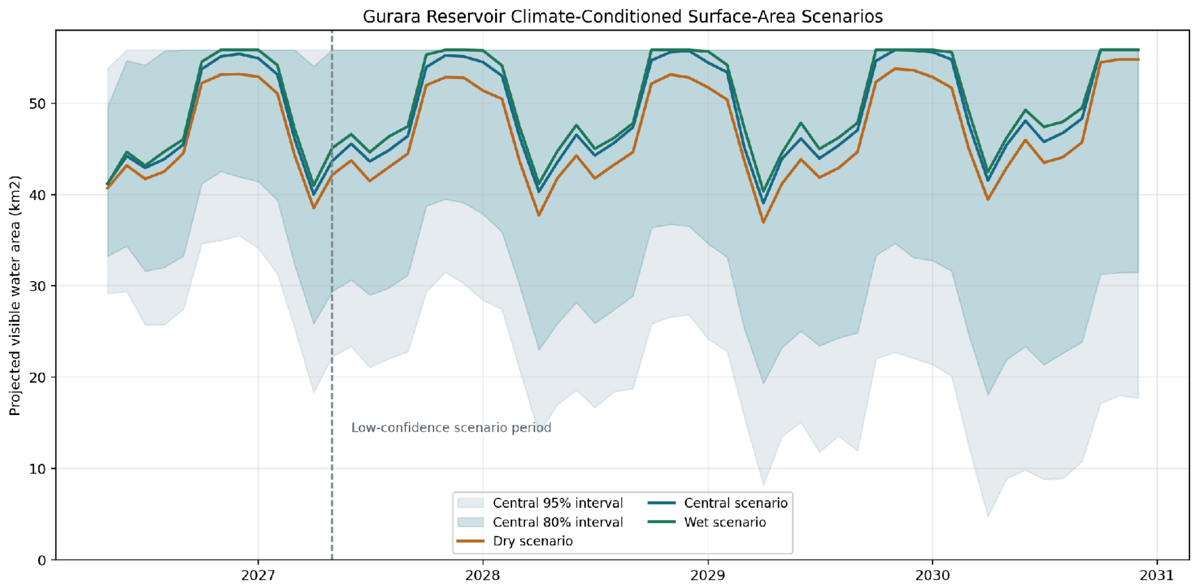
December 2030 conditional surface-area outcomes

Scenario	Point	Lower 80%	Upper 80%	Lower 95%	Upper 95%
Central	55.837	31.481	55.837	17.695	55.837
Dry	54.797	27.983	55.837	14.763	55.837
Wet	55.837	35.005	55.837	20.639	55.837

Note. All values are km²; horizon = 56 months; confidence label = low. Upper outcomes are bounded by the 55.837 km² recurring-water footprint.

Figure 9

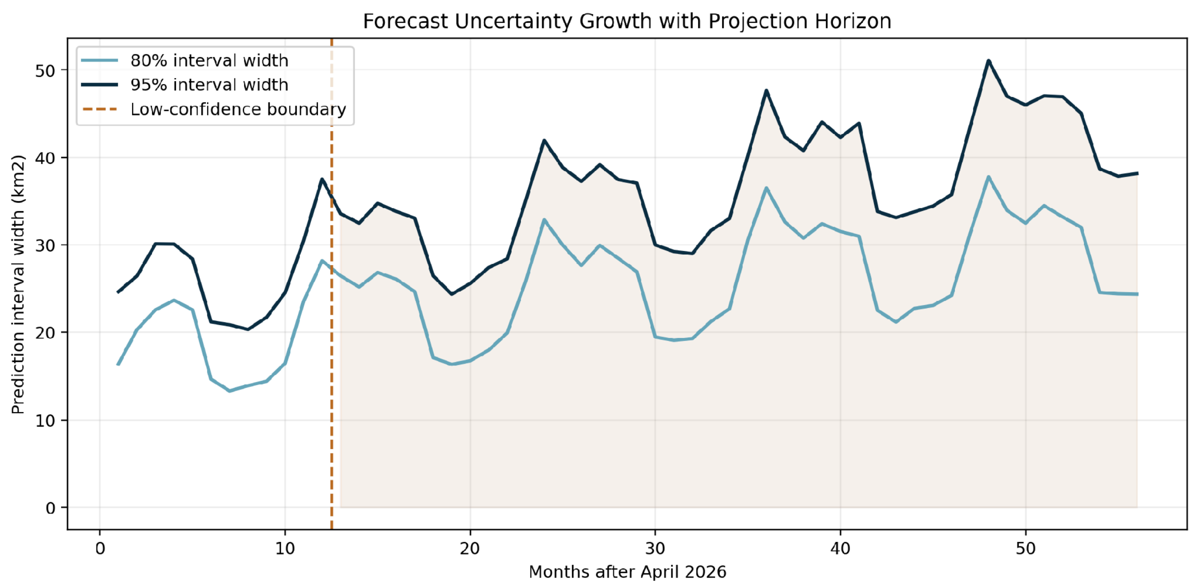
Monthly surface-area scenarios through December 2030



Note. Central, dry, and wet cases are conditional perturbations of the bias-corrected CMIP6-informed climate response.

