TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 24, No. 5, October 2026, pp. 1611~1631
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v24i5.27848

A comparative assessment of machine learning, statistical, and hybrid approaches in amending ECMWF rainfall bias in the Batanghari river basin

Arif Ma'rufi[1, 2], Tania June[1], I Putu Santikayasa[1, 5], Apip[3], Siswanto[4]
[1]Department of Geophysics and Meteorology, Faculty of Mathematics and Natural Sciences, IPB University, Bogor, Indonesia
[2]Climatology Station of Jambi, The Agency for Meteorology, Climatology and Geophysics (BMKG), Jambi, Indonesia
[3]Research Center for Limnology and Water Resources, National Research and Innovation Agency (BRIN), Cibinong, Indonesia
[4]Directorate of Applied Climate Services, The Agency for Meteorology, Climatology, and Geophysics (BMKG), Jakarta, Indonesia
[5]Centre for Climate Risk and Opportunity Management in Southeast Asia Pacific (CCROM–SEAP), IPB University, Bogor, Indonesia

Article Info

Article history:

Received Feb 3, 2026
Revised Jun 18, 2026
Accepted Jun 30, 2026

Keywords:

Bias correction
ECMWF
Extreme rainfall
Flood forecasting
Machine learning
Quantile mapping

ABSTRACT

In tropical basins, rainfall predictions using international numerical weather prediction (NWP) models, like the European Centre for Medium-Range Weather Forecast (ECMWF), exhibit significant persistent biases and insufficient ensemble dispersion. This study evaluates statistical, machine learning (ML)-based, and hybrid bias correction methods to improve daily rainfall estimates for Indonesia's Batanghari River Basin. Findings indicate that the ECMWF model significantly overestimates wet-day frequency (WDF) (81.5% versus 43.4% observed) and underestimates extreme maximum magnitudes, exhibiting an annual maximum rainfall bias of -15.6 mm/day. Standalone ML-based methods, such as random forest (RF) and extreme gradient boosting (XGBoost), experience significant variance deflation, inadequately addressing these substantial underestimations. In contrast, quantile mapping (QM) significantly reduces the intrinsic drizzle bias and achieves the highest monthly temporal stability among the tested frameworks (Kling-Gupta efficiency (KGE) = 0.341) with low processing requirements. Moreover, hybrid methodologies, including RF-QM and XGBoost-QM, successfully restore the natural ensemble spread and robustly reconstruct the upper tails of extreme rainfall, albeit at the expense of long-term volume conservation (resulting in negative monthly KGE). Ultimately, QM is favored as an effective baseline for continuous hydrological modeling, whereas hybrid methodologies are highly suited for evaluating localized flash flood hazards and short-term severe intensities.

Corresponding Author:

Tania June
Department of Geophysics and Meteorology, Faculty of Mathematics and Natural Sciences
IPB University
Raya Dramaga St., Bogor, 16680, West Java, Indonesia
Email: taniajune@apps.ipb.ac.id

1. INTRODUCTION

Accurate rainfall forecasting is fundamental for hydro-meteorological disaster early warning systems, water resource management, and agricultural planning [1], [2]. Global numerical weather prediction (NWP) models, particularly from the European Centre for Medium-Range Weather Forecasts (ECMWF), serve as the backbone of modern forecasting [3]. In the tropics, the coarse resolution of these models leads to noticeable systematic biases, primarily because they fail to resolve local convective dynamics [4], [5]. This creates two main issues: an overestimation of light rain (the 'drizzle effect') and an underestimation of heavy extremes. For catchments with complex terrain, relying on such biased data can undermine both hydrological modeling and flood mitigation efforts [6]–[9]. These impacts are especially critical in basins with intricate topography and high convective activity, such as the Batanghari River Basin in Indonesia.

Traditional bias correction typically relies on statistical methods like linear scaling (LS) and quantile mapping (QM). Although QM serves as a standard approach for aligning model outputs with empirical observations [10], its strictly univariate architecture inherently limits its capacity to resolve extreme convective anomalies and adapt to non-stationary climate dynamics [11]. To address this, machine learning (ML) algorithms like random forest (RF) [12] and extreme gradient boosting (XGBoost) [13] are increasingly applied to extract complex, non-linear patterns [14]–[16]. However, standalone ML models share a critical flaw: they tend to collapse data variance and distort the underlying statistical distribution [17], [18]. Combining the predictive strength of ML with the distributional stability of QM into a hybrid framework presents a promising alternative. This approach is particularly relevant for tropical climates, an environment where such hybrid evaluations are still very limited [19], [20].

Despite recent advancements in bias correction modeling, a significant research gap remains. Existing comparative assessments predominantly target mid-latitude regimes, creating a systematic knowledge deficit regarding convective-dominated tropical catchments [3], [7]. Moreover, current literature lacks an integrated benchmarking protocol capable of concurrently evaluating classical statistical techniques, fundamental linear baselines such as multiple linear regression (MLR) [21], and advanced hybrid architectures (random forest-quantile mapping (RFQM), and XGBoostQM). Prior investigations frequently neglect operational constraints, specifically the critical assessment of extreme precipitation indices (Rx1day), inter-member variance recovery, and computational overhead. Driven by intense monsoonal forcing and a documented history of severe pluvial hazards [22], the Batanghari River Basin serves as a robust empirical testbed for executing this multidimensional evaluation.

Motivated by these methodological gaps, this study executes a comprehensive benchmarking of eight distinct bias correction frameworks. The evaluated suite encompasses traditional statistical models (LS, QM, and MLR), standalone ML configurations (RF and XGBoost), and hybrid architectures (RFQM, XGBoostQM, and LS empirical QM (LSEQM)). Although existing literature has investigated a broad spectrum of techniques, spanning univariate statistical mapping [3], [23] and hybrid statistical combinations [20] highly complex deep learning models [1], [4], [11], a critical void persists regarding the integration of ML algorithms with QM to specifically resolve extreme daily convective anomalies in tropical regimes. The comparison of recent rainfall bias correction studies and the proposed work is shown in Table 1.

Table 1. Comparison of recent rainfall bias correction studies and the proposed work

References	Methods	Data source and region	Research gap / limitation context for this study
Pierce *et al*. [24]	Statistical (QM, EDCDFm, PresRat)	GCMs (United States)	Focuses on climate change scales; less effective for daily convective peaks.
Um *et al*. [20]	Statistical and Hybrid (QM, LS, and Spline)	GCMs (South Korea)	Restricted to statistical combinations; lacks ML integration.
Ayugi *et al*. [25]	Statistical (QMBC)	RCMs (Kenya)	Utilizes single method; misses hybrid approaches for complex topographies.
Han *et al*. [5]	ML (CU-net)	ECMWF (North China)	Excludes rainfall, emphasizing spatial grid corrections over extreme point evaluations.
Wang *et al*. [1]	ML (customized DL) and statistical (QM)	Reanalysis / Radar (Gulf of Mexico)	Focuses on hourly spatial downscaling using complex deep learning, rather than daily hybrid corrections.
Seo and Ahn [3]	Statistical (QM) and ML (long short-term memory (LSTM))	WRF / ERA5 (South Korea)	Compares standalone models, explicitly noting that ML struggles with extreme rainfall (>50 mm/day).
Tadase and Tekile [11]	ML (XGBoost) and statistical (EQM)	CMIP6 / CHIRPS (Unspecified)	Evaluates standalone XGBoost and EQM, concluding that ML underestimates rare extremes while EQM preserves them. Strongly justifies the need for a hybrid framework.
Proposed study	Hybrid (ML-QM), ML (RF and XGB), statistical (LS, QM, LSEQM, and MLR)	ECMWF (Batanghari River Basin, Indonesia)	Evaluates QM as a robust baseline and integrates it with ML (hybrid ML-QM) to preserve daily extremes in a tropical convective basin while balancing computational cost.

The innovation of this research lies in its multi-criteria evaluation using a leave-one-year-out (LOYO) cross-validation and ensemble-wide approach to ensure predictive stability. By identifying the trade-offs between volumetric accuracy (at daily and monthly scales) and the precise reconstruction of extreme rainfall magnitudes, this work delivers new value in determining the optimal bias correction strategy for flood early warning systems (FEWS) within large-scale river basins. This is critical for the downstream

Batanghari River Basin, where reliable rainfall accumulation drives hydrological models predicting cumulative flood discharge and supporting disaster risk management along the primary river network.

From a computing and telecommunication perspective, accurate and bias-corrected rainfall data serve as the vital computational backbone for modern hydro-informatics. Generating high-fidelity precipitation datasets are essential for driving cloud-based, real-time FEWS and optimizing internet of things (IoT) based environmental sensor networks. Reducing predictive uncertainty at the computational level directly enhances the efficiency of telemetry infrastructures. This improvement allows sensor networks to optimize data transmission rates and significantly lower the risk of false alarms, ultimately delivering highly reliable early warnings to at-risk populations.

The remainder of this manuscript is organized as follows. Section 2 delineates the geographical attributes of the study catchment, the hydro-meteorological datasets, and the mathematical foundations of the evaluated bias correction frameworks. Section 3 expounds upon the empirical findings, offering a rigorous diagnostic assessment of relative model performance across alternative scenarios. We wrap things up in section 4 by bringing together our key takeaways and suggesting how these tools can be deployed in real-world operations.

2. METHOD

This analysis focused on the Batanghari River Basin. Our analysis heavily depends on ECMWF daily rainfall forecasts (2020–2024) across 51 ensemble members. Once extracted at a daily scale, we spatially paired this forecast data with our target basin to prepare for the bias correction step. We then benchmarked the model results against real-world observations. This ground-truth data was sourced from 86 rainfall stations operated by the Meteorology, Climatology, and Geophysics Agency (BMKG) and the River Basin Organization of Sumatra VI (BWSS VI). The geographic domain of the investigated Batanghari River Basin within the Indonesian archipelago is delineated in Figure 1, with the right-hand inset specifically highlighting its macro-scale orientation. Meanwhile, the left panel maps out the basin borders, the ECMWF grid, and the spread of the rain gauges.

