

Article

Enhancing GEOGLOWS River Forecast System with a High-Resolution Pre-Processing Approach for Runoff Bias Correction

Juseth E. Chancay ^{1,2} , Jorge Luis Sánchez-Lozano ^{3,4,*} , Bryan G. Valencia ¹ , Mario Germán Trujillo-Vela ⁵ , E. James Nelson ⁴ , Riley C. Hales ⁴  and Angélica L. Gutiérrez ⁶

- ¹ Facultad de Ciencias de la Tierra y Agua, Universidad Regional Amazónica Ikiam, Km 7, Vía Muyuna, Tena 150150, Napo, Ecuador; juseth.chancay@est.ikiam.edu.ec (J.E.C.); bryan.valencia@ikiam.edu.ec (B.G.V.)
² Fundación Ecuatoriana de Estudios Ecológicos Ecociencia, Quito 170523, Pichincha, Ecuador
³ Aquaveo LLC., Provo, UT 84604, USA
⁴ Department of Civil and Construction Engineering, Brigham Young University, Provo, UT 84601, USA; jimn@byu.edu (E.J.N.); rchales@byu.edu (R.C.H.)
⁵ Facultad de Ingeniería, Universidad Surcolombiana, Neiva 410001, Colombia; mario.trujillo@usco.edu.co
⁶ National Oceanic and Atmospheric Administration (NOAA) (Ret.), Silver Springs, MD 20910, USA; angelica.gutierrez@geogloWS.org
* Correspondence: jsanchez@aquaveo.com; Tel.: +1-801-691-5531 (ext. 403)

Abstract

Accurate streamflow information is critical for early flood and drought warning. However, global hydrological forecasting systems are affected by residual errors in meteorological forcing, model structure, and routing, which propagate into simulated streamflow. Within the GEOGLOWS River Forecast System (RFS), ERA5 runoff biases are routed into streamflow simulations. The most effective operational bias-correction method, MFDC-QM, requires local discharge observations and cannot be applied consistently in ungauged basins. This study evaluates a pre-routing, grid-scale runoff bias-correction framework that adjusts ERA5 runoff before routing by combining Flow Duration Curve (FDC) mapping and Sparse Cumulative Distribution Function (CDF) matching, using GSCD as a spatially distributed reference runoff data. Baseline GEOGLOWS RFS, pre-routing correction, and MFDC-QM were compared for 1980–2025 using 16,517 gauging stations, Kling–Gupta Efficiency (KGE), and paired significance tests. Globally, the median KGE increased modestly from 0.16 to 0.22, compared with 0.48 for MFDC-QM. Results demonstrate a clear regional dependence: pre-routing correction produced statistically significant gains in South America and Africa ($p < 0.05$), where ERA5 runoff exhibits stronger residual biases, but had limited effects in Europe and North America, where dense hydrometeorological networks likely impose stronger observational constraints on the underlying reanalysis. These patterns show that pre-routing correction is most valuable where residual forcing bias is large and observational constraints are limited, complementing observation-based post-processing in ungauged, data-limited regions.

Keywords: GEOGLOWS RFS; runoff bias correction; ERA5 runoff; flow duration curve; global hydrological modeling; ungauged basins



Academic Editor: Alina Bărbulescu

Received: 12 February 2026

Revised: 2 May 2026

Accepted: 5 May 2026

Published: 10 May 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Accurate streamflow predictions and understanding their spatiotemporal dynamics are critical for anticipating hydrological extreme events, such as floods and droughts,

especially in low- and middle-income regions with limited hydrometeorological gauge networks [1–3]. This need is highlighted by the growing socioeconomic impacts of these hazards in recent years [4]. For instance, in 2019, floods and droughts affected 60.2 million people and caused USD 36.9 billion in economic losses [5]. By 2024, these numbers increased to 78.3 million people affected and USD 46.1 billion in economic losses [6]. Recent studies show that 89% of the population exposed to floods and droughts lives in low- and middle-income countries [7]. For instance, in South Asia, the 2022 Pakistan floods affected more than 33 million people [8], whereas in South America, the 2024 floods in Rio Grande do Sul, Brazil, affected 2.3 million people and displaced approximately 600,000 [9]. In Sub-Saharan Africa, drought remains a major driver of persistent food insecurity, particularly in pastoral regions [10].

Early warning systems (EWS) based on hydrological modelling have consolidated as practical tools to reduce the impacts of floods and droughts [11–13]. Their growing operational relevance has been enabled by advances in computational capacity, which now support large-domain simulations at increasingly fine resolutions [14,15]. Earth observation technologies have also influenced EWS based on hydrological modelling to provide near-real-time feedback, improving how we monitor and forecast hydrometeorological processes at the global scale [16]. The European Centre for Medium-Range Weather Forecasts (ECMWF) demonstrates this approach by assimilating satellite-based observations into its models using advanced data assimilation techniques [17].

By providing globally consistent atmospheric forcing, supported by satellite data assimilation, modelling systems such as Integrated Forecasting System (IFS) developed by ECMWF have facilitated the emergence of global-scale hydrological modelling frameworks, such as Global Flood Awareness System (GloFAS) [18] and GEOGLOWS River Forecast System (RFS) [19]. This reflects progress in hydrology with stronger integration with meteorological forcing and a better understanding of land-surface processes [20]. By coupling numerical weather prediction with terrestrial hydrological components, these integrated systems have strengthened our ability to anticipate hydrological extremes, while ensemble-based probabilistic approaches provide a practical way to represent uncertainty in meteorological forcing and model structure [21,22].