Figure 10

Growth of projection uncertainty with lead time



Discussion

What the Satellite Record Reveals

The Gurara record demonstrates the principal strength of satellite monitoring: it reconstructs a long history where consistent in situ records are not available to the analysis. The observed annual and seasonal dynamics are large enough to be operationally relevant, and the monthly climatology supplies a transparent seasonal reference. However, the same record also demonstrates why remote sensing must be quality-aware. More than one third of months were invalid, and even two modern optical sensors disagreed by several square kilometres in individual months.

The positive Sen slope should be read as a history-of-impoundment statistic. A trend beginning in 2001 crosses a major structural transition from pre-impoundment conditions to the later reservoir state. Detrending or fitting one stationary process across that transition would not remove the underlying regime change. A future causal study should separate construction and filling, stable-operation years, drought episodes, and any known rule changes using documented intervention dates.

Why Simple Forecasts Remained Competitive

SARIMA's short-horizon advantage is consistent with a system dominated by autocorrelation and seasonality. Its deterioration by three months and the return of persistence-based winners at six and twelve months reveal a limited predictive information set, not a failure of machine learning alone. Tree ensembles cannot infer release schedules or basin state that are absent from their inputs. With only 23 consecutive valid months at maximum, recurrent and transformer architectures would be particularly vulnerable to memorization, unstable validation, and exaggerated apparent skill.

The negative longer-horizon NSE values are substantively important. Ranking first among candidate models does not imply adequacy for consequential operating decisions. The correct interpretation is that six- and twelve-month point forecasts should be used as broad seasonal references, accompanied by uncertainty and updated whenever a new valid observation becomes available.

Climate Signal and Scenario Interpretation

Weak same-month rainfall correlations do not establish that rainfall is unimportant. Reservoir inflow integrates rainfall over space and time, antecedent soil moisture, routing, and upstream abstractions. Conversely, the stronger PDSI association may reflect persistent basin water balance, seasonal covariance, and the structural evolution of the reservoir. Lagged distributed-basin predictors and observed inflow would be required to distinguish these mechanisms.

The modest separation among 2030 scenario means relative to their uncertainty suggests that scenario forcing is not the dominant source of near-term predictive variance in the present framework. Data gaps, model residuals, unknown operations, and a capped area response are at least as important. The scenario envelope should therefore support stress testing for example, whether monitoring or demand-management thresholds remain robust across dry and wet cases rather than claims about the exact surface area in a future month.

Implications for Reservoir Monitoring and Management

A practical implementation should operate as an evidence ladder. First, compare each new valid satellite estimate with the month-specific climatology. Second, classify departures using both point estimates and cross-sensor disagreement. Third, update the short-range SARIMA forecast. Fourth, escalate persistent anomalies for field verification using gauge levels, intake observations, and operator logs. This design uses satellite information to target scarce field effort instead of treating a remotely sensed signal as an automatic operating instruction.

For planning, managers can use the central, dry, and wet trajectories to test whether decisions remain acceptable across plausible climates. Examples include timing supplementary surveys, scheduling shoreline inspections, anticipating intake-access constraints, and prioritizing data acquisition before uncertainty becomes too wide. Storage, yield, and release decisions require additional calibration to bathymetry and operating constraints.

Strengths, Limitations, and Future Research

The study's strengths are its long observation window, explicit missing-data treatment, independent sensor comparison, chronological out-of-sample testing, inclusion of persistence baselines, horizon-specific evaluation, physical clipping, and transparent separation between operational and scenario products. The reproducible output tables allow every value reported here to be traced to the implemented workflow.

Five limitations bound interpretation. First, surface area is not storage; without an area–elevation–volume relation, volumetric change cannot be estimated. Second, cross-sensor agreement is not ground truth. Third, optical imagery remains vulnerable to clouds, haze, shoreline vegetation, and mixed pixels. Fourth, the climate analysis omits observed inflow, releases, abstraction, groundwater exchange, and operating rules. Fifth, the series contains structural change and limited continuous data, reducing the reliability of high-capacity and long-horizon models.

Priority future work should collect bathymetry, gauge level, inflow, release, abstraction, and water-demand data; obtain GPS shoreline observations across wet and dry seasons; incorporate Sentinel-1 radar to reduce cloud gaps; evaluate state-space models that propagate observation error; test lagged and catchment-aggregated climate predictors; detect operational regimes explicitly; and establish pre-registered rolling-origin validation. Only after a sufficiently long, contiguous, and independently validated series exists should deep sequence models be compared.

Conclusion

This study provides a comprehensive, uncertainty-aware account of Gurara Reservoir surface-water dynamics from 2001 to 2026 and conditional trajectories through 2030. MNDWI produced the most defensible cross-sensor series, revealing pronounced seasonality and a positive long-term change dominated in part by reservoir establishment. SARIMA offered the best short-horizon performance, whereas simple persistence remained difficult to beat at longer horizons. Climate correlations were generally weak except for PDSI, and scenario differences were smaller than the expanding uncertainty envelope.