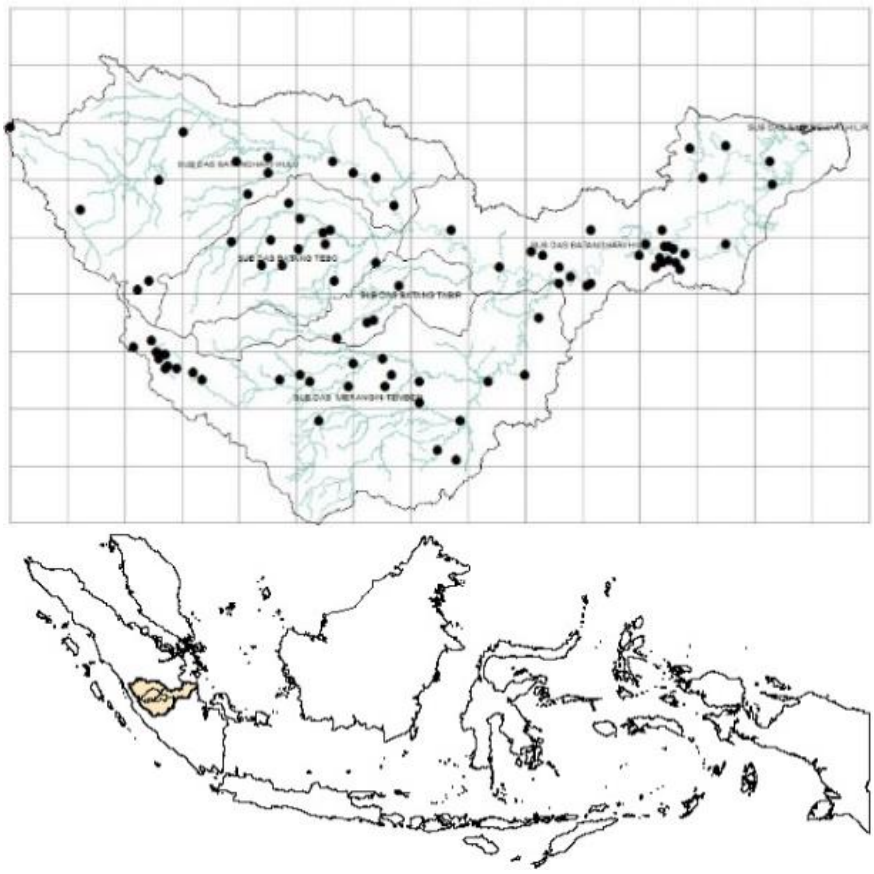


Figure 1. Study area of the Batanghari River Basin with ECMWF model grids and observation stations

To strictly preserve data quality, only observation stations with missing records under the 5% threshold were analyzed. We applied a K-nearest neighbors (KNN) algorithm to impute these minimal temporal gaps based on neighboring daily similarities, carefully avoiding any distortion to the original rainfall distribution. The structural topography of the basin was then validated using the 90-m resolution multi-error-removed improved-terrain (MERIT) digital elevation model (DEM). The computational experiments for all bias correction frameworks were executed on a local workstation equipped with a 13th Gen Intel Core i5-1335U processor (1.30 GHz), 16.0 GB of RAM, and a 64-bit operating system.

To mitigate non-linear errors in the modeled rainfall data, this study systematically evaluates three distinct categories of bias correction algorithms. The first category comprises tree-based ensemble ML models, namely RF and XGBoost. Second, we have traditional statistical methods like LS, QM, LSEQM, and MLR. Finally, we tested hybrid models (RFQM and XGBoostQM) that try to combine the strengths of ML-based techniques with statistical distribution fixes. We purposely brought MLR into the mix to act as a solid baseline. While QM just tries to match statistical distributions, MLR actually builds a functional link between predictors and real rainfall. By doing this, we can see if we need heavy, complex non-linear models to handle tropical convective storms.

The ML algorithms and their hybrid counterparts were configured using a single predictor: the raw daily rainfall estimates from the ECMWF. This single-feature approach was explicitly selected to directly correct the magnitude and distribution of model biases against observed data, ensuring a straightforward and replicable correction process for tropical climates. To establish a fundamental data-driven benchmark, the RF regressor was deployed [12], [23]. Rather than relying on simple linear scaling, this architecture leverages an aggregation of uncoupled decision trees to systematically resolve complex, non-linear error structures within the precipitation fields.

A randomized search cross-validation protocol was deployed to tune the RF algorithm and balance the bias-variance trade-off. Within the optimization space, the number of decision trees spanned from 100 to 500, and maximum structural depths were tested across values of 10, 20, 30, 40, and 50, in addition to an unconstrained baseline. Feature selection criteria evaluated both automated (‘auto’) and square-root (‘sqrt’) modalities. Node splitting complexity was explicitly governed by setting minimum internal split samples to 2, 5, and 10, alongside terminal leaf limits of 1, 2, and 4. The parameter combination that maximized validation accuracy was ultimately selected to train the operational architecture.

We also implemented XGBoost [13], as an alternative tree-based algorithm. Following the same randomized search procedure used for RF, the XGBoost tuning grid evaluated $n\_estimators$ at 100, 200, and 300, and $learning\_rate$ values of 0.05, 0.1, and 0.2. To constrain individual tree complexity and prevent overfitting, max_depth was strictly limited to 3, 5, 7, and 10. To prevent overfitting and introduce stochasticity into the training process, data and feature subsampling ratios (subsample and $colsample\_bytree$) were tested at 0.7, 0.8, and 1.0. Upon completion of the evaluation phase, the highest-performing parameter set was selected for the final predictive simulations.

After describing ML-based methods, the next step is statistical methods. In this study the statistical methods include parametric and non-parametric techniques: LS for mean adjustment, QM for data distribution correction, the combination of LSEQM, and an MLR method that incorporates geographical factors into the correction model. On the statistical side, our main baseline method is LS. This straightforward technique adjusts the model data based on the differences in monthly averages between the forecasts and actual observations. The LS method adjusts daily simulated rainfall by applying a multiplicative scaling factor, derived from the ratio of observed to modeled long-term means. Although this approach effectively corrects the mean bias (MB) at a monthly scale, its assumption of a constant multiplicative error limits its ability to adjust the variance or correct extreme precipitation magnitudes. The mathematical formulation used to shift the model’s mean is defined as (1) [24]:

$$P_d^* = P_d \times \left(\frac{\mu(P_{obs,m})}{\mu(P_{mod,m})}\right) \tag{1}$$

where P_{d^*} represents the corrected daily rainfall and P_d denotes the raw model output. The variables $\mu(P_{obs,m})$ and $\mu(P_{mod,m})$ indicate the long-term monthly means of the observed and simulated data, respectively.

Addressing the first-moment constraints of linear scaling, QM recalibrates the entire empirical frequency distribution of the precipitation dataset rather than restricting adjustments to the mean. Mechanically, this methodology establishes a non-linear transfer function that aligns the cumulative distribution function (CDF) of the simulated outputs with that of the historical baseline, translating the simulated probability density directly into the observed sample space. This direct mapping makes it a powerful tool for adjusting the occurrence of rainy days as well as those tricky, intense precipitation spikes. The CDF-matching process is (2) [26]:

$$P_d^* = F_{obs}^{-1}\left(F_{mod}(P_d)\right) \tag{2}$$

where F_{mod} is the CDF of the model data, F_{obs}^{-1} is the inverse CDF of the observation data, and P_d is the rainfall value to be corrected.

This research employs a combination of LSEQM methodology. This integrated approach is implemented in two sequential stages. In the initial stage, a LS operator is enforced to rectify systemic mean anomalies and preserve monthly volumetric mass-balance integrity. The intermediate scaled outputs are then subjected to empirical QM to refine the sub-monthly frequency distribution. This cascading integration simultaneously counteracts cumulative volumetric distortions while reconstructing the higher-order daily precipitation variability that standalone approaches typically fail to resolve. Consequently, the coupled algorithmic execution governing this two-step procedure is (3) [19], [20]:

$$P_{LSEQM}^* = QM_{empirical}\left(P_d \times \frac{\mu(P_{obs})}{\mu(P_{mod})}\right) \tag{3}$$

where $Q\mathrm{M}_{empirical}$ denotes the empirical QM applied to the linearly scaled output.

To explicitly model spatial heterogeneity and elevation-dependent variance, a multivariate MLR formulation serves as the baseline. Unlike its univariate counterparts, the MLR framework ingests geographic coordinates concurrently with terrain elevation to function as independent predictors. This configuration operates on the understanding that systemic precipitation anomalies are heavily modulated by localized orographic features and geometric positioning across the catchment. By accounting for these spatial gradients, the MLR algorithm estimates the co rrected rainfall through a linear combination of the modeled data and the spatial variables, formulated as (4) [21]:

$$P_d^* = \beta_0 + \beta_1 P_d + \beta_2(\mathrm{Lon}) + \beta_3(\mathrm{Lat}) + \beta_4(\mathrm{MDPL}) + \epsilon \tag{4}$$

where P_d represents the raw model rainfall and β_0 is the intercept. The parameters β_1 through β_4 denote the regression coefficients for each predictor, and ϵ accounts for the residual error term.

While ML algorithms (RF and XGBoost) effectively reduce prediction variance through non-linear feature extraction, empirical statistical methods like QM are fundamentally better suited for reconstructing historical distribution curves. However, relying on either approach in isolation exposes inherent structural limitations. ML-based techniques may sometimes produce under-dispersive forecasts, leading to an underestimation of extreme events, while quantitative models generally overlook interactions between meteorological variables. This study formulates a multi-stage hybrid methodology aimed at integrating the advantages of both areas. Hybrid methods were developed by applying QM as a post-processing step to the rainfall outputs generated by the ML-based models. In this framework, RF and XGBoost were first trained to predict rainfall based on ECMWF inputs, after which QM was applied to the model outputs to further correct distributional biases.

The first hybrid configuration, RFQM, sequentially couples the non-linear flexibility of RF with the statistical precision of QM. The initial phase trains the RF model using predictor variables (model rainfall and time indices) to generate preliminary forecasts minimizing the root mean square error (RMSE). Because RF forecasts, while reliable, tend to diminish extreme values via ensemble averaging, the second phase applies QM to the RF outputs. To force the residuals to follow the observed CDF and restore extreme event variance, this hybrid formulation passes the RF predictions through a QM transformation [20], [27]:

$$P_{RFQM}^* = F_{obs}^{-1}\left(F_{RF}\left(\widehat{P_{RF}}\right)\right) \tag{5}$$

where $\widehat{P_{RF}}$ is the rainfall prediction output from the RF model, F_{RF} is the CDF of these RF predictions, and F_{obs}^{-1} is the inverse CDF of the observation data. Consequently, RFQM rectifies both structural and distributional biases concurrently.

Following the same logic as RFQM, our second hybrid model, XGBoostQM, relies on XGBoost as its core engine. We specifically chose XGBoost because its boosting mechanism excels at driving down bias by fixing previous errors step-by-step. While this algorithm does a great job at catching mean trends, its residual forecasts often struggle at the extreme tails—meaning it can miss the mark during intense storms. To fix this, we apply QM to the XGBoost output, forcing the predicted statistics to match the actual observed profile. To ensure chronological accuracy and statistical validity in the corrected series, the following (6) defines the distributional adjustment [13], [28]:

$$P^{*}_{XGBQM} = F^{-1}_{obs}\left(F_{XGB}\left(\widehat{P_{XGB}}\right)\right) \tag{6}$$

where $\widehat{P_{XGB}}$ represents the predicted rainfall from XGBoost, and the composite mapping function $F^{-1}_{obs} \circ F_{XGB}$ recalibrates these values to align with the empirical distribution of ground-based observations.

To validate the performance of the post-processed datasets, this study deploys a bifurcated evaluation protocol targeting both the statistical properties and physical attributes of the precipitation fields. Within this framework, the statistical tier systematically quantifies temporal co-variance and numerical fidelity. Concurrently, the physical validation evaluates how well the models capture local climatic frequencies and extreme distribution profiles.

To evaluate the temporal stability of the bias correction models, a LOYO cross-validation framework was implemented. This iterative procedure systematically trains the algorithms on all historical data while withholding a single, independent year for validation, repeating the sequence across each year of the dataset to ensure robust generalization. This allows us to squeeze every drop of utility out of the data—a must-have in hydrology, where long-term records are notoriously scarce [29]. More importantly, it acts as a proxy for real-time operational forecasting. Evaluating the algorithms against independent, out-of-sample years ensures that temporal dependence bias is effectively mitigated [30].

2.1. Statistical performance and evaluation metrics

Framework efficacy is evaluated using a complementary mix of statistical and physical metrics. To ensure this evaluation framework is accessible to a cross-disciplinary audience, including researchers in control engineering and telemetry, the physical and operational meanings of each metric are explicitly integrated. Chronological alignment, variance, and systematic errors are quantified via MB, Kling-Gupta efficiency (KGE), and RMSE [31], the latter leveraged for its quadratic sensitivity to large convective deviations. To validate upper-tail extremes, these criteria are coupled with physical descriptors, namely wet-day frequency (WDF) and Rx1day. Operating in tandem, this configuration ensures a robust diagnostic baseline critical for operational tropical flood forecasting.

