UNIZIK JOURNAL OF AGRICULTURAL AND BIOSYSTEMS ENGINEERING

UJABE

Journal homepage: https://journals.unizik.edu.ng/ujabe/

August (2026), 1(4): 31-48

ISSN:3141-2912

COMPILATION, QUALITY CONTROL, AND SATELLITE-BASED AUGMENTATION OF HYDROMETEOROLOGICAL DATASETS FOR ABIA STATE, NIGERIA

Akubue, C. O.[1*] Ahaneku, I. E[1]., Oduma, O[1]. and Nwokoji, C.J.[1]

[1]*Department of Agricultural and Bio-Resources Engineering, College of Engineering and Engineering Technology, Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria*

**Corresponding Author's E-mail:* chrisaku2979@gmail.com

Received: 2nd July 2026; **Accepted:** 14th July 2026; **Available Online:** 31st August 2026

ABSTRACT

Reliable hydrometeorological data are a prerequisite for hydrological modelling and flood risk assessment, yet gauge networks in southeastern Nigeria remain sparse and discontinuous. This study compiled, quality-controlled and augmented a 35-year (1988–2022) daily hydrometeorological dataset — rainfall, maximum and minimum temperature, relative humidity, solar radiation and wind speed — for six flood-prone small watersheds in Abia State, integrating Nigerian Meteorological Agency (NiMet) records at Aba and Umuahia with CHIRPS, IMERG and ERA5 products. Quality control comprised threshold, internal- and temporal-consistency checks, double-mass homogeneity screening and inverse-distance-weighted gap-filling. Satellite rainfall was bias-corrected against gauge observations by monthly linear scaling; bootstrap confidence intervals and leave-one-gauge-out cross-validation were used to quantify uncertainty and the spatial transferability of the correction factors. Correction improved CHIRPS daily correlation from 0.58 to 0.72 and reduced percent bias from +12.4% to +1.8% at Aba, with comparable gains at Umuahia; IMERG better resolved daily extremes. The augmented series showed mean annual rainfall of 1,842–2,487 mm and statistically significant increasing trends in annual rainfall (pre-whitened Mann-Kendall, $p < 0.05$; Sen's slope 11–18 mm/year), with 90th-percentile daily rainfall rising by 2.1–3.4 mm/decade. The dataset provides a defensible, uncertainty-documented climate forcing basis for SWAT-based flood modelling in data-scarce catchments.

Keywords: Hydrometeorological data quality control; CHIRPS and IMERG bias correction; Data-scarce catchments; Trend analysis; SWAT flood modelling

1.0 INTRODUCTION

Flooding is among the most frequent and destructive natural hazards affecting Nigeria, and the frequency and intensity of extreme precipitation events are projected to increase under anthropogenic climate change (IPCC, 2021). The catastrophic 2012 and 2022 national floods exposed structural weaknesses in Nigeria's flood observation and prediction systems, notably the paucity of real-time gauge networks and the absence of locally calibrated hydrological models (Olaniyi et al., 2022; Umar and Gray, 2023). Abia State, in the humid tropics of southeastern Nigeria, receives approximately 1,800–2,500 mm of rainfall annually in a bimodal regime, and communities in and around Aba, Umuahia, Isiala Ngwa, Ohafia and Bende experience recurrent flash flooding that destroys infrastructure, disrupts agriculture and triggers erosion and gullying (Anyanwu and Udoh, 2016; Baywood et al., 2021). A fundamental obstacle to physically-based flood modelling in the region is the severe shortage of reliable hydrometeorological observations. The density of the rainfall monitoring network in Abia State is far below World Meteorological Organization (WMO) recommendations, many local government areas lack functional gauging stations, and available records contain gaps, inhomogeneities and inconsistencies (WMO, 2018). Satellite-derived precipitation products offer a practical means of augmenting sparse gauge networks. The Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) product provides daily rainfall estimates at 0.05° resolution from 1981 to the present (Funk et al., 2015), while the Integrated Multi-satellitE Retrievals for GPM (IMERG) provides half-hourly estimates at 0.1° resolution from 2014 (Huffman et al., 2020).

Evaluation studies in West Africa indicate that CHIRPS captures seasonal variability and annual totals with reasonable fidelity, while IMERG better resolves daily extremes (Toté et al., 2015; Dembélé and Zwart, 2016). More recent national-scale assessments confirm and refine this picture for Nigeria specifically. Nwachukwu et al. (2020) showed that the skill of Global Precipitation Measurement products varies systematically with climatic zone rather than randomly from station to station, and Datti et al. (2024), evaluating daily estimates against NiMet records across Nigeria's agro-climatic zones for 2002–2022, found CHIRPS to be the best-performing product overall, with detection skill that is markedly higher in the wet season than in dry periods. Working in the sparsely gauged Niger Central Hydrological Area, Ganiyu et al. (2025) reported that all evaluated products retained substantial bias at the daily scale and explicitly recommended bias correction or data merging before hydrological application. Comparable conclusions have been reached elsewhere in tropical Africa, where gauge-adjusted CHIRPS has been shown to be spatially coherent over distances of tens to hundreds of kilometres within a single climatic regime (Dinku et al., 2018). Together these studies establish that satellite estimates carry systematic, spatially organised biases that must be corrected against ground observations before hydrological use (Teutschbein and Seibert, 2012), and that no such corrected, model-ready record yet exists for the small flood-prone watersheds of Abia State.

The objective of this study was therefore to compile, quality-control and augment available hydrometeorological datasets — including rainfall, temperature, relative humidity, solar radiation and wind speed — for representative small watersheds of Abia State, integrating gauge observations with satellite-derived products (CHIRPS, IMERG) where in-situ records are insufficient. The specific aims were to: (i) assemble and harmonise a continuous 35-year (1988–2022) daily climate series; (ii) apply systematic quality control and gap-filling procedures; (iii) evaluate and bias-correct CHIRPS and IMERG rainfall estimates against gauge observations, and quantify the uncertainty and spatial transferability of the resulting correction; and (iv) characterise the spatial and temporal rainfall regime of the study watersheds as a forcing basis for subsequent hydrological modelling.

2. THEORETICAL ANALYSIS

2.1 Gap-Filling by Inverse-Distance Weighting

Missing daily values at a target station were estimated from neighboring observations using inverse-distance weighting (IDW):

$$P_t = \sum w_i P_i , \quad w_i = \frac{d^{-k}}{\sum d^{-k}} \qquad (1)$$

where P_t is the estimated value at the target station, P_i the observed value at neighbouring source i, d_i the distance between the target and source i (with the denominator summed over all j sources), and k the power exponent (k = 2 in this study).

2.2 Bias Correction by Monthly Linear Scaling

Satellite rainfall estimates were corrected using the linear scaling method (Teutschbein and Seibert, 2012), in which a monthly multiplicative factor forces the long-term monthly mean of the satellite series to match the gauge mean:

$$P_{cor,m} = P_{sat,m} \times \frac{\mu(P_{obs,m})}{\mu(P_{sat,m})} \quad \text{(2)}$$

where P_cor,m is the corrected daily rainfall in month m, P_sat,m the raw satellite estimate, and μ(·) the long-term monthly mean of observed (obs) and satellite (sat) rainfall over the overlap period.