The principal result is therefore not a claim of precise long-range prediction. It is a defensible monitoring-and-forecasting framework that exposes data gaps, compares methods honestly, and makes uncertainty visible. Coupled with bathymetry and operating records, the same framework could evolve from surface-area intelligence into reservoir storage and decision support. Until then, its appropriate role is screening, seasonal context, anomaly detection, and conditional planning.

Data and Code Availability

The processed datasets, model-evaluation tables, forecast tables, scenario outputs, figures, and executable analysis scripts are available in the associated Gurara project repository. A public repository URL and archived DOI should be inserted before submission.

References

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, *5*(1), 170191. https://doi.org/10.1038/sdata.2017.191

Adeniyi, M. O. (2020). Possible influence of climate change on water balance over West Africa under the global warming levels of 2 and 3 °C. *Journal of Water and Climate Change*, *12*(5), 1654-1673. https://doi.org/10.2166/wcc.2020.094

Alemu, M. G., Wubneh, M. A., Sahlu, D., & Zimale, F. A. (2023). Spatiotemporal change of climate extremes under the projection of CMIP6 model analysis over Awash Basin, Ethiopia. *Sustainable Water Resources Management*, *9*(6), 195. https://doi.org/10.1007/s40899-023-00981-5

Alsilibe, F., Bene, K., Bilal, G., Alghafli, K., & Shi, X. (2023). Accuracy Assessment and Validation of Multi-Source CHIRPS Precipitation Estimates for Water Resource Management in the Barada Basin, Syria. *Remote Sensing*, *15*(7), 1778. https://doi.org/10.3390/rs15071778

American Society of Agricultural and Biological Engineers (ASABE) (2015). Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria. *Transactions of the ASABE*, *58*(6), 1763-1785. https://doi.org/10.13031/trans.58.10715

Astagneau, P. C., Bourgin, F., Andréassian, V., & Perrin, C. (2024). Lead-time-dependent calibration of a flood forecasting model. *Journal of Hydrology*, *644*, 132119. https://doi.org/10.1016/j.jhydrol.2024.132119

Azadi, F., Ashofteh, P. S., & Loáiciga, H. A. (2018). Reservoir Water-Quality Projections under Climate-Change Conditions. *Water Resources Management*, *33*(1), 401-421. https://doi.org/10.1007/s11269-018-2109-z

Azadi, F., Ashofteh, P. S., Shokri, A., & Loáiciga, H. A. (2021). Simulation-Optimization of Reservoir Water Quality under Climate Change. *Journal of Water Resources Planning and Management*, *147*(9), 04021054. https://doi.org/10.1061/(asce)wr.1943-5452.0001436

Banda, V. D., Dzwairo, R. B., Singh, S. K., & Kanyerere, T. (2022). Hydrological Modelling and Climate Adaptation under Changing Climate: A Review with a Focus in Sub-Saharan Africa. *Water*, *14*(24), 4031. https://doi.org/10.3390/w14244031

Barbetta, S., Coccia, G., Moramarco, T., Brocca, L., & Todini, E. (2017). The multi temporal/multi-model approach to predictive uncertainty assessment in real-time flood forecasting. *Journal of Hydrology*, *551*, 555-576. https://doi.org/10.1016/j.jhydrol.2017.06.030

Barbetta, S., Sahoo, B., Bonaccorsi, B., Nanda, T., Chatterjee, C., Moramarco, T., & Todini, E. (2023). Addressing effective real-time forecasting inflows to dams through predictive uncertainty estimate. *Journal of Hydrology*, *620*, 129512. https://doi.org/10.1016/j.jhydrol.2023.129512

Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, *120*, 70-83. https://doi.org/10.1016/j.csda.2017.11.003

Biancamaria, S., Lettenmaier, D. P., & Pavelsky, T. M. (2016). The SWOT Mission and Its Capabilities for Land Hydrology. *Surveys in Geophysics*, *37*(2), 307-337. https://doi.org/10.1007/s10712-015-9346-y

Bisung, E. (2021). Community based collective action for safe water in rural sub-Saharan Africa. *Water Security*, *13*, 100088. https://doi.org/10.1016/j.wasec.2021.100088

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5-32. https://doi.org/10.1023/a:1010933404324

Busker, T., de Roo, A., Gelati, E., Schwatke, C., Adamovic, M., Bisselink, B., Pekel, J. F., & Cottam, A. (2019). A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry. *Hydrology and Earth System Sciences*, *23*(2), 669-690. https://doi.org/10.5194/hess-23-669-2019

Cai, Y., Shi, Q., & Liu, X. (2024). Spatiotemporal Mapping of Surface Water Using Landsat Images and Spectral Mixture Analysis on Google Earth Engine. *Journal of Remote Sensing*, *4*, 0117. https://doi.org/10.34133/remotesensing.0117

Cannon, A. J. (2018). Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dynamics*, *50*(1-2), 31-49. https://doi.org/10.1007/s00382-017-3580-6