2.1.1. RMSE

The aggregate scale of residuals between model estimates and field observations is quantified via the RMSE, which is mathematically expressed as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2} \tag{7}$$

where M_i represents the modeled rainfall, O_i denotes the observed rainfall at time step i, and N is the total number of observations.

The quadratic penalization inherent in error computation renders RMSE exceptionally responsive to extreme value spikes. From a signal processing perspective, RMSE functions as a primary indicator of overall forecast variance and underlying signal noise. Within tropical catchments, this sensitivity proves highly advantageous; minimizing RMSE guarantees that the model reliably replicates intense convective cloudbursts establishing an indispensable parameter for operational flood forecasting and disaster risk reduction.

2.1.2. KGE

Simultaneous assessment of phase alignment, dispersion fidelity, and mean mass-balance is secured by adopting the KGE as the principal aggregate metric. For control engineering applications, KGE acts as a composite performance index that balances timing accuracy, signal amplitude fidelity, and temporal stability. The KGE framework is (8):

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \tag{8}$$

where r denotes the Pearson correlation coefficient between the simulated and observed series, α represents the variability ratio (the ratio of modeled to observed standard deviations), and β is the volumetric bias ratio defined by the ratio of their respective means.

Unlike traditional residual metrics, KGE prevents the artificial inflation of performance scores by isolating phase shifts, amplitude errors, and volume biases. As the KGE metric converges toward unity, it demonstrates enhanced model capacity to reproduce concurrent seasonal climate cycles and the high-frequency variability inherent in tropical convective precipitation regimes.

2.1.3. MB

The propensity of the framework to introduce directional overestimation or underestimation across the precipitation fields is quantified via the MB, which is mathematically structured as:

$$MB = \frac{1}{N}\sum_{i=1}^{N}(M_i - O_i) \quad (9)$$

where M_i represents the modeled precipitation, O_i denotes the observed rainfall, and N indicates the total number of evaluation days.

In operational telemetry and reservoir management, this metric is vital to determine whether the predictive model consistently injects false water volumes into the system or misses actual water inputs. Rectifying these long-term systematic biases directly prevents distortions in the basin's volumetric water budget. In the context of tropical watersheds such as the Batanghari, a positive MB inadvertently compromises flood forecasting efficiency by elevating false alarms, whereas a negative bias downscales extreme precipitation intensities to the point of obscuring critical disaster thresholds. Consequently, suppressing the MB value serves as a crucial calibration benchmark for aligning downstream hydrological inputs with empirical field observations.

2.2. Verification of physical distribution

WDF: this diagnostic is essential for identifying the widely documented 'drizzle effect', which represents a systemic modeling artifact where simulation frameworks consistently overreport high frequency and trace level precipitation events. By enforcing a standard operational threshold of 1.0 mm to define a wet day, the analysis systematically filters out spurious rainfall traces to preserve strict meteorological relevance. Accurately replicating this frequency is an absolute prerequisite for ensuring the fidelity of multidisciplinary downstream applications, including agrometeorological planning and drought risk assessment. To quantify the proportion of these meteorologically significant days across the evaluation timeframe, WDF is formalized via the following relationship:

$$Freq_{wet} = \frac{1}{N}\sum_{i=1}^{N} I(P_i \geq 1.0 \text{ mm}) \quad (10)$$

where $I(\cdot)$ denotes the binary indicator function that yields 1 if the wet-day condition is satisfied, and 0 otherwise.

The annual maximum one day precipitation (Rx1day): this metric isolate extreme upper tail anomalies to assess how well the model reproduces high magnitude and rare storm events over an annual cycle. The evaluation focuses on comparing the probability distribution of these annual peak values between the simulated outputs and empirical observations. Engineers and system operators interpret this metric as the capacity of the predictive model to accurately capture severe sudden impulse inputs and peak system stress levels. Precision in this metric directly influences flood risk management and the engineering design of civil infrastructure such as hydraulic structures. To capture this extreme index, the framework isolates the absolute highest daily precipitation accumulation within each distinct year according to the following mathematical relationship:

$$Rx1day_y = \max\left(P_{d,y}\right) \quad (11)$$

where $max\left(P_{d,y}\right)$ represents the operation that extracts the single highest daily rainfall value ($P_{d,y}$) out of all days (d) within a specific year (y).

To accurately preserve the magnitude of extreme events and avoid the smoothing effect of ensemble averaging, the Rx1day metric for the ECMWF forecasts was computed by first identifying the annual maximum daily rainfall for each of the 51 individual ensemble members. Following this extraction, the ensemble mean of these 51 maximum values was calculated to represent the final model Rx1day. Figure 2 delineates the entire methodological architecture of this research, tracking the operational progression from initial data preprocessing through the deployment of the three bias correction pathways to the final dual validation stages.

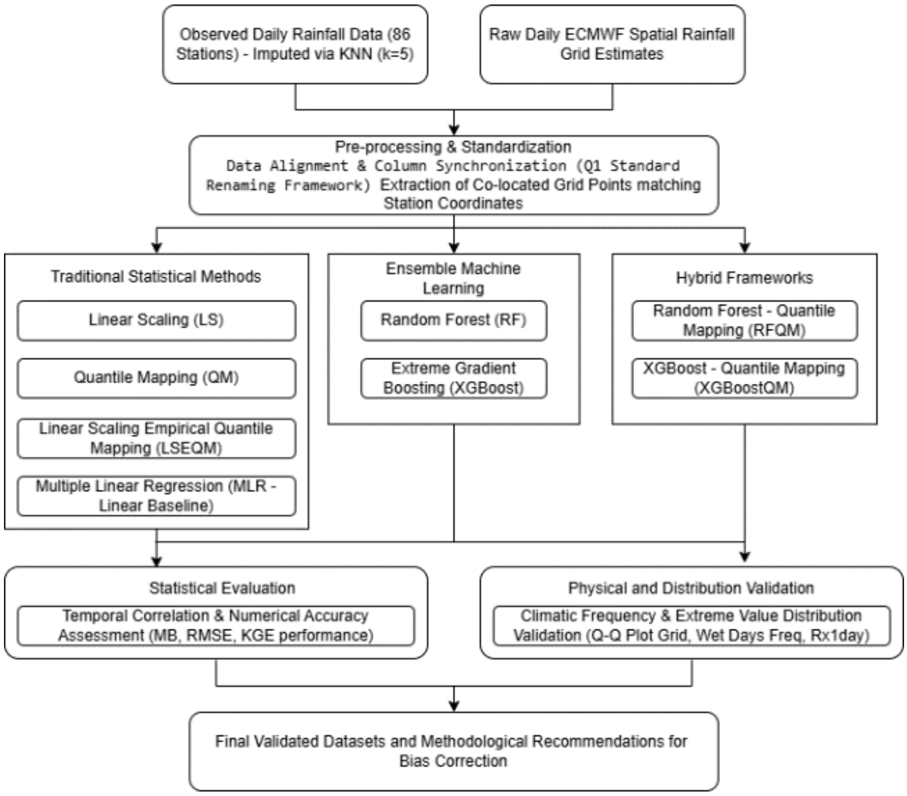


Figure 2. Methodological flowchart of the bias correction and evaluation framework

3. RESULTS AND DISCUSSION

3.1. Characterization of observational and raw ECMWF data

The initial phase of this assessment examines how effectively the simulation framework replicates seasonal rainfall climatology. Figure 3 indicates that the raw ECMWF outputs generally mirror the seasonal macrocycles of the Batanghari river basin by identifying shifting wet and dry periods. Nevertheless, substantial monthly discrepancies persist, demonstrating that the capacity to resolve localized rainfall remains limited despite the successful identification of overarching weather patterns. This systemic bottleneck aligns with regional findings [32], who identified similar performance degradation in tropical reanalysis products. Such variations in volume suggest that the convective parameterization architecture requires modification to accommodate the complex localized topography governing the Batanghari region. The amplitude errors, even with accurate phase timing, correspond to the technical constraints of the ECMWF system as identified by [33].

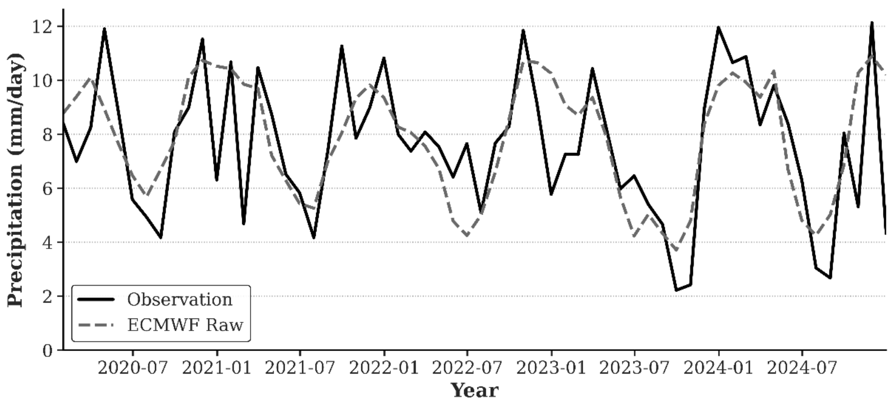


Figure 3. Comparison of monthly average rainfall between observations and ECMWF raw

While raw ECMWF tends to underestimate high-intensity rainfall events (extremes), it exhibits a systematic wet bias in terms of total accumulated rainfall. This contradiction is largely a byproduct of the 'drizzle effect', where the model frequently overpredicts low-intensity rainfall compared to ground observations. Table 2 and Figures 4(a) and (b) highlight these substantial statistical differences. Although the mean daily precipitation magnitudes are comparable (7.6 vs. 8.1 mm/day), the uncorrected model introduces severe systematic distortions. This 'drizzle bias' is evidenced by a highly elevated median (+5.1 mm/day bias) and a substantial overestimation of wet-day frequency, which reaches 81.5% compared to the observed baseline of 43.4% (Figure 4(b)). This verifies that the model frequently simulates sustained light rainfall due to its coarse spatial resolution, a feature already recorded in global reanalysis studies [31]-[33]. This situation occurs physically as the model disperses moisture throughout an extensive grid cell instead of concentrating it into violent, localized storms.

Furthermore, the raw dataset displays significant underdispersion highlighted by a truncated standard deviation of 10.6 versus 15.2 mm per day, validating earlier observations [33]. This dampening of natural variability directly degrades how effectively the system captures extreme events, causing the average annual maximum rainfall to be severely underestimated with a systematic negative bias of 15.6 mm per day according to Figure 4(a). This underestimation poses a significant risk for hydrological modeling, corroborating the findings of [25] about the effectiveness of climate models in tropical environments.