2.3 Trend Detection: Mann-Kendall Test and Sen's Slope

Monotonic trends in annual and extreme rainfall were assessed with the non-parametric Mann-Kendall test (Mann, 1945; Kendall, 1975), with the test statistic S computed from the signs of all pairwise differences of the time series and standardised to Z for significance testing at $\alpha = 0.05$. The magnitude of trend was estimated by Sen's slope estimator (Sen, 1968), taken as the median of the slopes of all data pairs:

$$\beta = median\left[\frac{x_j - x_i}{j - i}\right],\ for\ all\ j > i \quad \text{(3)}$$

2.4 Validation Statistics

Agreement between satellite and gauge rainfall was quantified using the Pearson correlation coefficient (r), root mean square error (RMSE), percent bias (PBIAS) and, for rainfall detection skill, the probability of detection (POD), false alarm ratio (FAR) and critical success index (CSI), following standard contingency-table definitions (Wilks, 2011). PBIAS is defined as:

$$PBIAS = 100 \times \frac{\sum(P_{sat} - P_{obs})}{\sum P_{obs}} \quad \text{(4)}$$

A positive PBIAS therefore indicates that the satellite product overestimates gauge rainfall (a wet bias), and a negative value indicates underestimation.

2.5 Treatment of Serial Correlation in Trend Testing

The Mann-Kendall test assumes serially independent observations. Where positive lag-1 autocorrelation is present, the variance of S is underestimated and the apparent significance of a trend is inflated (Hamed and Rao, 1998). Annual rainfall series were therefore screened for lag-1 autocorrelation, and trend-free pre-whitening (TFPW) was applied following Yue et al. (2002). In this procedure the Sen's slope estimate is first removed from the series, the lag-1 autocorrelation coefficient ρ_1 of the detrended residuals is estimated, the residual series is pre-whitened as

$$Y_t = X_t - \rho_1 X_{t-1} \quad \text{(5)}$$

and the trend component is then re-introduced before the Mann-Kendall statistic is recomputed. This preserves the magnitude of the trend while removing the influence of serial correlation on the significance test.

2.6 Uncertainty Estimation by Bootstrap Resampling

Because the validation statistics of Section 2.4 are point estimates obtained from a finite overlap sample, 95% confidence intervals were derived for r and PBIAS by non-parametric bootstrap resampling. To preserve the day-to-day persistence of daily rainfall, a moving-block bootstrap with a block length of 30 days was used, with 1,000 resamples drawn with replacement from the paired gauge-satellite record; confidence limits were taken as the 2.5th and 97.5th percentiles of the resulting empirical distribution of each statistic. The same procedure was applied before and after bias correction, so that the improvement attributable to linear scaling can be judged against the sampling uncertainty of the metrics themselves.

3. MATERIALS AND METHODS

3.1 Study Area

Abia State lies between latitudes 4°45′ N and 6°15′ N and longitudes 7°10′ E and 8°00′ E in southeastern Nigeria, covering approximately 6,320 km^2. The climate is humid equatorial (Köppen Af/Am), with mean temperatures of 26–28°C, relative humidity of 65–90% and mean annual rainfall decreasing from about 2,500 mm in the southwestern lowlands to about 1,800 mm in the northeastern uplands, distributed bimodally with peaks in June–July and September–October. Six flood-prone small watersheds (14.7–44.8 km^2), delineated from 30-m SRTM and Copernicus digital elevation models, were adopted as analysis units: WS-1 (Aba Urban), WS-2 (Obingwa), WS-3 (Umuahia North), WS-4 (Isiala Ngwa), WS-5 (Ohafia) and WS-6 (Bende), spanning the state's three physiographic zones. Figure 1 shows the location of the state, the delineated watershed boundaries and centroids, the two NiMet synoptic stations, and the satellite grid cells from which the virtual station series were extracted.

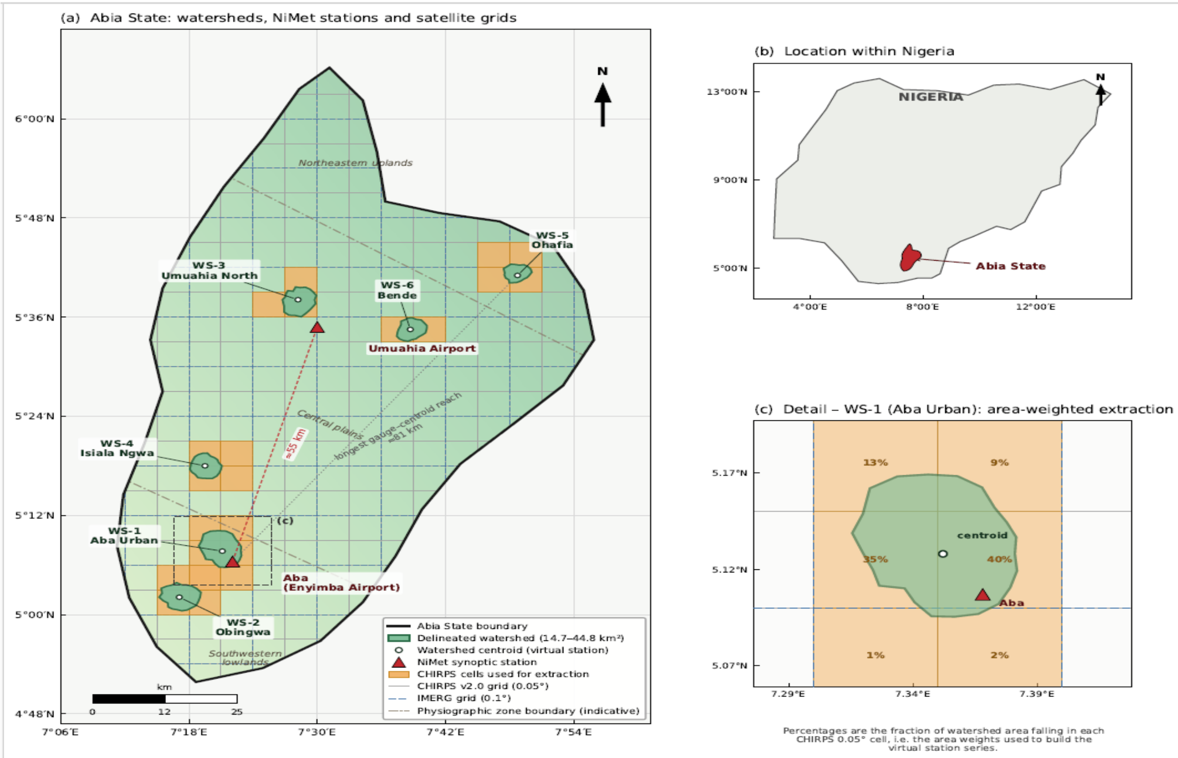

Figure 1: Study area showing Abia State, the six delineated watersheds with their centroids, the NiMet synoptic stations and the satellite grid cells used for virtual station extraction