Cecchi, P., Forkuor, G., Cofie, O., Lalanne, F., Poussin, J. C., & Jamin, J. Y. (2020). Small Reservoirs, Landscape Changes and Water Quality in Sub-Saharan West Africa. *Water*, *12*(7), 1967. https://doi.org/10.3390/w12071967

Chen, T., & Guestrin, C. (2016). XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. https://doi.org/10.1145/2939672.2939785

Chiarelli, D. D., Davis, K. F., Rulli, M. C., & D'Odorico, P. (2016). Climate change and large-scale land acquisitions in Africa: Quantifying the future impact on acquired water resources. *Advances in Water Resources*, *94*, 231-237. https://doi.org/10.1016/j.advwatres.2016.05.016

Chipman, J. W. (2019). A Multisensor Approach to Satellite Monitoring of Trends in Lake Area, Water Level, and Volume. *Remote Sensing*, *11*(2), 158. https://doi.org/10.3390/rs11020158

Cho, H., & Ryu, D. (2024). Reservoir Water Surface Area Estimation Using Sentinel-1 and Sentinel-2 Imagery. *Korean Journal of Remote Sensing*, *40*(5-1), 643-656. https://doi.org/10.7780/kjrs.2024.40.5.1.18

Das, P., Hossain, F., Khan, S., Biswas, N. K., Lee, H., Piman, T., Meechaiya, C., Ghimire, U., & Hosen, K. (2022). Reservoir Assessment Tool 2.0: Stakeholder driven improvements to satellite remote sensing based reservoir monitoring. *Environmental Modelling & Software*, *157*, 105533. https://doi.org/10.1016/j.envsoft.2022.105533

Declaro, A., & Kanae, S. (2024). Enhancing Surface Water Monitoring through Multi-Satellite Data-Fusion of Landsat-8/9, Sentinel-2, and Sentinel-1 SAR. *Remote Sensing*, *16*(17), 3329. https://doi.org/10.3390/rs16173329

Dery, F., & Bisung, E. (2022). Exploring informalities, inequalities, and health risk in the informal water sector in Sub-Saharan Africa. *Water Security*, *17*, 100128. https://doi.org/10.1016/j.wasec.2022.100128

Du, J., Zhou, H., Jacinthe, P. A., & Song, K. (2023). Retrieval of lake water surface albedo from Sentinel-2 remote sensing imagery. *Journal of Hydrology*, *617*, 128904. https://doi.org/10.1016/j.jhydrol.2022.128904

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937-1958. https://doi.org/10.5194/gmd-9-1937-2016

Faye, A., & Akinsanola, A. A. (2021). Evaluation of extreme precipitation indices over West Africa in CMIP6 models. *Climate Dynamics*, *58*(3-4), 925-939. https://doi.org/10.1007/s00382-021-05942-2

Fessehaye, M., Franke, J., & Brönnimann, S. (2022). Evaluation of satellite-based (CHIRPS and GPM) and reanalysis (ERA5-Land) precipitation estimates over Eritrea. *Meteorologische Zeitschrift*, *31*(5), 401-413. https://doi.org/10.1127/metz/2022/1111

Feyisa, G. L., Meilby, H., Fensholt, R., & Proud, S. R. (2014). Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sensing of Environment*, *140*, 23-35. https://doi.org/10.1016/j.rse.2013.08.029

Fisher, A., Flood, N., & Danaher, T. (2016). Comparing Landsat water index methods for automated water classification in eastern Australia. *Remote Sensing of Environment*, *175*, 167-182. https://doi.org/10.1016/j.rse.2015.12.055

Flood, N. (2017). Comparing Sentinel-2A and Landsat 7 and 8 Using Surface Reflectance over Australia. *Remote Sensing*, *9*(7), 659. https://doi.org/10.3390/rs9070659

Foga, S., Scaramuzza, P. L., Guo, S., Zhu, Z., Dilley, R. D., Beckmann, T., Schmidt, G. L., Dwyer, J. L., Joseph Hughes, M., & Laue, B. (2017). Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sensing of Environment*, *194*, 379-390. https://doi.org/10.1016/j.rse.2017.03.026

Frantz, D. (2019). FORCE—Landsat + Sentinel-2 Analysis Ready Data and Beyond. *Remote Sensing*, *11*(9), 1124. https://doi.org/10.3390/rs11091124

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, *2*(1), 150066. https://doi.org/10.1038/sdata.2015.66

Gauch, M., Kratzert, F., Klotz, D., Nearing, G., Lin, J., & Hochreiter, S. (2021). Rainfall–runoff prediction at multiple timescales with a single Long Short-Term Memory network. *Hydrology and Earth System Sciences*, *25*(4), 2045-2062. https://doi.org/10.5194/hess-25-2045-2021

Golfam, P., & Ashofteh, P. S. (2022). Performance Indexes Analysis of the Reservoir-Hydropower Plant System Affected by Climate Change. *Water Resources Management*, *36*(13), 5127-5162. https://doi.org/10.1007/s11269-022-03295-x

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18-27. https://doi.org/10.1016/j.rse.2017.06.031