Table 2. A comparison of basic statistics between observations and ECMWF raw

Parameter	Observation (mm/day)	ECMWF raw (mm/day)	Bias (mm/day)
Mean (mm/day)	7.6	8.1	0.5
Median (mm/day)	0.3	5.4	5.1
Standard deviation	15.2	10.6	-4.6
Average annual max (Rx1day)	103.0	87.4	-15.6
Minimum (mm/day)	0.0	0.0	0.0
Wet days (%)	43.4	81.5	38.1

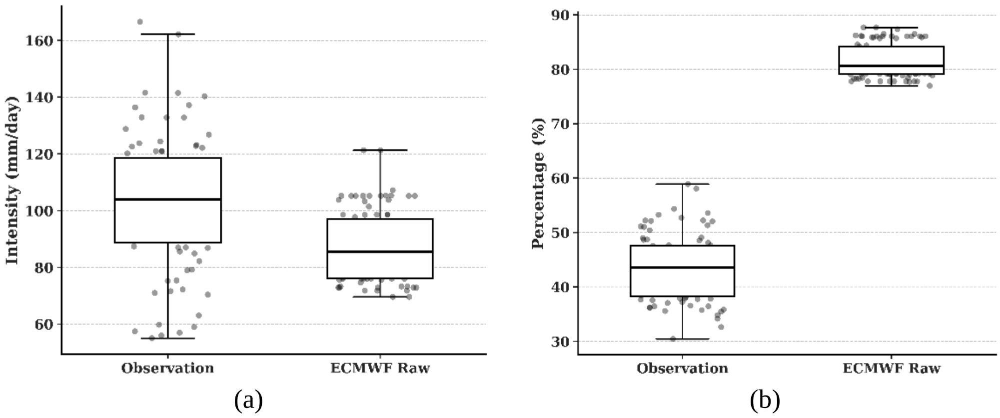


Figure 4. Comparative boxplots of observational and ECMWF raw distributions: (a) annual maximum rainfall (Rx1day) and (b) wet-days percentage (%)

3.2. Comparison of average performance of bias correction methods

To systematically assess the reliability of the proposed bias correction methods, a comprehensive evaluation was conducted using a LOYO cross-validation scheme across 86 observational stations in the Batanghari River Basin. We designed this benchmark to test each method's ability to clean up statistical errors and tighten temporal correlations, all while keeping an eye on operational feasibility. In Table 3, we summarize the performance of every approach relative to the raw ECMWF data. The data highlights a clear divergence: while some methods excel at slashing errors or mimicking variability, they often differ drastically in the computational effort they require.

The assessment of daily and monthly measures (Table 3) indicates clear performance trade-offs. While the hybrid ML-based architectures encompassing RFQM and XGBoostQM achieved success in aligning daily temporal trends, they paradoxically compromised volume precision by generating RMSE values greater than 18 mm per day alongside elevated computational costs. This particular conclusion perfectly illustrates the 'double penalty' phenomenon articulated by [34]. In hydrometeorological evaluations, this phenomenon occurs when a high-resolution model accurately predicts the magnitude of a severe rainfall event but slightly misplaces it in exact time or location. Consequently, traditional point-by-point metrics like RMSE penalize the model twice: once for a missed event at the observation point, and once for a false alarm where the rain was actually predicted. This dynamic corroborates the operational impediments of utilizing overly simplistic evaluation metrics for complex ML-based outputs, as identified by [35].

Table 3. Comparison of performance indicators among various methods

Methods	KGE daily	KGE monthly	MB (mm/day)	RMSE (mm/day)	Calibration time (minutes)	Opr. time (second)
ECMWF raw	-0.111	0.246	6.817	17.323	-	-
RF	-0.267	0.035	1.290	15.423	59.07	358.66
XGBoost	-0.331	-0.066	1.315	15.430	64.99	1087.28
LS	-0.224	0.262	0.275	15.241	3.93	3.94
QM	-0.161	0.341	1.208	15.498	3.33	5.09
LSEQM	-0.227	0.155	-0.952	15.865	3.48	7.13
MLR	-0.309	-0.182	0.000	15.203	3.84	3.29
RFQM	-0.071	-0.111	3.206	18.158	68.51	322.71
XGBoostQM	-0.108	-0.204	4.071	18.348	77.85	731.67

Interestingly, although standalone models like RF and XGBoost posted the lowest RMSE, we found these numbers to be somewhat of a statistical illusion. This phenomenon is driven by variance deflation, an optimization characteristic where the algorithms prioritize conditional mean estimation by smoothing out natural data fluctuations. Although this smoothing mechanism optimizes overarching residual error scores, it does so at the cost of failing to capture high magnitude and extreme precipitation events. Dependence on variance deflated outputs introduces profound operational risks into civil engineering and disaster management applications. This anomaly can prompt critical underestimations within flood early warning frameworks or lead to structural deficits in agricultural drainage planning, representing a systemic limitation highlighted by [36]. In contrast, QM proved to be the most balanced and operationally viable standalone technique within this study. Through the concurrent reduction of volumetric biases and preservation of physical consistency, QM ensures that mean water availability assessments for agrometeorological planning remain robustly valid, aligning with the conclusions drawn by [26]. The comparative statistical performance of these standalone frameworks is detailed in Table 3.

The investigation indicates that although statistical approaches are nearly instantaneous, the operational requirements of hybrid models remain feasible for real-time forecasting, despite their rigorous training phases. Shifting the evaluation to a monthly resolution uncovers a profound divergence in predictive accuracy. Figure 5 maps the average KGE trajectories across all configurations to illustrate these monthly fluctuations. Empirical comparisons highlight a severe performance deficit among the standalone and hybrid ML-based architectures, as their efficiency metrics deteriorate significantly and frequently plunge into negative KGE territory. This collapse exposes their structural instability at an aggregated temporal scale. Furthermore, such degradation reveals a systemic limitation where purely algorithmic models struggle to independently sustain long term mass balance, reflecting the inherent difficulty of embedding physical mass conservation constraints into algorithmic frameworks [37]. In stark contrast, QM demonstrated exceptional temporal robustness by securing the highest monthly KGE score of 0.341. By proficiently preserving volume aggregates through direct cumulative distribution adjustment, QM proves to be intrinsically more dependable for long-term water balance evaluations in this basin [38].

Ultimately, while all evaluated bias correction techniques substantially modified the raw ECMWF precipitation forecasts, identifying the optimal framework requires a careful reconciliation of statistical performance and operational constraints. Our findings indicate that the standalone QM approach provides a superior operational compromise compared to alternative methods. This framework effectively minimizes systematic bias while sustaining a balanced equilibrium among statistical precision, mass balance, and computational efficiency. Such characteristics make it highly ideal as a foundational baseline for general operational forecasting scenarios. In contrast, standalone ML configurations encompassing RF and XGBoost proved inadequate at replicating extreme precipitation variability due to pronounced variance deflation. Although the hybrid frameworks (RFQM and XGBoostQM) successfully resolved these upper tail limitations and delivered peak accuracy improvements for high intensity events, they simultaneously introduced substantial architectural complexity and compromised monthly mass balance conservation. Ultimately, these divergent outcomes underscore the necessity of aligning the chosen methodology with distinct operational objectives. This strategy entails prioritizing QM for continuous hydrological simulations while dedicating hybrid methods exclusively toward targeted extreme flood risk mitigation.

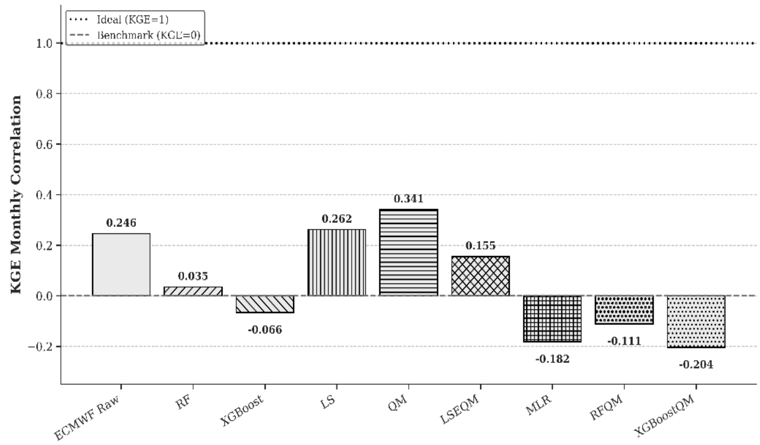


Figure 5. Comparison of monthly KGE performance across bias correction methods

3.3. Analysis of ensemble performance and uncertainty spread

Since ensemble forecasting provides a range of possible future states rather than a single deterministic value, evaluating the consistency of the 51 ECMWF ensemble members is vital for assessing predictive uncertainty. While the deterministic mean provides a baseline, it does not capture the full predictive uncertainty. Here, we further investigate the distribution of performance, examining how the metrics vary across every member of the ensemble. Evaluating the distribution of KGE scores (Figure 6) determines whether the bias correction frameworks preserve critical ensemble dispersion while enhancing deterministic accuracy, or conversely, induce an artificial reduction in ensemble spread.

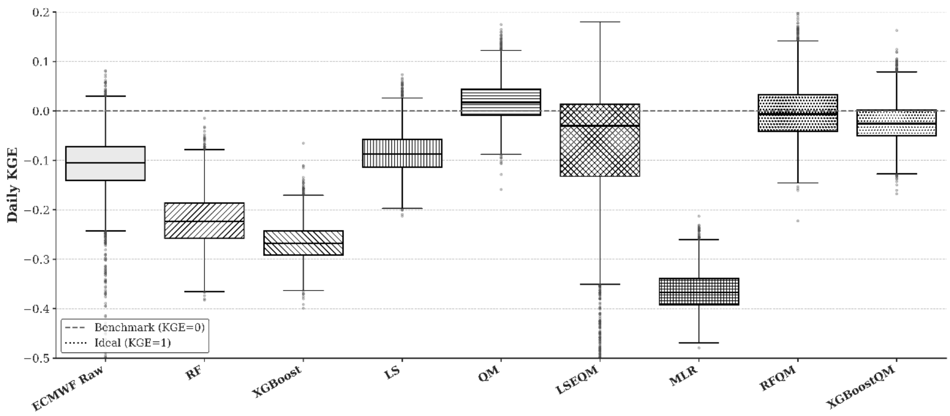


Figure 6. Ensemble diversity and daily KGE spread across bias correction methods

The probabilistic behavior of the ensemble system was evaluated by analyzing the distributional consistency of member performance (Figure 6). The raw ECMWF output exhibits inadequate baseline capabilities, reflecting a daily median KGE of approximately negative 0.1. Such a deficiency reinforces the critical necessity of implementing bias correction frameworks, corroborating the earlier conclusions documented by [37]. Standalone ML methodologies and MLR paradoxically trigger severe performance degradation by pushing the KGE distributions deeper into negative territory. The origin of this restriction lies within their algorithmic objective functions that depend heavily on mean squared error (MSE) minimization. This mathematical optimization inherently flattens individual predictions and eliminates the vital ensemble spread required for accurate physical representation. In response to these challenges, the hybrid architectures including RFQM and XGBoostQM, alongside the standalone QM technique, successfully neutralized this constraint and propelled the comprehensive performance distributions toward vastly superior efficiency levels. While the XGBoostQM framework achieved substantial gains over its standalone version by raising the median KGE near the zero boundary, RFQM ultimately delivered the most robust equilibrium between global distributional stability and effective variance reduction. It managed to boost the accuracy across all 51 members smoothly, avoiding the high volatility observed in the LSEQM results.

The dependability of the ensemble spread was further evaluated using the distribution of the mean annual maximum daily rainfall (Rx1day) (Figure 7). The observational data reveals considerable natural variability across the tropical basin, with a median intensity near 100 mm/day. The ECMWF Raw model moderately captures this spread but exhibits general under-dispersion, reflecting an 'overconfident' forecast that falsely assures low extreme rainfall probabilities, a misleading characteristic commonly observed in global weather prediction models [37]. Furthermore, a critical phenomenon, known as 'spread collapse', is glaringly evident in the standalone ML-based models (RF and XGBoost) and MLR. Their interquartile ranges are exceptionally narrow and severely underestimate the extreme rainfall magnitudes (median < 30 mm/day). This demonstrates that pure deterministic ML-based modifications aggressively dampen natural extreme events, rendering them unsuitable for extreme weather forecasting. standalone QM and the hybrid models (RFQM and XGBoostQM) were far more successful at recovering a realistic uncertainty framework. By expanding the ensemble spread to align with observed variability, these distribution-based approaches deliver a truer depiction of future flood risks. This directly satisfies the main goal of probabilistic bias correction ensuring that hydrological risk assessments are not just mathematical exercises, but are both robust and grounded in reality [38].