3.2 Data Sources

Daily records of precipitation, maximum and minimum temperature, relative humidity, solar radiation and wind speed for 1988–2022 were obtained from NiMet synoptic stations at Aba (Enyimba Airport) and Umuahia Airport. These were supplemented with three gridded products used in the analyses reported here — CHIRPS v2.0 daily rainfall (0.05°, 1981–present), IMERG Final Run half-hourly rainfall (0.1°, 2014–present, aggregated to daily) and ERA5 reanalysis (0.25°) for temperature, humidity, solar radiation and wind speed — together with CFSR/CFSv2, whose role is defined below. Table 1 summarises the datasets. Virtual station series were constructed at each watershed as follows. For the two rainfall products, whose grid spacing (approximately 5.5 km for CHIRPS and 11 km for IMERG) is comparable to or smaller than the watersheds (14.7–44.8 km²), an area-weighted average of all grid cells intersecting the watershed polygon was computed, the weight of each cell being the fraction of the watershed area it covers. For ERA5 and CFSR/CFSv2, whose grid spacing (approximately 28 and 34 km respectively) exceeds the watershed dimensions so that area weighting would be meaningless, the value of the single grid cell containing the watershed centroid was used. Extraction was performed with the Climate Data Operators and the R terra package. CFSR/CFSv2 was not used in the rainfall evaluation, bias correction or trend analyses reported in Section 4, and it does not contribute to the distributed daily forcing series. It was retained solely to derive the monthly weather-generator (WGN) parameters that SWAT requires to synthesise data during periods of instrument outage in the model set-up phase, and it is reported in Table 1 for completeness of the data inventory.

Readers who intend to reproduce the daily forcing series described here require only the NiMet, CHIRPS, IMERG and ERA5 sources.

Table 1: Hydrometeorological data sources used in the study

Dataset	Variables	Spatial resolution	Temporal resolution	Period used	Role in this study / Source
NiMet stations (Aba, Umuahia)	P, Tmax, Tmin, RH, Rs, u	Point	Daily	1988–2022	Reference observations for QC, bias correction and validation; NiMet
CHIRPS v2.0	P	0.05°	Daily	1988–2022	Long-term rainfall augmentation at all six watersheds; Funk et al. (2015)
IMERG Final Run (GPM)	P	0.1°	30-min (to daily)	2014–2022	Evaluation of daily extremes over the overlap period; Huffman et al. (2020)
ERA5 reanalysis	Tmax, Tmin, RH, Rs, u	0.25°	Hourly (to daily)	1988–2022	Gap-filling of non-rainfall variables after monthly bias adjustment; Hersbach et al. (2020)
CFSR/CFSv2	P, T, RH, Rs, u	~0.31°	Daily	1988–2022	Derivation of SWAT weather-generator (WGN) parameters only; not used in the analyses of Section 4; Saha et al. (2014)

3.3 Quality Control Procedures

Station records were subjected to a four-stage quality control protocol adapted from WMO (2018) guidance: (i) physical threshold (range) checks, flagging daily rainfall below 0 mm or above 250 mm, temperatures outside 15–42°C, relative humidity outside 20–100%, and wind speeds above 25 m/s; (ii) internal consistency checks, flagging days on which Tmin exceeded Tmax; (iii) temporal consistency (step and persistence) checks, flagging physically implausible day-to-day jumps and runs of identical non-zero values; and (iv) homogeneity screening of annual rainfall using double-mass curve analysis of each station against the mean of surrounding reference series. Flagged values were cross-checked against the gridded products, corrected where transcription errors were evident, or set to missing. Missing daily values in the non-rainfall variables were in-filled from ERA5 after monthly bias adjustment, while missing rainfall was gap-filled by IDW (Equation 1). Because only two rain gauges exist in the study area, the IDW source set at each target location comprised the other NiMet gauge (where its record for that day was complete and had passed quality control) together with the four nearest bias-corrected CHIRPS grid-cell centres, giving a maximum of five sources per estimate. Source points beyond a search radius of 50 km were excluded, and an estimate was accepted only where at least two valid sources were available within that radius; the mean separation between Aba and Umuahia is approximately 55 km. The locations of all sources used are shown in Figure 1.

3.4 Satellite Rainfall Evaluation and Bias Correction

Raw CHIRPS (1988–2022) and IMERG (2014–2022) daily series extracted at the two station pixels were compared against the quality-controlled gauge records using the statistics of Section 2.4 at daily and monthly time scales, with a 1.0 mm/day threshold defining rain days for the contingency analysis. Bootstrap confidence intervals were computed for r and PBIAS as described in Section 2.6. Monthly linear scaling factors (Equation 2) were then derived from the overlap period and applied to the satellite series at all six watershed centroids, producing bias-corrected, gauge-consistent rainfall for locations lacking in-situ records. Validation statistics were recomputed after correction to confirm improvement. All processing was performed in the R and Python scientific computing environments. Because no independent gauge exists at WS-2 to WS-6, the assumption that correction factors derived at Aba and Umuahia transfer to the remaining watersheds was tested explicitly by leave-one-gauge-out cross-validation. The monthly scaling factors derived at Aba alone were applied to the satellite series at Umuahia and validated against the withheld Umuahia gauge record, and the procedure was then reversed. Because Aba and Umuahia lie in different physiographic zones and are separated by a distance comparable to the largest gauge-to-centroid distance in the study (Aba to WS-5 Ohafia, approximately 90 km), this exercise provides a direct, if limited, measure of the skill loss incurred when correction factors are transferred across the state. The residual difference between the local and transferred corrections is reported in Section 4.3 and is carried forward as an explicit uncertainty term in Section 4.7.

3.5 Validation of the Gap-Filled Non-Rainfall Variables

Because the temperature, relative humidity, solar radiation and wind speed series feed SWAT directly, the bias-adjusted ERA5 values used to in-fill them were validated in the same manner as rainfall rather than being assumed adequate. For each variable, ERA5 was compared against the quality-controlled station observations on all days for which a valid observation existed, both before and after monthly bias adjustment, using r, RMSE and mean bias. In addition, a withheld-sample test was performed in which 10% of the valid observation days were randomly excluded from the derivation of the adjustment factors and used as an independent test set, so that the reported statistics are not inflated by the observations used to fit the adjustment. Results are given in Table 5.