Gyamfi, C., Tindan, J. Z. O., & Kifanyi, G. E. (2021). Evaluation of CORDEX Africa multi-model precipitation simulations over the Pra River Basin, Ghana. *Journal of Hydrology: Regional Studies*, *35*, 100815. https://doi.org/10.1016/j.ejrh.2021.100815

Hafeez, S., Wong, M. S., Abbas, S., & Asim, M. (2022). Evaluating Landsat-8 and Sentinel-2 Data Consistency for High Spatiotemporal Inland and Coastal Water Quality Monitoring. *Remote Sensing*, *14*(13), 3155. https://doi.org/10.3390/rs14133155

Hollstein, A., Segl, K., Guanter, L., Brell, M., & Enesco, M. (2016). Ready-to-Use Methods for the Detection of Clouds, Cirrus, Snow, Shadow, Water and Clear Sky Pixels in Sentinel-2 MSI Images. *Remote Sensing*, *8*(8), 666. https://doi.org/10.3390/rs8080666

Hu, Q., Li, C., Wang, Z., Liu, Y., & Liu, W. (2022). Continuous Monitoring of the Surface Water Area in the Yellow River Basin during 1986–2019 Using Available Landsat Imagery and the Google Earth Engine. *ISPRS International Journal of Geo-Information*, *11*(5), 305. https://doi.org/10.3390/ijgi11050305

Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, *27*(3). https://doi.org/10.18637/jss.v027.i03

Inman, V. L., & Lyons, M. B. (2020). Automated Inundation Mapping Over Large Areas Using Landsat Data and Google Earth Engine. *Remote Sensing*, *12*(8), 1348. https://doi.org/10.3390/rs12081348

Jeon, Y., & Seong, S. (2022). Robust recurrent network model for intermittent time-series forecasting. *International Journal of Forecasting*, *38*(4), 1415-1425. https://doi.org/10.1016/j.ijforecast.2021.07.004

Jiang, C., Zhang, H., Wang, C., Ge, J., & Wu, F. (2022). Water Surface Mapping from Sentinel-1 Imagery Based on Attention-UNet3+: A Case Study of Poyang Lake Region. *Remote Sensing*, *14*(19), 4708. https://doi.org/10.3390/rs14194708

Katsanos, D., Retalis, A., & Michaelides, S. (2016). Validation of a high-resolution precipitation database (CHIRPS) over Cyprus for a 30-year period. *Atmospheric Research*, *169*, 459-464. https://doi.org/10.1016/j.atmosres.2015.05.015

Keys, T. A., & Scott, D. T. (2017). Monitoring volumetric fluctuations in tropical lakes and reservoirs using satellite remote sensing. *Lake and Reservoir Management*, *34*(2), 154-166. https://doi.org/10.1080/10402381.2017.1402226

Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, *23*(10), 4323-4331. https://doi.org/10.5194/hess-23-4323-2019

Kolarik, N. E., Roopsind, A., Pickens, A., & Brandt, J. S. (2023). A satellite-based monitoring system for quantifying surface water and mesic vegetation dynamics in a semi-arid region. *Ecological Indicators*, *147*, 109965. https://doi.org/10.1016/j.ecolind.2023.109965

Kouakou, C., Paturel, J. E., Satgé, F., Tramblay, Y., Defrance, D., & Rouché, N. (2023). Comparison of gridded precipitation estimates for regional hydrological modeling in West and Central Africa. *Journal of Hydrology: Regional Studies*, *47*, 101409. https://doi.org/10.1016/j.ejrh.2023.101409

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*, *22*(11), 6005-6022. https://doi.org/10.5194/hess-22-6005-2018

Leffrang, D., & Müller, O. (2024). Visualizing Uncertainty in Time Series Forecasts: The Impact of Uncertainty Visualization on Users' Confidence, Algorithmic Advice Utilization, and Forecasting Performance. *Journal of Forecasting*, *44*(4), 1235-1246. https://doi.org/10.1002/for.3222

Li, J., & Roy, D. (2017). A Global Analysis of Sentinel-2A, Sentinel-2B and Landsat-8 Data Revisit Intervals and Implications for Terrestrial Monitoring. *Remote Sensing*, *9*(9), 902. https://doi.org/10.3390/rs9090902

Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, *37*(4), 1748-1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

Lu, L., & Sun, H. (2023). Dynamic monitoring of surface water areas of nine plateau lakes in Yunnan Province using long time-series Landsat imagery based on the Google Earth Engine platform. *Geocarto International*, *38*(1), 2253196. https://doi.org/10.1080/10106049.2023.2253196

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, *34*(4), 802-808. https://doi.org/10.1016/j.ijforecast.2018.06.001

Marra, F., Levizzani, V., & Cattani, E. (2022). Changes in extreme daily precipitation over Africa: Insights from a non-asymptotic statistical approach. *Journal of Hydrology X*, *16*, 100130. https://doi.org/10.1016/j.hydroa.2022.100130

Mateus, M. C., & Tullos, D. (2017). Reliability, Sensitivity, and Vulnerability of Reservoir Operations under Climate Change. *Journal of Water Resources Planning and Management*, *143*(4), 04016085. https://doi.org/10.1061/(asce)wr.1943-5452.0000742