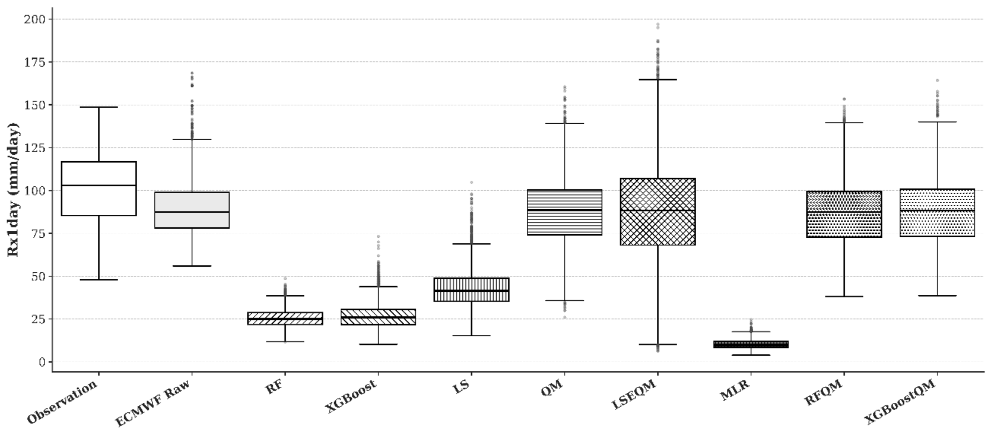


Figure 7. Analysis of ensemble spread and uncertainty in annual maximum rainfall (Rx1day) distributions

3.4. Validation of statistical distribution attributes

In this section, we look beyond simple error reduction to assess the physical coherence of the corrected rainfall, ensuring the results remain hydrologically sound. To verify that the regional water balance remained intact, we first mapped how mean daily rainfall is distributed across the study area. The spatial baseline visualized in Figure 8 sets the foundational benchmark for the upcoming evaluation phases. The majority of the bias correction architectures successfully preserve the volumetric water budget by producing spatial distributions that closely mirror empirical observations. Such consistency proves that these methodologies properly resolve magnitude errors without distorting the cumulative catchment inflow, a conclusion that corroborates the long-term statistical trends documented by [39]. On the other hand, the LSEQM technique demonstrates severe spatial instability and completely fails to generate a coherent precipitation field. This localized divergence reveals that the LS component within LSEQM remains highly vulnerable to spatial outliers, effectively crippling its reliability for water balance assessments across complex topography [36].

In addition to securing volumetric accuracy, an effective simulation architecture needs to accurately reproduce the temporal intermittency of local rainfall. Assessing the WDF as presented in Figure 9 confirms the pervasive drizzle anomaly within the uncorrected ECMWF output. This raw dataset overestimates rainy day occurrences at nearly 80%, contrasting sharply with the empirically recorded baseline of 43%. In this metric, the standalone ML-based models (RF and XGBoost) and MLR demonstrated a significant failure by predicting low-intensity rainfall nearly every day (frequency approaching 100%). This 'drizzle effect' completely fails to capture the natural dry-day sequences critical for agricultural planning. On the other hand, QM and the hybrid approaches (RFQM and XGBoostQM) successfully corrected this bias. By explicitly altering the CDF at low thresholds, these methods restored a realistic likelihood of dry days.

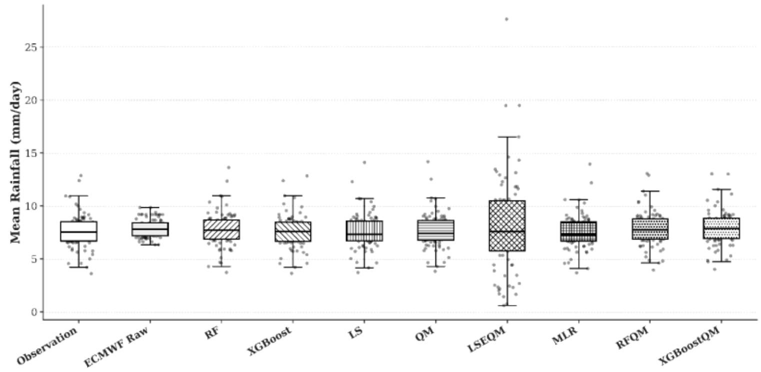


Figure 8. Comparative analysis of mean daily rainfall distributions with various methodologies

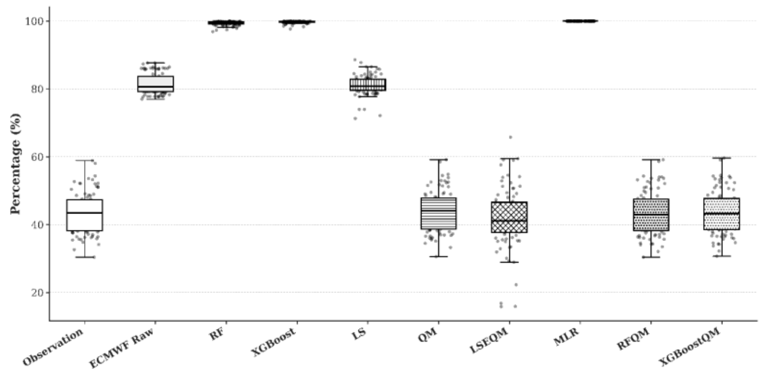


Figure 9. Analysis of the frequency distributions of rainfall days utilizing various methodologies

Figure 10 provides a detailed comparison of the station level distributions for the mean annual maximum precipitation (Rx1day), explicitly highlighting how effectively each bias correction framework replicates extreme rainfall variability. The observational data and the ECMWF Raw dataset exhibit a broad distribution, reflecting the high spatial variability of extreme convective rainfall. However, a severe phenomenon known as ‘spread collapse’ is glaringly evident in the standalone ML-based models (RF and XGBoost) and MLR. Because these algorithms are fundamentally optimized to minimize the MSE, they tend to predict values converging toward the mean, thereby systematically underestimating extreme rainfall magnitudes (median < 30 mm/day) and eliminating the spatial variance. Conversely, standalone QM and the hybrid approaches (RFQM and XGBoostQM) successfully overcome this limitation. By integrating QM as a primary adjustment or post-processing, these methods correct the distributional tails that are not fully captured by ML-based models alone, bringing the extreme rainfall magnitudes and natural variance back into close agreement with the observed distribution.

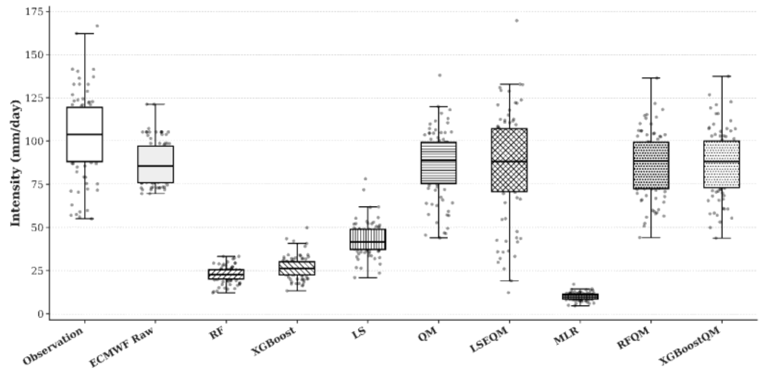
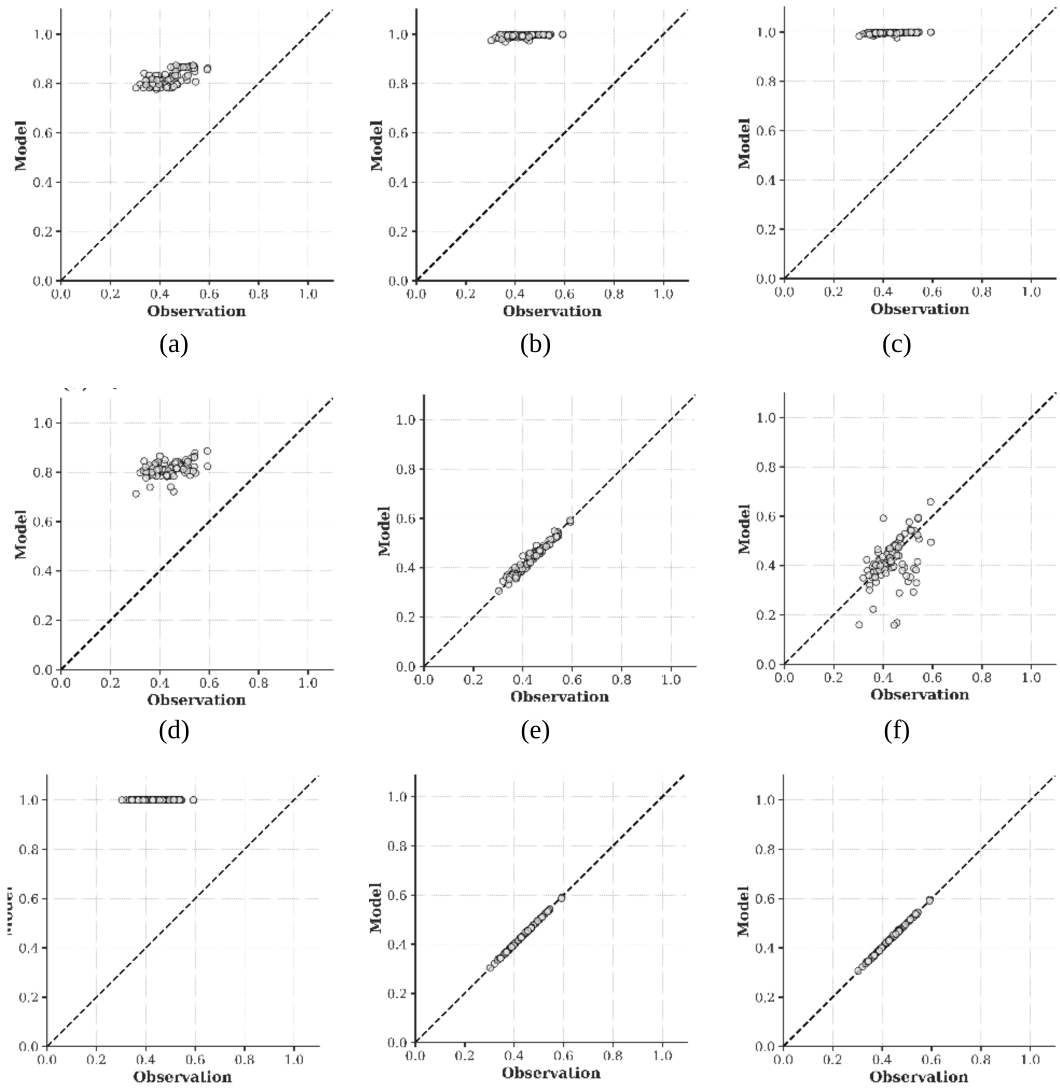


Figure 10. Distribution of Rx1day across bias correction methods

3.5. Analysis of model consistency per station and extreme quantiles

To thoroughly evaluate the methods, Figures 11 through 12 present comparative point-to-point and distributional analyses, where each specific method is visualized in a dedicated subfigure from Figures 11(a) to (i) and Figures 12(a) to (i): (a) ECMWF raw, (b) RF, (c) XGBoost, (d) LS, (e) QM, (f) LSEQM, (g) MLR, (h) RFQM, and (i) XGBoostQM.