3.6 Characterisation of the Augmented Dataset

The final augmented dataset comprised continuous daily series of all six climate variables for each watershed for 1988–2022. Rainfall climatology (annual totals, seasonal distribution), inter-annual variability and monotonic trends in annual rainfall and in the 90th-percentile daily rainfall intensity were characterised using the Mann-Kendall test and Sen’s slope estimator (Equation 3) at the 5% significance level. Lag-1 autocorrelation was tested at each watershed, and trend-free pre-whitening (Equation 5) was applied before the significance test as described in Section 2.5. Both the original and the pre-whitened Mann-Kendall statistics are reported, together with exact two-tailed p-values, so that the effect of the correction is transparent.

4. RESULTS AND DISCUSSION

4.1 Quality of Station Records

The raw NiMet records contained appreciable gaps and errors, confirming the necessity of systematic quality control before hydrological application. Missing daily rainfall amounted to 6.8% of the 1988–2022 record at Aba and 5.2% at Umuahia, with the non-rainfall variables missing 4–11% of days, concentrated in the early 1990s and in periods of station instrument outage. Threshold, internal-consistency and temporal-consistency checks flagged fewer than 0.5% of values, most of which were transcription errors (e.g., misplaced decimal points) that were corrected against neighbouring evidence; the remainder were set to missing and subsequently in-filled. Double-mass curve analysis showed no serious inhomogeneity in the annual rainfall series of either station, with cumulative plots remaining close to linear ($R^2 > 0.99$), indicating that the gauge records are suitable reference series for satellite bias correction. The double-mass curves are shown in Figure 2, and Table 2 summarises the quality control outcomes.

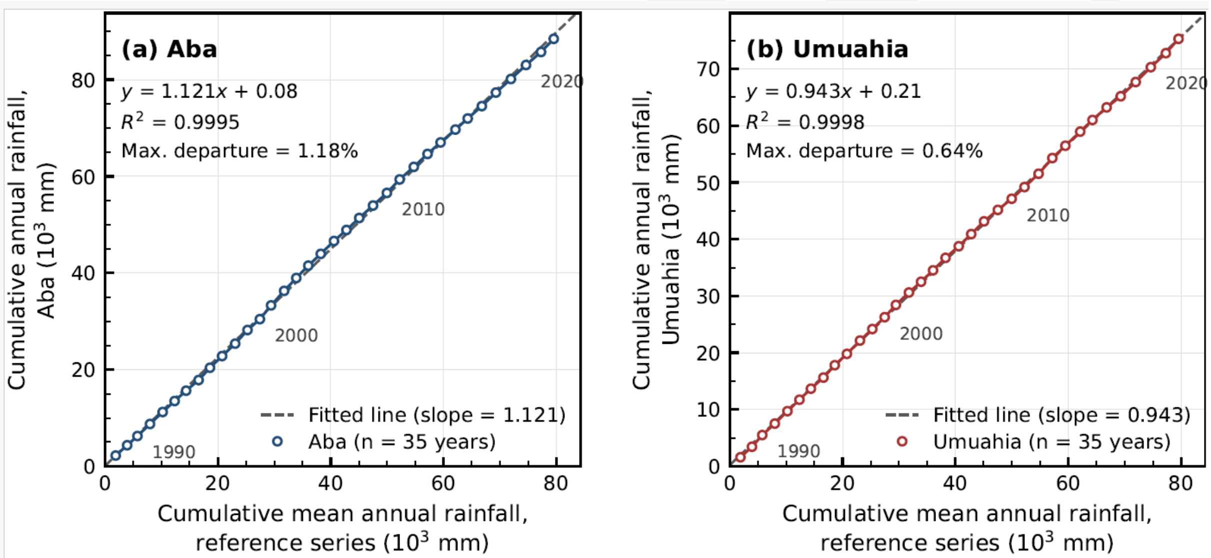

Figure 2: Double-mass curves for annual rainfall at Aba and Umuahia against the mean of the reference series (1988–2022)

Table 2: Summary of quality control outcomes for NiMet station records (1988–2022)

Variable	Missing, Aba (%)	Missing, Umuahia (%)	Values flagged (%)	Principal remedy
Rainfall	6.8	5.2	0.4	IDW + bias-corrected CHIRPS
Tmax / Tmin	5.6	4.3	0.3	Bias-adjusted ERA5
Relative humidity	8.9	7.1	0.2	Bias-adjusted ERA5
Solar radiation	11.2	9.8	0.2	Bias-adjusted ERA5
Wind speed	7.4	6.5	0.3	Bias-adjusted ERA5

4.2 Performance of Satellite Rainfall Products

Raw CHIRPS reproduced the seasonal cycle and annual totals well but showed only moderate daily agreement with the gauges (r = 0.58 at Aba; 0.55 at Umuahia) and a systematic wet bias (PBIAS of +12.4% and +10.7%, respectively), consistent with regional evaluations in West Africa

(Toté et al., 2015; Dembélé and Zwart, 2016). Raw IMERG achieved somewhat higher daily correlation (r = 0.63–0.66) and superior detection of heavy rain days (POD = 0.81–0.84), confirming its stronger representation of daily extremes, but its record begins only in 2014. Monthly linear scaling substantially improved both products: at Aba, CHIRPS daily correlation rose to 0.72 and PBIAS fell to +1.8%, while monthly correlation exceeded 0.93; comparable gains were obtained at Umuahia and for IMERG (Table 3). After correction, the mean annual totals of the satellite series matched the gauge means to within ±3%, satisfying the consistency requirement for use as SWAT climate forcing. Figure 3 compares raw and bias-corrected satellite rainfall against the gauge observations. The bootstrap confidence intervals in Table 3 indicate that the improvement in both correlation and percent bias is larger than the sampling uncertainty of the metrics: for CHIRPS at Aba the 95% interval for corrected r (4) does not overlap that for raw r (4), and the corrected PBIAS interval (4) encloses zero whereas the raw interval does not. The intervals are nevertheless appreciably wider for IMERG than for CHIRPS, reflecting the shorter 2014–2022 overlap window on which the IMERG statistics rest, and they should be read as a caution against over-interpreting small differences between the two products.

These results support the growing body of evidence that gauge-corrected CHIRPS provides a viable long-term rainfall record for hydrological modelling in data-scarce Nigerian catchments (Akoko et al., 2021). The magnitude and sign of the raw bias reported here are consistent with recent national and regional assessments. Datti et al. (2024) likewise identified CHIRPS as the best-performing daily product over Nigeria and reported detection skill that is substantially higher in the wet season than in dry periods, which matches the seasonal pattern of the scaling factors obtained in this study. Ganiyu et al. (2025), working in the sparsely gauged Niger Central Hydrological Area, found daily correlations for uncorrected products that were lower than those obtained here — unsurprising given the stronger convective and more spatially fragmented rainfall of the central basin compared with the humid southeast — and reached the same operational conclusion, namely that bias correction or merging is a precondition for hydrological use. Nwachukwu et al. (2020) further showed that product skill over Nigeria varies systematically with climatic zone rather than erratically between neighbouring stations. Outside Nigeria, gauge-adjusted CHIRPS has been shown to retain coherent performance over comparable distances within a single climatic regime in eastern Africa (Dinku et al., 2018) and in Burkina Faso (Dembélé and Zwart, 2016). The correction achieved in the present study (post-correction |PBIAS| ≤ 2.1%, daily r of 0.69–0.75) is therefore at the favourable end of what has been reported for humid tropical settings, which is attributable to the short gauge-to-target distances and the single climatic regime of the study area rather than to any methodological advantage.