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

Mendoza Paz, S., & Willems, P. (2022). Uncovering the strengths and weaknesses of an ensemble of quantile mapping methods for downscaling precipitation change in Southern Africa. *Journal of Hydrology: Regional Studies*, *41*, 101104. https://doi.org/10.1016/j.ejrh.2022.101104

Monteiro, M., & Costa, M. (2018). A Time Series Model Comparison for Monitoring and Forecasting Water Quality Variables. *Hydrology*, *5*(3), 37. https://doi.org/10.3390/hydrology5030037

Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood Prediction Using Machine Learning Models: Literature Review. *Water*, *10*(11), 1536. https://doi.org/10.3390/w10111536

Mutanda, G. W., & Nhamo, G. (2024). Gendered perspective on water security, rights and conflicts in sub-Saharan Africa: a systematic review. *Frontiers in Water*, *6*, 1399415. https://doi.org/10.3389/frwa.2024.1399415

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., & Thépaut, J. N. (2021). ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, *13*(9), 4349-4383. https://doi.org/10.5194/essd-13-4349-2021

Nawaz, M., Iqbal, M. F., & Mahmood, I. (2021). Validation of CHIRPS satellite-based precipitation dataset over Pakistan. *Atmospheric Research*, *248*, 105289. https://doi.org/10.1016/j.atmosres.2020.105289

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., & Gupta, H. V. (2021). What Role Does Hydrological Science Play in the Age of Machine Learning?. *Water Resources Research*, *57*(3), e2020WR028091. https://doi.org/10.1029/2020wr028091

Nguyen, H., Mehrotra, R., & Sharma, A. (2020). Assessment of Climate Change Impacts on Reservoir Storage Reliability, Resilience, and Vulnerability Using a Multivariate Frequency Bias Correction Approach. *Water Resources Research*, *56*(2), e2019WR026022. https://doi.org/10.1029/2019wr026022

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J. F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., & Sanderson, B. M. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461-3482. https://doi.org/10.5194/gmd-9-3461-2016

Onyutha, C. (2024). Climate change impacts on hydrology and water resources in East Africa considering CMIP3, CMIP5, and CMIP6. *Frontiers in Climate*, *6*, 1453726. https://doi.org/10.3389/fclim.2024.1453726

Op de Hipt, F., Diekkrüger, B., Steup, G., Yira, Y., Hoffmann, T., & Rode, M. (2018). Modeling the impact of climate change on water resources and soil erosion in a tropical catchment in Burkina Faso, West Africa. *CATENA*, *163*, 63-77. https://doi.org/10.1016/j.catena.2017.11.023

Owusu, C., Snigdha, N. J., Martin, M. T., & Kalyanapu, A. J. (2022). PyGEE-SWToolbox: A Python Jupyter Notebook Toolbox for Interactive Surface Water Mapping and Analysis Using Google Earth Engine. *Sustainability*, *14*(5), 2557. https://doi.org/10.3390/su14052557

Page, B. P., Olmanson, L. G., & Mishra, D. R. (2019). A harmonized image processing workflow using Sentinel-2/MSI and Landsat-8/OLI for mapping water clarity in optically variable lake systems. *Remote Sensing of Environment*, *231*, 111284. https://doi.org/10.1016/j.rse.2019.111284

Pan, F. (2024). Accuracy Assessment of Estimated River Water Surface Elevations from Landsat 8 and 9 Imagery among Twenty Water Indices. *Remote Sensing*, *16*(16), 3054. https://doi.org/10.3390/rs16163054

Papacharalampous, G., & Tyralis, H. (2020). Hydrological time series forecasting using simple combinations: Big data testing and investigations on one-year ahead river flow predictability. *Journal of Hydrology*, *590*, 125205. https://doi.org/10.1016/j.jhydrol.2020.125205

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

Perez, M., & Vitale, M. (2023). Landsat-7 ETM+, Landsat-8 OLI, and Sentinel-2 MSI Surface Reflectance Cross-Comparison and Harmonization over the Mediterranean Basin Area. *Remote Sensing*, *15*(16), 4008. https://doi.org/10.3390/rs15164008

Pham, H. T., Marshall, L., & Johnson, F. (2021). Daily time series of river water levels derived from a seasonal linear model using multisource satellite products under uncertainty. *Journal of Hydrology*, *602*, 126783. https://doi.org/10.1016/j.jhydrol.2021.126783

Pickens, A. H., Hansen, M. C., Hancher, M., Stehman, S. V., Tyukavina, A., Potapov, P., Marroquin, B., & Sherani, Z. (2020). Mapping and sampling to characterize global inland water dynamics from 1999 to 2018 with full Landsat time-series. *Remote Sensing of Environment*, *243*, 111792. https://doi.org/10.1016/j.rse.2020.111792

Poméon, T., Jackisch, D., & Diekkrüger, B. (2017). Evaluating the performance of remotely sensed and reanalysed precipitation data over West Africa using HBV light. *Journal of Hydrology*, *547*, 222-235. https://doi.org/10.1016/j.jhydrol.2017.01.055