The point-to-point consistency analysis of wet-day frequencies (Figure 11) demonstrates a clear performance polarization. The ECMWF raw (Figure 11(a)) and LS (Figure 11(d)) models exhibit a consistent wet bias. This widespread overestimation corresponds with the intrinsic drizzle bias of global models recorded by [32]. Notably, the standalone ML-based models RF (Figure 11(b)) and XGBoost (Figure 11(c)) alongside MLR (Figure 11(g)), yield a horizontal line at a frequency near 1.0, indicating a total lack of dry-day variability. This signifies that deterministic ML-based models, when devoid of suitable thresholding algorithms, suffer from a severe 'drizzle effect'. They are unable to distinguish between 'rain' and 'no-rain' conditions in a tropical dataset characterized by zero values. Unlike the preceding configurations, QM (Figure 11(e)), LSEQM (Figure 11(f)), and the hybrid frameworks containing RFQM (Figure 11(h)) alongside XGBoostQM (Figure 11(i)) exhibit high fidelity adjustments. Their corresponding data points cluster tightly around the 1:1 identity line. Such pronounced alignment confirms that these specific methodologies effectively simulate localized precipitation intermittency.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 11. Comparing scatter plots of wet-day frequency: (a) ECMWF raw, (b) RF, (c) XGBoost, (d) LS, (e) QM, (f) LSEQM, (g) MLR, (h) RFQM, and (i) XGBoostQM

Evaluating extreme value consistency through the annual maximum rainfall (Rx1day) uncovers a profound contrast in predictive accuracy throughout the multi panel visualizations presented in Figure 12. Scatter plot examinations confirm that the raw ECMWF (Figure 12(a)) and LS (Figure 12(d)) frameworks routinely underestimate severe precipitation occurrences, proving that conventional scaling modifiers remain inadequate for reproducing upper tail magnitudes [36]. Furthermore, standalone ML models comprising RF (Figure 12(b)) and XGBoost (Figure 12(c)) along with MLR (Figure 12(g)) completely miss peak rainfall intensities. Their predictions collapse into horizontal clusters restricted between 20 mm and 40 mm per day. This systemic negative bias originates directly from the variance smoothing properties embedded within MSE objective functions, a mathematical optimization process that lowers global residuals while simultaneously dampening extreme distributional values [36]. On the other hand, the results for QM (Figure 12(e)), LSEQM (Figure 12(f)), and our hybrid frameworks, RFQM (Figure 12(h)) and XGBoostQM (Figure 12(i)), sit symmetrically along the 1:1 diagonal. This alignment shows they have successfully overcome the underestimation biases that affect the other models. Unlike standalone tree-based algorithms that inherently optimize towards the conditional mean and compress data variance, the hybrid ML-QM framework successfully recovers the necessary ensemble spread. This recovery occurs because the secondary QM step explicitly forces the compressed ML outputs to align with the empirical CDF of the historical observations, mathematically ensuring that the upper-tail magnitudes of extreme convective events remain intact.

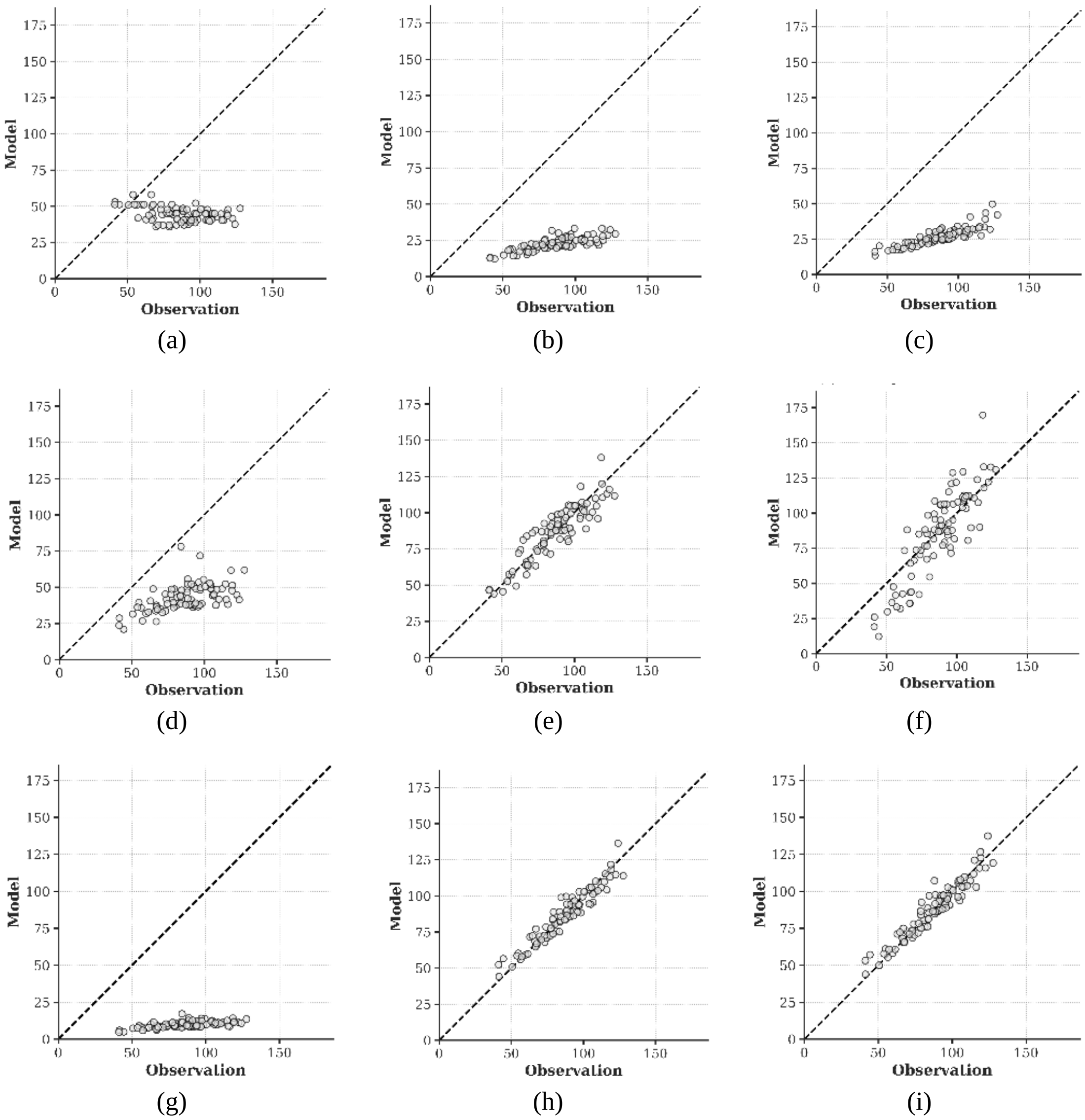


Figure 12. Comparing scatter plots of annual maximum rainfall (Rx1day) between methods: (a) ECMWF raw, (b) RF, (c) XGBoost, (d) LS, (e) QM, (f) LSEQM, (g) MLR, (h) RFQM, and (i) XGBoostQM

While scatter plots are highly effective for evaluating exact point-to-point temporal accuracy, they are notoriously susceptible to the ‘double penalty’ phenomenon. In highly convective tropical regions, a forecasting model might accurately predict the magnitude of an extreme storm but misplace it by a few kilometers or hours, leading to severe penalization in paired scatter analyses. Therefore, inter-station quantile-quantile (Q-Q) plots (Figures 13(a) to (i)) are critical for assessing the models’ capacity to preserve the physical probability distribution of historical storms, independent of exact spatiotemporal matches. The Q-Q plot analysis corroborates the previous findings: the curves for RF (Figure 13(b)), XGBoost (Figure 13(c)), and MLR (Figure 13(g)) plateau prematurely, failing entirely to depict the upper tail of the extreme distribution. In sharp contrast, the QM, LSEQM, and hybrid configurations presented in Figures 13(h) and 13(i) produce outputs that converge strictly upon the 1:1 identity line. Such uniform alignment across the entire observational spectrum proves that these specific frameworks successfully sustain high fidelity predictions during both low magnitude rainfall and severe high intensity precipitation events. The linearity at these upper quantiles confirms that the distribution-corrected data preserves the true probability of historical extreme storms, which is a crucial requirement for reliable return-period analysis in flood forecasting, as emphasized by [26].

Standalone QM and LSEQM might keep the overall distribution accurate, but they often miss the mark on timing a flaw highlighted by their lower daily KGE scores in Table 3. By successfully merging the improved daily temporal sequencing of ML-based models (reflected by the higher daily KGE in Table 3) with the robust probability reconstruction from Figure 13, the hybrid methodologies (RFQM and XGBoostQM) are explicitly recommended for forecasting short-term extreme rainfall and assessing localized flash flood risks. In such critical disaster risk reduction applications, capturing the precise magnitude of the intensity peak is far more vital than preserving long-term volume conservation.