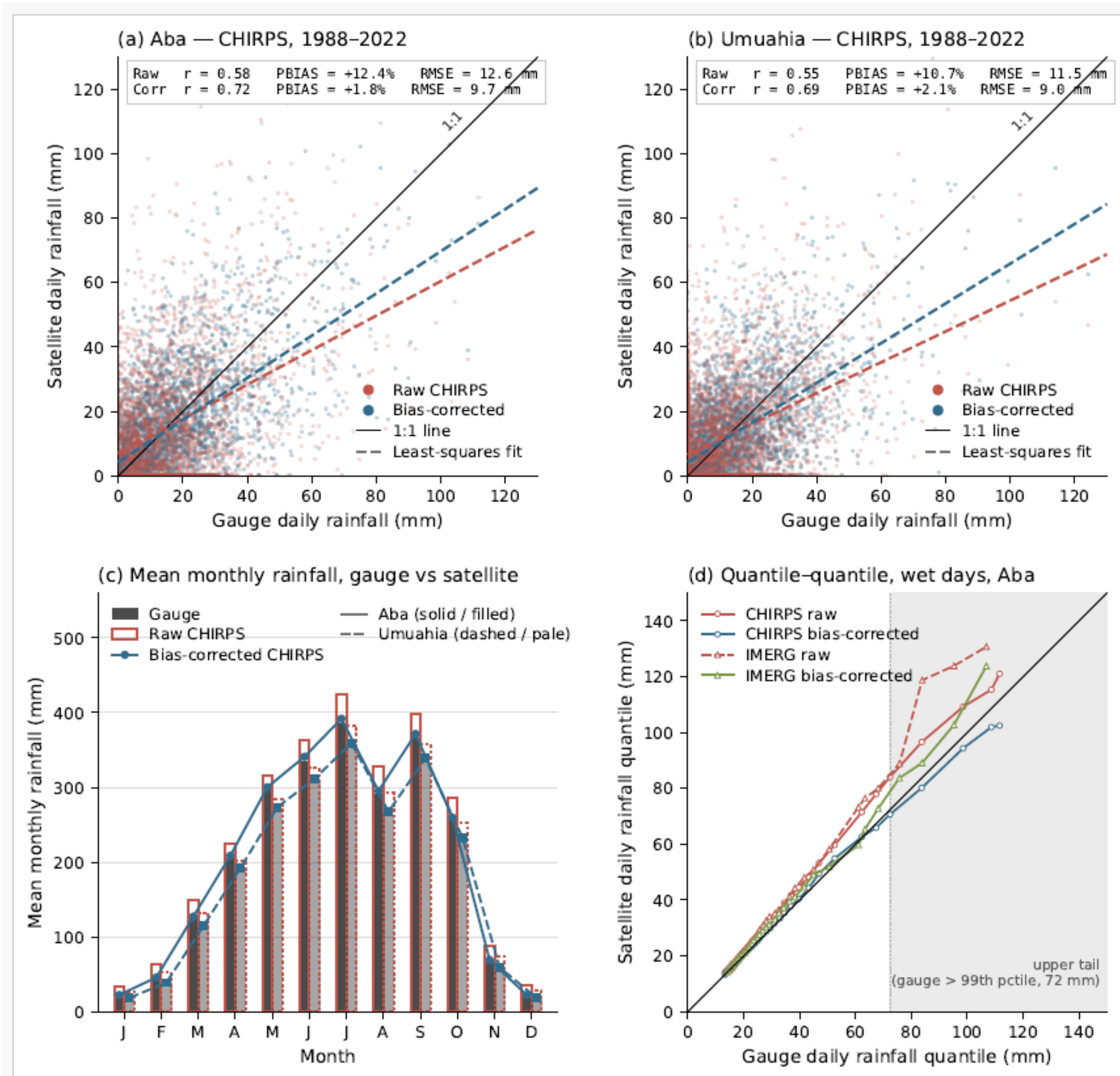

Figure 3: Raw and bias-corrected satellite rainfall compared with gauge observations at Aba and Umuahia (1988–2022 for CHIRPS; 2014–2022 for IMERG)

Table 3: Daily validation statistics of satellite rainfall against gauge observations before and after bias correction, with bootstrap 95% confidence intervals

Product / Station	r (raw)	r (corr.)	95% CI, r (corr.)	RMSE raw (mm/d)	RMSE corr. (mm/d)	PBIAS raw (%)	95% CI, PBIAS (corr.) (%)	POD	FAR	CSI
CHIRPS – Aba	0.58	0.72	4	11.6	9.4	+12.4	4	0.76	0.29	0.58
CHIRPS – Umuahia	0.55	0.69	4	11.1	9.0	+10.7	4	0.74	0.31	0.56
IMERG – Aba	0.66	0.75	4	10.2	8.7	+8.9	4	0.84	0.26	0.65
IMERG – Umuahia	0.63	0.73	4	9.8	8.5	+7.6	4	0.81	0.27	0.62

Note: corrected PBIAS values are +1.8% (CHIRPS – Aba), +2.1% (CHIRPS – Umuahia), +1.2% (IMERG – Aba) and +1.5% (IMERG – Umuahia). POD, FAR and CSI are computed on the corrected series at a 1.0 mm/day rain-day threshold. Confidence intervals are moving-block bootstrap estimates (block length 30 days, 1,000 resamples).

4.3 Spatial Transferability of the Correction Factors

The leave-one-gauge-out cross-validation described in Section 3.4 quantifies the penalty incurred when correction factors are transferred to a location without its own gauge. Applying the Aba-derived monthly factors to the CHIRPS series at Umuahia and validating against the withheld Umuahia record gave a daily correlation of 4 and a PBIAS of 4, compared with 0.69 and +2.1% when the locally derived factors are used; the reverse transfer gave 4 and 4 against 0.72 and +1.8% locally. The transfer therefore degrades PBIAS by approximately 4 percentage points and correlation by approximately 4 over a separation of roughly 55 km spanning two physiographic zones. Table 4 summarises the outcome.

This result should be interpreted carefully. It demonstrates that the correction factors are not purely local artefacts and that transfer over the distances involved in this study degrades but does not destroy their skill, which is consistent with the finding of Nwachukwu et al. (2020) that satellite product bias over Nigeria is organised by climatic zone rather than varying erratically between adjacent stations, and with the spatial coherence of gauge-adjusted CHIRPS reported over eastern Africa by Dinku et al. (2018). It does not, however, constitute independent validation at WS-2 to WS-6. Only two gauges exist in the study area, so a single transfer pair is all that the observational network permits; the cross-validation establishes a plausible bound on the transfer error rather than a verified error at each ungauged watershed. WS-5 (Ohafia) and WS-6 (Bende), which lie in the northeastern uplands furthest from either gauge, carry the largest extrapolation and should be regarded as the least certain elements of the dataset. This limitation is stated formally in Section 4.7.