Prakash, S. (2019). Performance assessment of CHIRPS, MSWEP, SM2RAIN-CCI, and TMPA precipitation products across India. *Journal of Hydrology*, *571*, 50-59. https://doi.org/10.1016/j.jhydrol.2019.01.036

Quan V. Dau, & Kittiwet Kuntiyawichai (2020). Identifying Adaptive Reservoir Operation for Future Climate Change Scenarios: A Case Study in Central Vietnam. *Water Resources*, *47*(2), 189-199. https://doi.org/10.1134/s009780782002013x

Rivera, J. A., Marianetti, G., & Hinrichs, S. (2018). Validation of CHIRPS precipitation dataset along the Central Andes of Argentina. *Atmospheric Research*, *213*, 437-449. https://doi.org/10.1016/j.atmosres.2018.06.023

Sacré Regis M., D., Mouhamed, L., Kouakou, K., Adeline, B., Arona, D., Houebagnon Saint. J., C., Koffi Claude A., K., Talnan Jean H., C., Salomon, O., & Issiaka, S. (2020). Using the CHIRPS Dataset to Investigate Historical Changes in Precipitation Extremes in West Africa. *Climate*, *8*(7), 84. https://doi.org/10.3390/cli8070084

Sang, J., Hou, B., Wang, H., & Ding, X. (2023). Prediction of water resources change trend in the Three Gorges Reservoir Area under future climate change. *Journal of Hydrology*, *617*, 128881. https://doi.org/10.1016/j.jhydrol.2022.128881

Satgé, F., Defrance, D., Sultan, B., Bonnet, M. P., Seyler, F., Rouché, N., Pierron, F., & Paturel, J. E. (2020). Evaluation of 23 gridded precipitation datasets across West Africa. *Journal of Hydrology*, *581*, 124412. https://doi.org/10.1016/j.jhydrol.2019.124412

Schilling, J., Hertig, E., Tramblay, Y., & Scheffran, J. (2020). Climate change vulnerability, water resources and social implications in North Africa. *Regional Environmental Change*, *20*(1), 15. https://doi.org/10.1007/s10113-020-01597-7

Schwatke, C., Scherer, D., & Dettmering, D. (2019). Automated Extraction of Consistent Time-Variable Water Surfaces of Lakes and Reservoirs Based on Landsat and Sentinel-2. *Remote Sensing*, *11*(9), 1010. https://doi.org/10.3390/rs11091010

Sen, P. K. (1968). Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of the American Statistical Association*, *63*(324), 1379-1389. https://doi.org/10.1080/01621459.1968.10480934

Shen, C. (2018). A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. *Water Resources Research*, *54*(11), 8558-8593. https://doi.org/10.1029/2018wr022643

Shen, Z., Yong, B., Gourley, J. J., Qi, W., Lu, D., Liu, J., Ren, L., Hong, Y., & Zhang, J. (2020). Recent global performance of the Climate Hazards group Infrared Precipitation (CHIRP) with Stations (CHIRPS). *Journal of Hydrology*, *591*, 125284. https://doi.org/10.1016/j.jhydrol.2020.125284

Solgi, R., Loáiciga, H. A., & Kram, M. (2021). Long short-term memory neural network (LSTM-NN) for aquifer level time series forecasting using in-situ piezometric observations. *Journal of Hydrology*, *601*, 126800. https://doi.org/10.1016/j.jhydrol.2021.126800

Sorkhabi, O. M., Shadmanfar, B., & Kiani, E. (2022). Monitoring of dam reservoir storage with multiple satellite sensors and artificial intelligence. *Results in Engineering*, *16*, 100542. https://doi.org/10.1016/j.rineng.2022.100542

Stumpf, A., Michéa, D., & Malet, J. P. (2018). Improved Co-Registration of Sentinel-2 and Landsat-8 Imagery for Earth Surface Motion Measurements. *Remote Sensing*, *10*(2), 160. https://doi.org/10.3390/rs10020160

Taheri Dehkordi, A., Valadan Zoej, M. J., Ghasemi, H., Ghaderpour, E., & Hassan, Q. K. (2022). A New Clustering Method to Generate Training Samples for Supervised Monitoring of Long-Term Water Surface Dynamics Using Landsat Data through Google Earth Engine. *Sustainability*, *14*(13), 8046. https://doi.org/10.3390/su14138046

Thrasher, B., Wang, W., Michaelis, A., Melton, F., Lee, T., & Nemani, R. (2022). NASA Global Daily Downscaled Projections, CMIP6. *Scientific Data*, *9*(1), 262. https://doi.org/10.1038/s41597-022-01393-4

Tsolakidis, I., & Vafiadis, M. (2019). Comparison of Hydrographic Survey and Satellite Bathymetry in Monitoring Kerkini Reservoir Storage. *Environmental Processes*, *6*(4), 1031-1049. https://doi.org/10.1007/s40710-019-00394-7

Tyralis, H., Papacharalampous, G., & Langousis, A. (2019). A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources. *Water*, *11*(5), 910. https://doi.org/10.3390/w11050910

Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*, *185*, 46-56. https://doi.org/10.1016/j.rse.2016.04.008