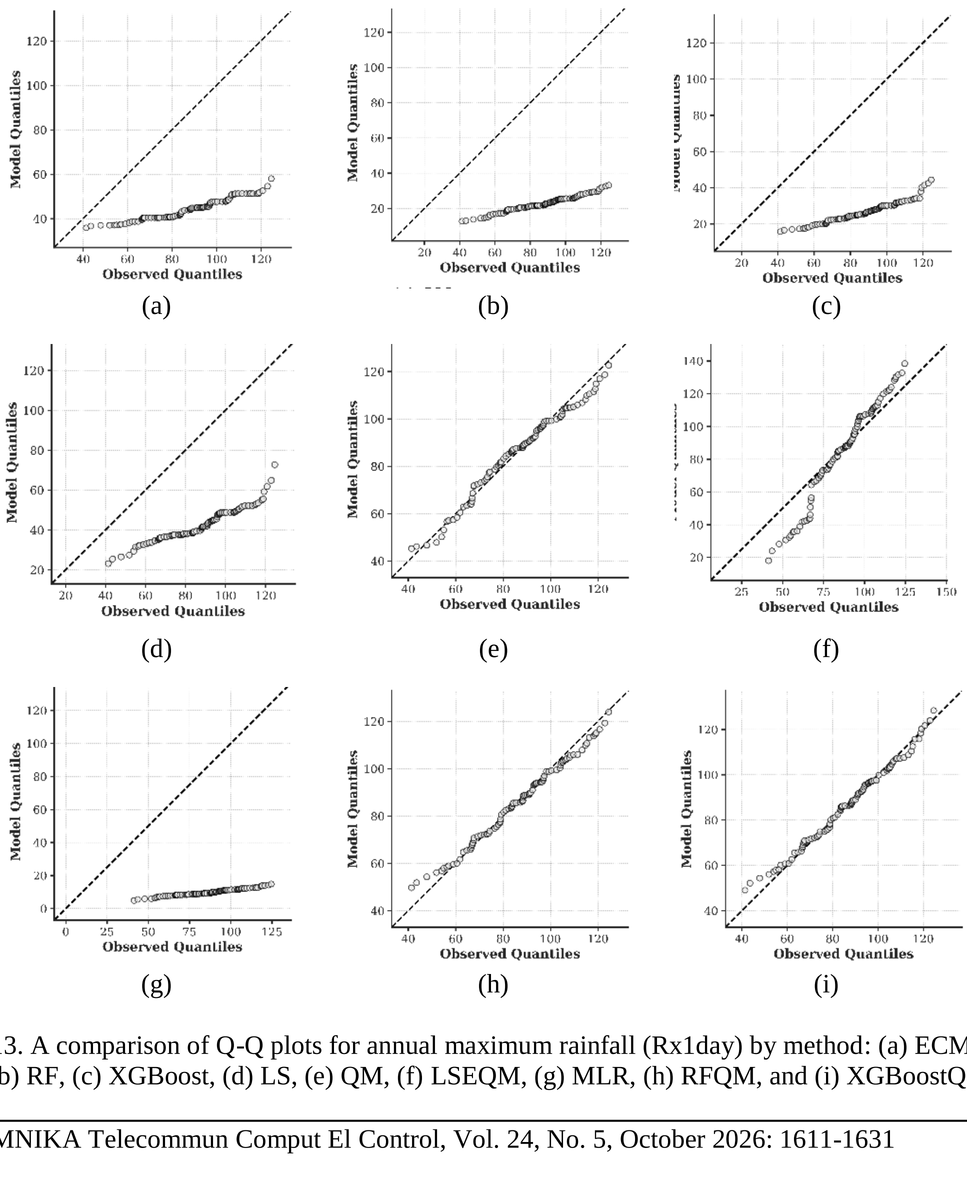


Figure 13. A comparison of Q-Q plots for annual maximum rainfall (Rx1day) by method: (a) ECMWF raw, (b) RF, (c) XGBoost, (d) LS, (e) QM, (f) LSEQM, (g) MLR, (h) RFQM, and (i) XGBoostQM

3.6. Spatial validation of QM model performance

The spatial distribution of performance enhancement (delta KGE), as illustrated in Figure 14, confirms the geographical reliability of the standalone QM approach. Diverging from the pronounced spatial instability observed within the LSEQM outputs, the standalone QM framework sustained uniform and robust bias reduction throughout the heterogeneous terrain of the Batanghari River Basin. This precise correction applies equally from the low elevation eastern floodplains up to the mountainous western highlands. Such spatial reliability proves that the QM methodology effectively resolves precipitation gradients driven by localized topography, directly solving a systemic regional modeling obstacle widely documented in existing literature [36]. The persistent stability of this method across various statistical metrics confirms that QM functions as both a mathematically rigorous architecture and a highly practical baseline for hydrometeorological evaluations within intricate tropical watersheds.

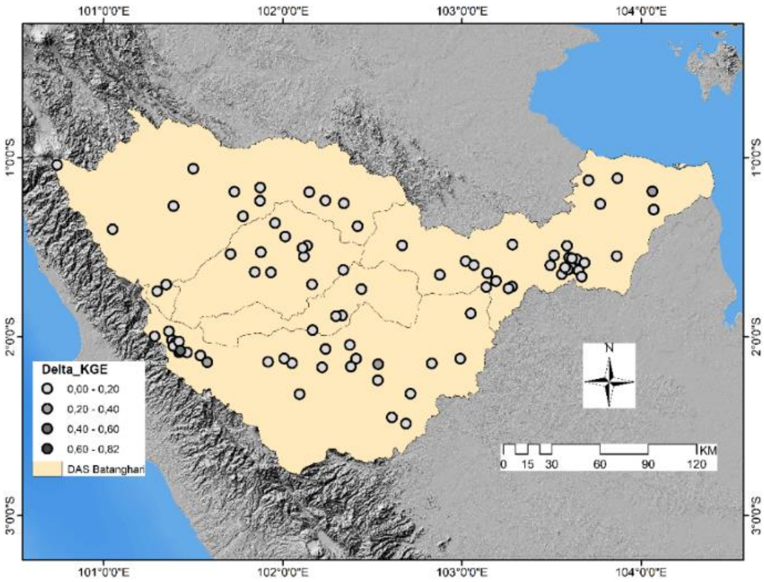


Figure 14. Differential KGE scores (QM vs. ECMWF Raw) in the Batanghari River Basin

The collective results establish a core scientific consensus that the operational success of a bias correction architecture depends strictly upon the distinct statistical and physical properties governing the local rainfall system. Whether focusing on mean states, temporal shifts, or extremes, no single method reigns supreme. For operational use, the key is to match the tool, whether it's a statistical baseline or a hybrid model, to the specific physical demands and objectives of the study area. For macro-scale water resource management, agro-meteorological planning, and long-duration fluvial flood forecasting within expansive basins, preserving the long-term regional mass balance is of primary importance. In these contexts, standalone QM provides a computationally streamlined and robust baseline, owing to its capacity to sustain monthly volumetric aggregates, as evidenced by its positive monthly KGE metrics. Conversely, for localized hazard mitigation and disaster risk reduction concerning convective flash floods, the extreme quantile analysis demonstrates that hybrid frameworks (RFQM and XGBoostQM) are structurally advantageous. By successfully preserving the natural ensemble spread and accurately reconstructing the extreme upper tails of the daily rainfall distribution (Rx1day), these hybrid methods provide the indispensable reliability required for assessing sudden, high-impact flood hazards, despite their limitations in long-term volume conservation.

4. CONCLUSION

This research evaluated the effectiveness of statistical, ML-based, and hybrid bias correction methods for enhancing ECMWF daily rainfall estimates across the Batanghari River Basin. Initial assessments confirmed that the raw ECMWF data suffered from an inherent drizzle anomaly and ensemble underdispersion. Resolving these baseline limitations, the systematic application of bias correction successfully elevated the overall predictive reliability. The final results expose observable operational compromises connecting computational demand, mass balance preservation, and extreme value representation.

The findings reveal that standalone ML-based models including RF and XGBoost, while adept at mapping non-linear atmospheric relationships, experienced variance deflation and pronounced spread collapse. As a result, they systematically underestimated extreme rainfall magnitudes when deployed without distributional post processing. Shifting to statistical methods, standalone QM delivered stable spatial mass balance correction, mitigated the drizzle anomaly, and achieved highest monthly temporal stability (KGE) alongside modest computational requirements. Advancing the analysis shows that hybrid ML frameworks encompassing RFQM and XGBoostQM successfully addressed the ensemble variance deflation problem. By integrating the temporal phase precision of ML algorithms with the distributional adjustments of QM, these configurations recovered the necessary ensemble spread and maintained upper tail precipitation dynamics. Despite these architectural enhancements, the models present a distinct operational cost. While generating improved accuracy for intense storm simulations, they demand increased computational resources and display weakened long term mass balance conservation, a limitation reflected in their negative monthly KGE metrics.

These collective observations validate the importance of matching bias correction methodologies with specific operational targets and catchment hydroclimatology. Standalone QM emerges as a practical and dependable baseline for continuous hydrological simulations, agrometeorological water management, and large basin fluvial flood forecasting, largely due to its consistent preservation of monthly volumetric aggregates. Despite their compromises in long term volume consistency, hybrid frameworks such as RFQM and XGBoostQM provide the targeted extreme value fidelity required for essential disaster risk reduction applications. These specific applications include resolving short duration convective intensities, mapping localized flash flood hazards, and estimating return periods for intense storm events.

Future research should extend this framework using deep learning architectures, such as convolutional neural networks (CNNs) and LSTM networks, to better capture non-linear spatiotemporal precipitation dynamics. Subsequent iterations must also incorporate multivariate bias correction to preserve the physical covariance among interconnected hydrometeorological variables, including temperature, humidity, and wind fields, without compromising long term mass balance constraints. Finally, deploying this hybrid architecture within a cloud computing environment can facilitate functional, real time localized flood forecasting.

ACKNOWLEDGMENTS

The authors express gratitude to the Agency for Meteorology, Climatology, and Geophysics (BMKG) for its financial assistance and provision of rainfall statistics. Gratitude is also extended to the Department of Geophysics and Meteorology (GFM) of IPB University for their oversight of the research throughout its progression.

FUNDING INFORMATION

This research was funded by the the Agency for Meteorology, Climatology, and Geophysics (BMKG) of Indonesia.

AUTHOR CONTRIBUTIONS STATEMENT

This journal utilizes the Contributor Roles Taxonomy (CRediT) to attribute credit to each author, mitigate authorship disputes, and facilitate collaboration among writers.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Arif Ma'rufi	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
Tania June	✓	✓		✓	✓	✓			✓	✓	✓	✓	✓	
I Putu Santikayasa	✓			✓	✓	✓			✓	✓		✓		
Apip	✓	✓		✓	✓	✓			✓	✓	✓	✓		
Siswanto	✓	✓		✓	✓	✓			✓	✓	✓	✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

DATA AVAILABILITY

The datasets produced and/or examined in the present work can be obtained from the first author upon reasonable request.

REFERENCES

[1] F. Wang, D. Tian, and M. Carroll, "Customized deep learning for precipitation bias correction and downscaling," *Geoscientific Model Development*, vol. 16, no. 2, pp. 535–556, Jan. 2023, doi: 10.5194/gmd-16-535-2023.

[2] T. Zhang, Z. Liang, W. Li, J. Wang, Y. Hu, and B. Li, "Statistical post-processing of precipitation forecasts using circulation classifications and spatiotemporal deep neural networks," *Hydrology and Earth System Sciences*, vol. 27, no. 10, pp. 1945–1960, May 2023, doi: 10.5194/hess-27-1945-2023.

[3] G. Y. Seo and J. B. Ahn, "Comparison of bias correction methods for summertime daily rainfall in South Korea using quantile mapping and machine learning model," *Atmosphere*, vol. 14, no. 7, p. 1057, Jun. 2023, doi: 10.3390/atmos14071057.

[4] A. Badrinath, L. Delle Monache, N. Hayatbini, W. Chapman, F. Cannon, and M. Ralph, "Improving precipitation forecasts with convolutional neural networks," *Weather and Forecasting*, Feb. 2023, doi: 10.1175/WAF-D-22-0002.1.

[5] L. Han *et al.*, "A deep learning method for bias correction of ECMWF 24–240 h forecasts," *Advances in Atmospheric Sciences*, vol. 38, no. 9, pp. 1444–1459, Sep. 2021, doi: 10.1007/s00376-021-0215-y.

[6] M. Aich, P. Hess, B. Pan, S. Bathiany, Y. Huang, and N. Boers, "Conditional diffusion models for downscaling and bias correction of Earth system model precipitation," 2024, *arXiv*. doi: 10.48550/ARXIV.2404.14416.

[7] P. Hess and N. Boers, "Deep learning for improving numerical weather prediction of heavy rainfall," *Journal of Advances in Modeling Earth Systems*, vol. 14, no. 3, Mar. 2022, doi: 10.1029/2021MS002765.

[8] H. Li, Y. Zhang, H. Lei, and X. Hao, "Machine learning-based bias correction of precipitation measurements at high altitude," *Remote Sensing*, vol. 15, no. 8, p. 2180, Apr. 2023, doi: 10.3390/rs15082180.

[9] C. R. Rajulapati and S. M. Papalexiou, "Precipitation bias correction: a novel semi-parametric quantile mapping method," *Earth and Space Science*, vol. 10, no. 4, Apr. 2023, doi: 10.1029/2023EA002823.

[10] G. Al-Rawas, M. R. Nikoo, N. Sadra, and F. Mousavi, "Integrated quantile mapping and spatial clustering for robust bias correction of satellite precipitation in data-sparse regions," *Sustainability (Switzerland)*, vol. 17, no. 18, p. 8321, Sep. 2025, doi: 10.3390/su17188321.