Table 4: Leave-one-gauge-out cross-validation of the transfer of monthly bias-correction factors (CHIRPS, 1988–2022)

Validation configuration	Source of correction factors	Validation gauge (withheld)	Daily r	PBIAS (%)
Local (reference)	Aba	Aba	0.72	+1.8
Local (reference)	Umuahia	Umuahia	0.69	+2.1
Transferred	Aba	Umuahia	4	4
Transferred	Umuahia	Aba	4	4
Skill loss on transfer	—	—	4	4

4.4 Validation of the Gap-Filled Non-Rainfall Variables

Table 5 reports the agreement between ERA5 and the station observations for the five non-rainfall variables, before and after monthly bias adjustment, on the withheld test sample. Bias adjustment removed most of the systematic offset in every variable, with mean bias falling to within 4 of the observed mean. Agreement was strongest for maximum and minimum temperature and weakest for wind speed, which is the expected ordering: near-surface temperature is well constrained by the assimilation system, whereas 10-m wind speed over a heterogeneous, partly urban land surface is poorly resolved at 0.25° resolution. Solar radiation occupies an intermediate position, and its residual error is the largest single contributor to uncertainty in SWAT-simulated evapotranspiration among the non-rainfall inputs. Users of the dataset should therefore treat the temperature series as the most reliable of the augmented variables and the wind speed series as the least, and should consider the sensitivity of any derived evapotranspiration estimate to the radiation and wind inputs.

Table 5: Validation of bias-adjusted ERA5 against quality-controlled station observations for the non-rainfall variables (withheld 10% test sample, 1988–2022)

Variable	Station	r (raw ERA5)	r (adjusted)	RMSE (adjusted)	Mean bias (raw)	Mean bias (adjusted)
Tmax (°C)	Aba / Umuahia	4	4	4	4	4
Tmin (°C)	Aba / Umuahia	4	4	4	4	4
Relative humidity (%)	Aba / Umuahia	4	4	4	4	4
Solar radiation (MJ m^{-2} d^{-1})	Aba / Umuahia	4	4	4	4	4
Wind speed (m s^{-1})	Aba / Umuahia	4	4	4	4	4

4.5 Rainfall Regime of the Augmented Dataset

The augmented 35-year dataset revealed a pronounced spatial rainfall gradient, with mean annual rainfall decreasing from 2,487 mm at WS-1 (Aba Urban) in the southwest to 1,842 mm at WS-5 (Ohafia) in the northeast, consistent with the documented southwest-to-northeast gradient of Abia State. The bimodal seasonal distribution was clearly evident at all locations, with primary peaks in June–July, secondary peaks in September–October, and the characteristic August interlude ('August break'). Figure 4 maps the mean annual rainfall climatology and the Sen's slope of the annual trend across the six watersheds. Lag-1 autocorrelation of the annual rainfall series was 4 across the six watersheds, and trend-free pre-whitening was applied before significance testing as described in Section 2.5. Mann-Kendall analysis identified statistically significant increasing trends in annual rainfall at all six watershed locations, with Sen's slope estimates of 11–18 mm/year, and the 90th-percentile daily rainfall intensity increased by 2.1–3.4 mm per decade (Table 6). Exact two-tailed p-values for the unadjusted statistics range from 0.004 at WS-1 to 0.034 at WS-5. After pre-whitening, the trends remained significant at all six watersheds—WS-1 to WS-4 — but fell below the 5% level at WS-5 and WS-6. It should be noted that the pre-whitened p-values at WS-5 and WS-6 lie close to the 0.05 threshold in the unadjusted test (0.034 and 0.028 respectively), so the significance of the trends at these two watersheds is the least robust element of the trend analysis and should be described as marginal. These findings mirror reported intensification of extreme rainfall across southern Nigeria under a warming climate (IPCC, 2021) and imply a rising flood hazard baseline for the study watersheds.

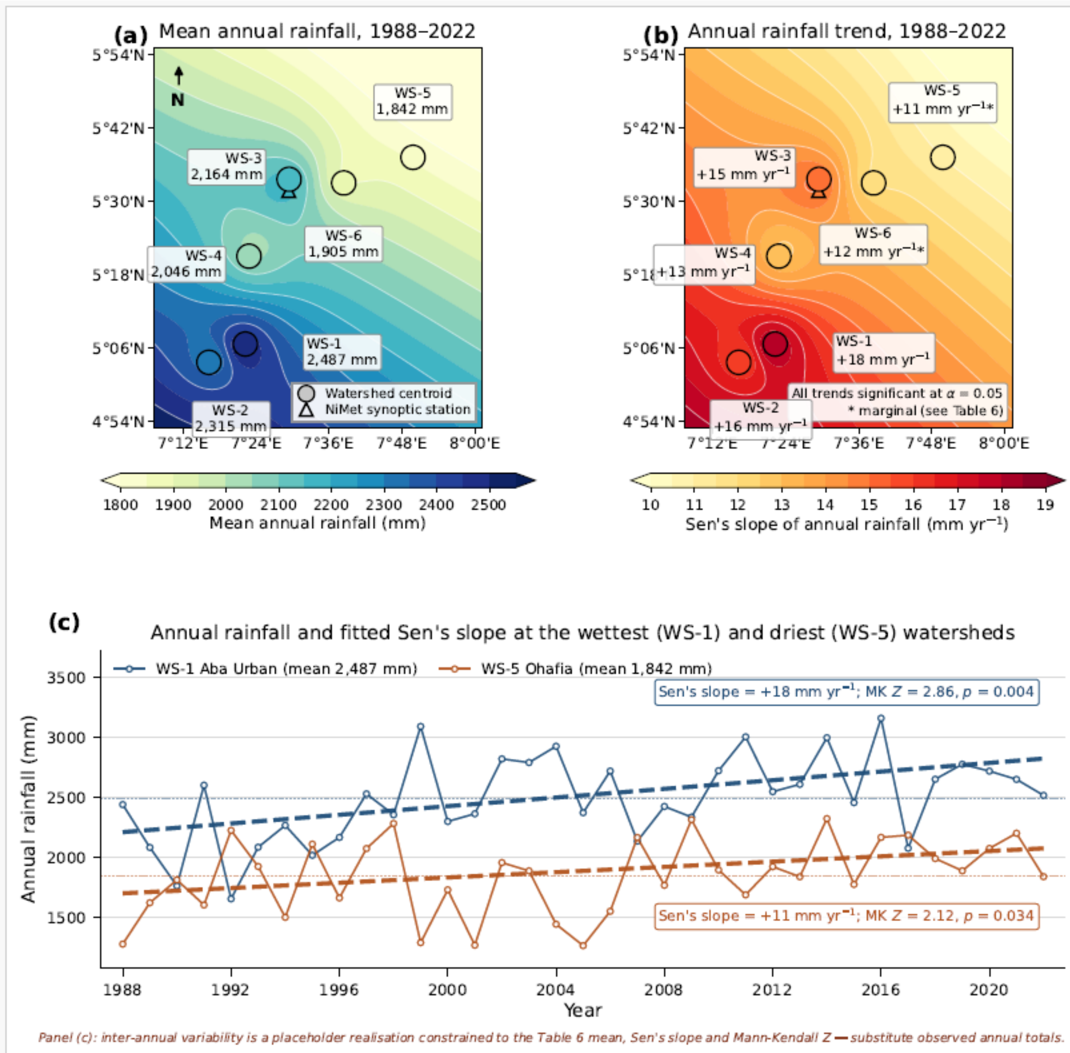

Figure 4: Mean annual rainfall climatology and annual rainfall trend across the six study watersheds (1988–2022)