Wang, C., Jia, M., Chen, N., & Wang, W. (2018). Long-Term Surface Water Dynamics Analysis Based on Landsat Imagery and the Google Earth Engine Platform: A Case Study in the Middle Yangtze River Basin. *Remote Sensing*, *10*(10), 1635. https://doi.org/10.3390/rs10101635

Wang, R., Xia, H., Qin, Y., Niu, W., Pan, L., Li, R., Zhao, X., Bian, X., & Fu, P. (2020). Dynamic Monitoring of Surface Water Area during 1989–2019 in the Hetao Plain Using Landsat Data in Google Earth Engine. *Water*, *12*(11), 3010. https://doi.org/10.3390/w12113010

Wang, Z., Xie, F., Ling, F., & Du, Y. (2022). Monitoring Surface Water Inundation of Poyang Lake and Dongting Lake in China Using Sentinel-1 SAR Images. *Remote Sensing*, *14*(14), 3473. https://doi.org/10.3390/rs14143473

Weng, P., Tian, Y., Liu, Y., & Zheng, Y. (2023). Time-series generative adversarial networks for flood forecasting. *Journal of Hydrology*, *622*, 129702. https://doi.org/10.1016/j.jhydrol.2023.129702

Wu, G., Chen, C., Liu, Y., Fan, X., Niu, H., & Liu, Y. (2023). Developing a High-Resolution Seamless Surface Water Extent Time-Series over Lake Victoria by Integrating MODIS and Landsat Data. *Remote Sensing*, *15*(14), 3500. https://doi.org/10.3390/rs15143500

Wulder, M. A., Roy, D. P., Radeloff, V. C., Loveland, T. R., Anderson, M. C., Johnson, D. M., Healey, S., Zhu, Z., Scambos, T. A., Pahlevan, N., Hansen, M., Gorelick, N., Crawford, C. J., Masek, J. G., Hermosilla, T., White, J. C., Belward, A. S., Schaaf, C., Woodcock, C. E., ..., & Huntington, J. L. (2022). Fifty years of Landsat science and impacts. *Remote Sensing of Environment*, *280*, 113195. https://doi.org/10.1016/j.rse.2022.113195

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

Yang, X., Chen, Y., & Wang, J. (2020). Combined use of Sentinel-2 and Landsat 8 to monitor water surface area dynamics using Google Earth Engine. *Remote Sensing Letters*, *11*(7), 687-696. https://doi.org/10.1080/2150704x.2020.1757780

Yao, F., Wang, J., Wang, C., & Crétaux, J. F. (2019). Constructing long-term high-frequency time series of global lake and reservoir areas using Landsat imagery. *Remote Sensing of Environment*, *232*, 111210. https://doi.org/10.1016/j.rse.2019.111210

Yao, J., Sun, S., Zhai, H., Feger, K. H., Zhang, L., Tang, X., Li, G., & Wang, Q. (2022). Dynamic monitoring of the largest reservoir in North China based on multi-source satellite remote sensing from 2013 to 2022: Water area, water level, water storage and water quality. *Ecological Indicators*, *144*, 109470. https://doi.org/10.1016/j.ecolind.2022.109470

Yu, X., Xu, Y. P., Gu, H., & Guo, Y. (2023). Multi-objective robust optimization of reservoir operation for real-time flood control under forecasting uncertainty. *Journal of Hydrology*, *620*, 129421. https://doi.org/10.1016/j.jhydrol.2023.129421

Yue, L., Li, B., Zhu, S., Yuan, Q., & Shen, H. (2023). A fully automatic and high-accuracy surface water mapping framework on Google Earth Engine using Landsat time-series. *International Journal of Digital Earth*, *16*(1), 210-233. https://doi.org/10.1080/17538947.2023.2166606

Zhang, J., Ma, S., & Song, Y. (2022). Hydrological and water quality simulation and future runoff prediction under CMIP6 scenario in the upstream basin of Miyun Reservoir. *Journal of Water and Climate Change*, *13*(7), 2505-2530. https://doi.org/10.2166/wcc.2022.389

Zhao, G., & Gao, H. (2018). Automatic Correction of Contaminated Images for Assessment of Reservoir Surface Area Dynamics. *Geophysical Research Letters*, *45*(12), 6092-6099. https://doi.org/10.1029/2018gl078343

Zhao, G., & Gao, H. (2019). Towards Global Hydrological Drought Monitoring Using Remotely Sensed Reservoir Surface Area. *Geophysical Research Letters*, *46*(22), 13027-13035. https://doi.org/10.1029/2019gl085345

Zhao, B., & Wang, L. (2024). Surface water monitoring from 1984 to 2021 based on Landsat time-series images and Google Earth Engine. *Heliyon*, *10*(17), e36660. https://doi.org/10.1016/j.heliyon.2024.e36660

Zou, Y., Wang, J., Lei, P., & Li, Y. (2023). A novel multi-step ahead forecasting model for flood based on time residual LSTM. *Journal of Hydrology*, *620*, 129521. https://doi.org/10.1016/j.jhydrol.2023.129521