[11] A. T. Tadase and A. K. Tekile, "XGBoost-based bias correction of CMIP6 precipitation in the Wabi Shebelle Basin, Ethiopia: a comparative study with empirical quantile mapping," *Journal of Water and Climate Change*, vol. 17, no. 3, pp. 683–700, Mar. 2026, doi: 10.2166/wcc.2026.103.

[12] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[13] T. Chen and C. Guestrin, "XGBoost," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[14] S. C. Mishra Sharma and A. Mitra, "Deep learning based bias correction of TRMM precipitation estimates using IMD-gridded precipitation as ground observation," *Environmental Data Science*, vol. 3, p. e41, Jan. 2024, doi: 10.1017/eds.2024.39.

[15] S. Rasp and S. Lerch, "Neural networks for postprocessing ensemble weather forecasts," *Monthly Weather Review*, vol. 146, no. 11, pp. 3885–3900, Nov. 2018, doi: 10.1175/MWR-D-18-0187.1.

[16] C. K. Sønderby *et al.*, "MetNet: A Neural Weather Model for Precipitation Forecasting," 2020, *arXiv*. doi: 10.48550/ARXIV.2003.12140.

[17] S. Ravuri *et al.*, "Skilful precipitation nowcasting using deep generative models of radar," *Nature*, vol. 597, no. 7878, pp. 672–677, Sep. 2021, doi: 10.1038/s41586-021-03854-z.

[18] M. Reichstein *et al.*, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, Feb. 2019, doi: 10.1038/s41586-019-0912-1.

[19] L. Rahimi *et al.*, "Future climate projections for south florida: improving the accuracy of air temperature and precipitation extremes with a hybrid statistical bias correction technique," *Earth's Future*, vol. 12, no. 8, Aug. 2024, doi: 10.1029/2024EF004531.

[20] M.-J. Um, H. Kim, and J.-H. Heo, "Hybrid approach in statistical bias correction of projected precipitation for the frequency analysis of extreme events," *Advances in Water Resources*, vol. 94, pp. 278–290, Aug. 2016, doi: 10.1016/j.advwatres.2016.05.021.

[21] C. Passow and R. V. Donner, "Regression-based distribution mapping for bias correction of climate model outputs using linear quantile regression," *Stochastic Environmental Research and Risk Assessment*, vol. 34, no. 1, pp. 87–102, Jan. 2020, doi: 10.1007/s00477-019-01750-7.

[22] Siswanto *et al.*, "A very unusual precipitation event associated with the 2015 floods in Jakarta: an analysis of the meteorological factors," *Weather and Climate Extremes*, vol. 16, pp. 23–28, Jun. 2017, doi: 10.1016/j.wace.2017.03.003.

[23] Y. H. Song, E.-S. Chung, and M. S. Shiru, "Uncertainty analysis of monthly precipitation in GCMs using multiple bias correction methods under different RCPs," *Sustainability*, vol. 12, no. 18, p. 7508, Sep. 2020, doi: 10.3390/su12187508.

[24] D. W. Pierce, D. R. Cayan, E. P. Maurer, J. T. Abatzoglou, and K. C. Hegewisch, "Improved bias correction techniques for hydrological simulations of climate change*," *Journal of Hydrometeorology*, vol. 16, no. 6, pp. 2421–2442, Dec. 2015, doi: 10.1175/JHM-D-14-0236.1.

[25] B. Ayugi *et al.*, "Quantile mapping bias correction on rossby centre regional climate models for precipitation analysis over Kenya, East Africa," *Water*, vol. 12, no. 3, p. 801, Mar. 2020, doi: 10.3390/w12030801.

[26] A. J. Cannon, S. R. Sobie, and T. Q. Murdock, "Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes?," *Journal of Climate*, vol. 28, no. 17, pp. 6938–6959, Sep. 2015, doi: 10.1175/JCLI-D-14-00754.1.

[27] C. Volosciuk, D. Maraun, M. Vrac, and M. Widmann, "A combined statistical bias correction and stochastic downscaling method for precipitation," *Hydrology and Earth System Sciences*, vol. 21, no. 3, pp. 1693–1719, Mar. 2017, doi: 10.5194/hess-21-1693-2017.

[28] Y. Zhang *et al.*, "Comparing quantile regression forest and mixture density long short-term memory models for probabilistic post-processing of satellite precipitation-driven streamflow simulations," *Hydrology and Earth System Sciences*, vol. 27, no. 24, pp. 4529–4550, Dec. 2023, doi: 10.5194/hess-27-4529-2023.

[29] X. Wenwei *et al.*, "Deep learning experiments for tropical cyclone intensity forecasts," *Weather and Forecasting*, Jun. 2021, doi: 10.1175/waf-d-20-0104.1.

[30] S. J. Lee, E. Sohn, M. Kim, K. H. Park, K. Park, and Y. Lee, "Real-time retrieval of daily soil moisture using IMERG and GK2A satellite images with NWP and topographic data: a machine learning approach for South Korea," *Remote Sensing*, vol. 15, no. 17, p. 4168, Aug. 2023, doi: 10.3390/rs15174168.

[31] W. J. M. Knoben, J. E. Freer, and R. A. Woods, "Technical note: inherent benchmark or not? comparing Nash-Sutcliffe and Kling-Gupta efficiency scores," *Hydrology and Earth System Sciences*, vol. 23, no. 10, pp. 4323–4331, Oct. 2019, doi: 10.5194/hess-23-4323-2019.

[32] H. Chen, B. Yong, Y. Shen, J. Liu, Y. Hong, and J. Zhang, "Comparison analysis of six purely satellite-derived global precipitation estimates," *Journal of Hydrology*, vol. 581, p. 124376, Feb. 2020, doi: 10.1016/j.jhydrol.2019.124376.

[33] H. Hersbach *et al.*, "The ERA5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, Jul. 2020, doi: 10.1002/qj.3803.

[34] W. M. Ridwan, M. Sapitang, A. Aziz, K. F. Kushiar, A. N. Ahmed, and A. El-Shafie, "Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia," *Ain Shams Engineering Journal*, vol. 12, no. 2, pp. 1651–1663, Jun. 2021, doi: 10.1016/j.asej.2020.09.011.

[35] J. Baño-Medina, R. Manzanas, and J. M. Gutiérrez, "Configuration and intercomparison of deep learning neural models for statistical downscaling," *Geoscientific Model Development*, vol. 13, no. 4, pp. 2109–2124, Apr. 2020, doi: 10.5194/gmd-13-2109-2020.

[36] G. Worku, E. Teferi, A. Bantider, and Y. T. Dile, "Statistical bias correction of regional climate model simulations for climate change projection in the Jemma sub-basin, upper Blue Nile Basin of Ethiopia," *Theoretical and Applied Climatology*, vol. 139, no. 3–4, pp. 1569–1588, Feb. 2020, doi: 10.1007/s00704-019-03053-x.

[37] S. Vannitsem *et al.*, "Statistical postprocessing for weather forecasts review, challenges, and avenues in a big data world," *Bulletin of the American Meteorological Society*, vol. 102, no. 3, pp. E681–E699, Mar. 2021, doi: 10.1175/BAMS-D-19-0308.1.

[38] C. Navarro-Racines, J. Tarapues, P. Thornton, A. Jarvis, and J. Ramirez-Villegas, "High-resolution and bias-corrected CMIP5 projections for climate change impact assessments," *Scientific Data*, vol. 7, no. 1, p. 7, Jan. 2020, doi: 10.1038/s41597-019-0343-8.

[39] D. N. Moriasi, M. W. Gitau, N. Pai, and P. Daggupati, "Hydrologic and water quality models: performance measures and evaluation criteria," *Transactions of the ASABE*, vol. 58, no. 6, pp. 1763–1785, Dec. 2015, doi: 10.13031/trans.58.10715.

BIOGRAPHIES OF AUTHORS

Arif Ma'rufi is pursuing a Ph.D. in Applied Climatology at the Department of Geophysics and Meteorology, IPB University, Indonesia. He holds a first Diploma in Meteorology from the Academy of Meteorology and Geophysics (AMG), Jakarta, in 1996; a Bachelor's degree in Soil Science Study Program, Faculty of Agriculture, from Jambi University in 2002; and a Master's degree in Earth Science, Faculty of Earth Sciences and Technology, from Bandung Institute of Technology (ITB) in 2014. From 1997 until now, he has been working at BMKG, with previous assignments at the Jambi Climatological Station. He can be contacted at email: arif.marufi@bmkg.go.id, marufiarif@apps.ipb.ac.id.

Tania June is a Professor at the Department of Geophysics and Meteorology (GFM), IPB University (Bogor Agricultural University), Indonesia. She has been a faculty member since 1988 and currently serves as the Head of the Doctoral Program in Applied Climatology. She also leads the agrometeorology division in the department. She earned her Bachelor's degree in Agrometeorology from IPB University, her Master's degree in Agricultural science from the University of Queensland, Australia, and her Doctoral degree (Ph.D.) in Plant Sciences/Environmental Biology from the Australian National University, Australia. She is a senior researcher in the Collaborative Research Centre (CRC) 990 hosted by the University of Göttingen, Germany, which explores ecological and socioeconomic transitions in tropical Indonesia. She can be contacted at email: taniajune@apps.ipb.ac.id.

I Putu Santikayasa received a Ph.D. in Water Engineering and Management from IPB University, with the dissertation "sustainability of water use in irrigated agriculture under future climate scenarios: a case study of Citarum River basin in Indonesia." He also holds a Master of Science (M.Si.) in Information Technology for Natural Resources Management (2003–2006) and a Bachelor's degree (S.P.) in Agrometeorology (1997–2002), both from IPB University, Indonesia. He is currently a lecturer at the Department of Geophysics and Meteorology, Faculty of Mathematics and Natural Science, IPB University, Indonesia. His research interests are hydrology modeling, climate change, water resources management, and climate risk assessment. He can be contacted at email: ipsantika@apps.ipb.ac.id.

Apip is a principal researcher at the Research Center for Limnology and Water Resources (PRLSDA), National Research and Innovation Agency (BRIN), Indonesia, with extensive expertise in ecohydrology and surface hydrological modeling. He earned his Bachelor of Science degree from the Department of Geophysics and Meteorology (GFM), Faculty of Mathematics and Natural Sciences, IPB University, before completing both his Master’s and Doctoral degrees at Kyoto University, Japan, where he obtained his Ph.D. in 2013. He as the Head of the Watershed Hydrology and Inland Waters Management Research Group, directs strategic research on flood forecasting systems, peatland hydrological dynamics, soil erosion and sediment transport modeling, and the assessment of climate-land use change impacts, thereby making substantial contributions to hydro-meteorological disaster mitigation and sustainable water management in Indonesia. He can be contacted at email: apip001@brin.go.id.

Siswanto a senior climate scientist at the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG) obtained his Ph.D. in Water and Climate Risk from Vrije Universiteit Amsterdam, focusing on extreme climate, climate change, meteorology, and oceanography. His academic credentials include a Master’s degree in Climate and Environmental Physics (Cum Laude) from Universität Bern, Switzerland (2008–2010), a Bachelor’s degree in Physics from Universitas Indonesia (2004–2006), and a Diploma in Meteorology from the Academy of Meteorology and Geophysics (AMG), Indonesia (1997–2000). He currently holds the position of senior researcher at the Center for Applied Climate Services, BMKG, where he significantly contributes to research on climate variability and water-related disaster risks. He can be contacted at email: siswanto@bmkg.go.id.