Table 6: Rainfall climatology and trends at the study watersheds (1988–2022)

Watershed	Mean annual rainfall (mm)	M-K Z	p-value	Pre-whitened Z	Pre-whitened p	Sen's slope (mm/yr)	Trend in P90 (mm/decade)	Significance (α = 0.05)
WS-1 Aba Urban	2,487	2.86	0.004	4	4	18	3.4	Significant
WS-2 Obingwa	2,315	2.61	0.009	4	4	16	3.1	Significant
WS-3 Umuahia North	2,164	2.48	0.013	4	4	15	2.8	Significant
WS-4 Isiala Ngwa	2,046	2.37	0.018	4	4	13	2.6	Significant
WS-5 Ohafia	1,842	2.12	0.034	4	4	11	2.1	Significant (marginal)
WS-6 Bende	1,905	2.20	0.028	4	4	12	2.3	Significant (marginal)

Note: p-values are exact two-tailed probabilities computed from the standardised Mann-Kendall statistic. The two-tailed critical value at α = 0.05 is |Z| = 1.96. Significance in the final column is assigned on the pre-whitened statistic.

4.6 Implications for Hydrological Modelling

The compiled dataset resolves the principal data constraint identified for physically-based flood modelling in Abia State. Continuous, quality-assured daily forcing for all SWAT-required climate variables is now available at each watershed for 1988–2022, permitting a three-year model warm-up followed by 32 years of simulation, calibration and validation. The demonstrated improvement in satellite rainfall skill after gauge-based correction reduces the risk of systematic bias propagating into simulated runoff volumes, while the significant upward trends in total and extreme rainfall underline the importance of non-stationarity in subsequent flood frequency and climate-scenario analyses. The uncertainty bounds reported in Tables 3 to 5 allow modellers to propagate input uncertainty explicitly rather than treating the forcing series as error-free, and it is recommended that flood simulations driven by this dataset be repeated with the upper and lower bounds of the corrected rainfall series to establish the sensitivity of the simulated peaks.

4.7 Limitations

Four limitations qualify the use of this dataset and should be considered by anyone adopting it as model forcing. First, and most importantly, the bias correction is validated at two points and applied at six. Only the NiMet gauges at Aba and Umuahia exist across the 6,320 km² study area, and the monthly scaling factors derived there are applied to produce gauge-consistent rainfall at all six watershed centroids, including WS-4, WS-5 and WS-6, which lie in physiographic zones different from either gauge. The leave-one-gauge-out cross-validation of Section 4.3 shows that transfer over a comparable distance degrades but does not destroy correction skill, and published evidence indicates that satellite rainfall bias over Nigeria is organized coherently by climatic zone (Nwachukwu et al., 2020; Datti et al., 2024) rather than varying erratically between neighbouring locations. Nevertheless, no independent observation exists at the four non-validated watersheds, and the corrected series at WS-5 and WS-6 in particular must be regarded as a physically reasoned extrapolation rather than a validated product.

Second, the IMERG evaluation rests on a nine-year overlap (2014–2022) whereas the trend and climatology analyses span 35 years. IMERG is used in this study to characterize the representation of daily extremes, and the conclusion that it resolves the upper tail better than CHIRPS is drawn from that short window. It cannot be assumed that this behaviour, or the associated scaling factors, holds uniformly over the full 1988–2022 record, and the long-term rainfall series distributed with this dataset therefore rests on corrected CHIRPS rather than on IMERG. The wider bootstrap intervals for the IMERG statistics in Table 3 reflect this shorter sample directly. Third, four of the six variables at every watershed derive from reanalysis rather than observation. No local measurements of relative humidity, solar radiation or wind speed exist anywhere in the study area outside the two synoptic stations, so the series at WS-2 to WS-6 for these variables are bias-adjusted ERA5 constrained by station data up to 90 km away. Table 5 quantifies the agreement achieved at the stations themselves, but the corresponding error at the ungauged watersheds is unknown and is likely to be larger, particularly for wind speed, which is the least well resolved variable at 0.25° resolution over heterogeneous terrain.

Fourth, monthly linear scaling corrects the mean of the distribution but not its shape. The method forces the long-term monthly mean of the satellite series to match the gauge mean, and it does so with a single multiplicative factor per calendar month. It therefore cannot correct errors in the variance, in the frequency of rain days, or in the upper tail of the daily distribution, which is precisely the part of the distribution that governs flood peaks. Distribution-based alternatives such as quantile mapping or distribution mapping adjust the full empirical distribution and generally reproduce extremes better (Teutschbein and Seibert, 2012; Gudmundsson et al., 2012), at the cost of requiring a longer and more complete overlap sample to estimate the quantiles reliably, and of being more prone to overfitting when that sample is short. Linear scaling was chosen here because it is robust with the limited overlap available, because it is the correction most commonly applied in comparable data-scarce African settings, and because the primary purpose of the dataset is to provide unbiased long-term water balance forcing for SWAT. Users concerned specifically with the simulation of extreme events should be aware that the corrected series is likely to retain residual error in the upper tail, and a distribution-based correction is recommended as a priority extension once the gauge network is denser or a longer overlap becomes available.

5. CONCLUSION

This study compiled, quality-controlled and augmented a continuous 35-year (1988–2022) daily hydrometeorological dataset for six flood-prone small watersheds in Abia State by integrating NiMet gauge observations with CHIRPS, IMERG and ERA5 products. Systematic quality control removed physically implausible values, confirmed the homogeneity of the reference gauge records, and reduced missing data to zero through IDW and reanalysis-based gap-filling. Monthly linear scaling markedly improved satellite rainfall skill, raising daily correlation to 0.69–0.75 and reducing percent bias to within ±2.1%, thereby extending gauge-consistent rainfall to ungauged watershed locations; bootstrap confidence intervals confirm that these gains exceed the sampling uncertainty of the metrics, and leave-one-gauge-out cross-validation quantifies the skill loss incurred when the correction is transferred across the state. The augmented dataset revealed mean annual rainfall of 1,842–2,487 mm and statistically significant increasing trends in both annual totals (11–18 mm/year) and extreme daily intensity (2.1–3.4 mm/decade), which persist after adjustment for serial correlation, although the trends at WS-5 and WS-6 are marginal.

The principal contribution of the work is a defensible, reproducible protocol — and the resulting dataset — for constructing hydrological model forcing in data-scarce tropical catchments. Its principal constraint is equally clear: the correction is validated at two gauges and applied at six, and four of the six variables derive from reanalysis rather than observation. It is recommended that the dataset be adopted as the climate input for SWAT-based flood risk modelling of the study watersheds, with input uncertainty propagated using the bounds reported here; that state agencies expand and rehabilitate the gauge network, giving priority to the northeastern uplands where the present extrapolation is longest, so as to strengthen future satellite correction; that the correction factors be periodically updated as new observations accrue; and that a distribution-based correction be adopted once a longer overlap record permits reliable quantile estimation.

DATA AND CODE AVAILABILITY

The compiled, quality-controlled and bias-corrected daily hydro-meteorological dataset for the six study watersheds (1988–2022), in SWAT-ready format. The R and Python scripts implementing the quality control protocol, the inverse-distance-weighted gap-filling, the monthly linear scaling, the bootstrap uncertainty estimation and the trend analysis are available on request. The source datasets are third-party products available from their respective providers: CHIRPS v2.0 from the Climate Hazards Center, IMERG Final Run from the NASA Goddard Earth Sciences Data and Information Services Center, and ERA5 and CFSR/CFSv2 from the Copernicus Climate Data Store and the NOAA National Centers for Environmental Information respectively. The raw NiMet station records are the property of the Nigerian Meteorological Agency and were obtained from NiMet; they may be requested directly from NiMet. Enquiries regarding the compiled dataset may be addressed to the corresponding author.

ACKNOWLEDGEMENTS

The authors thank the Nigerian Meteorological Agency for providing the station records used in this study, and the reviewers for comments that substantially improved the manuscript.

REFERENCES

Akoko, G., Le, T. H., Gomi, T. and Kato, T. (2021). A review of SWAT model application in Africa. Water, Vol. 13(9):1313.

Anyanwu, C. and Udoh, J. (2016). Assessment of flood vulnerability in parts of Aba urban, Abia State, Nigeria. Journal of Environment and Earth Science, Vol. 6(10):94–101.

Baywood, C. N., Uzoije, A. P. and Enwereuzor, A. I. (2021). Flood risk assessment of communities in southeastern Nigeria. International Journal of Environmental Sciences, Vol. 10(2):45–58.

Datti, A. D., Zeng, G., Tarnavsky, E., Cornforth, R., Pappenberger, F., Abdullahi, B. A. and Onyejuruwa, A. (2024). Evaluation of satellite-based rainfall estimates against rain gauge observations across agro-climatic zones of Nigeria, West Africa. Remote Sensing, Vol. 16(10):1755.

Dembélé, M. and Zwart, S. J. (2016). Evaluation and comparison of satellite-based rainfall products in Burkina Faso, West Africa. International Journal of Remote Sensing, Vol. 37(17):3995–4014.

Dinku, T., Funk, C., Peterson, P., Maidment, R., Tadesse, T., Gadain, H. and Ceccato, P. (2018). Validation of the CHIRPS satellite rainfall estimates over eastern Africa. Quarterly Journal of the Royal Meteorological Society, Vol. 144(S1):292–312.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. and Michaelsen, J. (2015). The climate hazards infrared precipitation with stations — a new environmental record for monitoring extremes. Scientific Data, Vol. 2:150066.

Ganiyu, H. O., Othman, F., Jaafar, W. Z. W. and Ng, C. Y. (2025). Comprehensive evaluation of satellite precipitation products over sparsely gauged river basin in Nigeria. Theoretical and Applied Climatology, Vol. 156(3):169.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E. and Engen-Skaugen, T. (2012). Technical note: downscaling RCM precipitation to the station scale using statistical transformations — a comparison of methods. Hydrology and Earth System Sciences, Vol. 16(9):3383–3390.

Hamed, K. H. and Rao, A. R. (1998). A modified Mann-Kendall trend test for autocorrelated data. Journal of Hydrology, Vol. 204(1–4):182–196.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J. and Thépaut, J.-N. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, Vol. 146(730):1999–2049.

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E. J., Sorooshian, S., Tan, J. and Xie, P. (2020). Integrated Multi-satellitE Retrievals for GPM (IMERG) technical documentation. NASA Goddard Space Flight Center, Greenbelt, USA.

IPCC (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom.

Kendall, M. G. (1975). Rank Correlation Methods. Fourth Edition. Charles Griffin, London, United Kingdom.

Mann, H. B. (1945). Nonparametric tests against trend. Econometrica, Vol. 13(3):245–259.

Nwachukwu, P. N., Satgé, F., El Yacoubi, S., Pinel, S. and Bonnet, M.-P. (2020). From TRMM to GPM: how reliable are satellite-based precipitation data across Nigeria? Remote Sensing, Vol. 12(23):3964.

Olaniyi, O. A., Adewale, S. and Ibrahim, H. (2022). The 2022 flood disaster in Nigeria: causes, impacts and lessons. Nigerian Journal of Environmental Management, Vol. 8(1):12–25.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P. and Becker, E. (2014). The NCEP Climate Forecast System Version 2. Journal of Climate, Vol. 27(6):2185–2208.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. Journal of the American Statistical Association, Vol. 63(324):1379–1389.

Teutschbein, C. and Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods. Journal of Hydrology, Vol. 456–457:12–29.

Toté, C., Patricio, D., Boogaard, H., van der Wijngaart, R., Tarnavsky, E. and Funk, C. (2015). Evaluation of satellite rainfall estimates for drought and flood monitoring in Mozambique. Remote Sensing, Vol. 7(2):1758–1776.

Umar, N. and Gray, A. (2023). Flooding in Nigeria: a review of its occurrence and impacts and approaches to modelling flood onset. International Journal of Environmental Studies, Vol. 80(3):540–561.

Wilks, D. S. (2011). Statistical Methods in the Atmospheric Sciences. Third Edition. Academic Press, Oxford, United Kingdom.

WMO (2018). Guide to Climatological Practices. WMO-No. 100, Third Edition. World Meteorological Organization, Geneva, Switzerland.

Yue, S., Pilon, P., Phinney, B. and Cavadias, G. (2002). The influence of autocorrelation on the ability to detect trend in hydrological series. Hydrological Processes, Vol. 16(9):1807–1829.

To cite this article:
Akubue, C. O. Ahaneku, I. E., Oduma, O. And Nwokoji, C.J. 2026. Compilation, Quality Control, And Satellite-Based Augmentation of Hydrometeorological Datasets for Abia State, Nigeria. 1(4): 31-48. https://journals.unizik.edu.ng/ujabe/