Despite these advances, translating global-scale hydrological predictions into actionable information at the local scale, where decisions are made, remains challenging [20,23]. Because these hydrological modelling systems operate globally, discharge-based calibration is not feasible and rarely provides parameterizations that remain robust across contrasting hydroclimatic conditions [24,25]. Consequently, residual errors in meteorological forcing, model structure, and routing propagate into streamflow simulations [26], leading to biased magnitudes and a systematic overestimation of peak flows and underestimation of base flows [27,28]. In operational contexts, these biases can directly affect decision-making by shifting the timing, magnitude, and frequency of threshold exceedances used for flood warnings and drought monitoring. Bias reduction can therefore improve early warning reliability, reduce false alarms, and support timely risk-management decisions [29]. In this regard, bridging this global-to-local gap requires approaches that explicitly reduce bias and improve reliability at operational local scales.

Bias-correction methods have become a widely used strategy to improve the reliability of global model predictions and make them more suitable for decision support [29]. Most existing approaches are implemented as post-processing steps that adjust simulated streamflow using available observations, commonly through distribution- or frequency-matching techniques such as Flow Duration Curve (FDC) mapping, where simulated values are replaced by observed-consistent counterparts at the same exceedance probability [30,31]. Sánchez Lozano et al. [32] extended this concept by introducing Monthly Flow Duration

Curve Quantile Mapping (MFDC-QM). This seasonality-aware approach corrects systematic biases in the GEOGloWS River Forecast System at locations where historical observed streamflow records are available. However, this dependence on local streamflow records limits the use of MFDC-QM as a stand-alone global operational solution, because several river reaches remain ungauged or lack sufficiently long and reliable historical observations.

Hales et al. [33] further advanced the FDC-based approach by developing SABER, a machine learning-based method that finds similarity relationships among basins (e.g., via k-means clustering and network connectivity aware decision tree) to transfer FDC-based correction factors from gauged to ungauged locations, thereby extending bias correction beyond sites with observed streamflow records. However, SABER still faces practical constraints for global operational use considering (i) its correction skill depends on the availability and quality of observed discharge records which cannot be ensured everywhere, and (ii) its transfer step remains uncertain given the selection of “donor” basins is sensitive to the similarity criteria and available attributes, and truly comparable basins are often scarce in practice [34–36].

In addition to streamflow bias-correction methods, Lin et al. [37] explored a pre-routing approach that applies Sparse Cumulative Distribution Function (CDF) matching to bias-correct runoff before routing, demonstrating its effectiveness in improving large-scale hydrological streamflow reconstructions. This approach differs from MFDC-QM and SABER in both its correction target and operational assumptions: rather than adjusting routed streamflow using local or transferred discharge information, it modifies the gridded runoff forcing that propagates through the river network. However, the framework proposed by Lin et al. [37] was developed and evaluated within offline reconstruction models and was not designed to assess the performance of operational river forecast systems. Therefore, its added value for systems such as GEOGLOWS RFS, where scalability, spatial consistency, and forecast-oriented implementation are central requirements, remains unexplored.

Building on this gap, the present study introduces and evaluates an operationally oriented pre-routing runoff bias-correction framework for global hydrological forecasting. In contrast to post-processing methods applied to routed streamflow, the proposed approach corrects ERA5 hydrological forcing at the grid scale prior to routing. This is achieved by leveraging the Global Streamflow Characteristics Dataset (GSCD) to adjust runoff distributions through Flow Duration Curve (FDC) mapping in combination with sparse Cumulative Distribution Function (CDF) matching. The framework is implemented within the GEOGLOWS River Forecast System (RFS) and its added value is assessed relative to the baseline configuration and the MFDC-QM post-processing approach, utilizing Kling–Gupta Efficiency (KGE), CDF-based diagnostics, and paired significance testing. The study specifically investigates whether pre-routing runoff correction enhances simulated streamflow, examines regional and seasonal variability in its effectiveness, and compares its performance with MFDC-QM in locations where discharge observations are available. It is hypothesized that the correction will yield the greatest benefits in data-scarce regions characterized by pronounced residual ERA5 runoff biases, while offering limited improvements in well-instrumented regions where reanalysis runoff is more strongly constrained by observations. By addressing forcing errors before they propagate through the river network, the framework offers a scalable, spatially consistent strategy that complements observation-based post-processing and supports operational flood and drought monitoring in both gauged and ungauged basins.

2. Data and Methods

2.1. Data

To implement the proposed pre-routing runoff bias-correction and quantify the performance of the resulting simulations, we used three datasets described as follows.

2.1.1. Ground-Observed Streamflow Data

Daily streamflow observations were obtained from >16,000 hydrological stations worldwide (Figure 1), belonging to National Hydrometeorological Institutions and compiled from open access hydrological repositories, such as the Global Runoff Data Centre (GRDC). Stations with at least 12 months of data within the 1980–2025 period were selected, a criterion established by Sánchez Lozano [32] for bias correction.

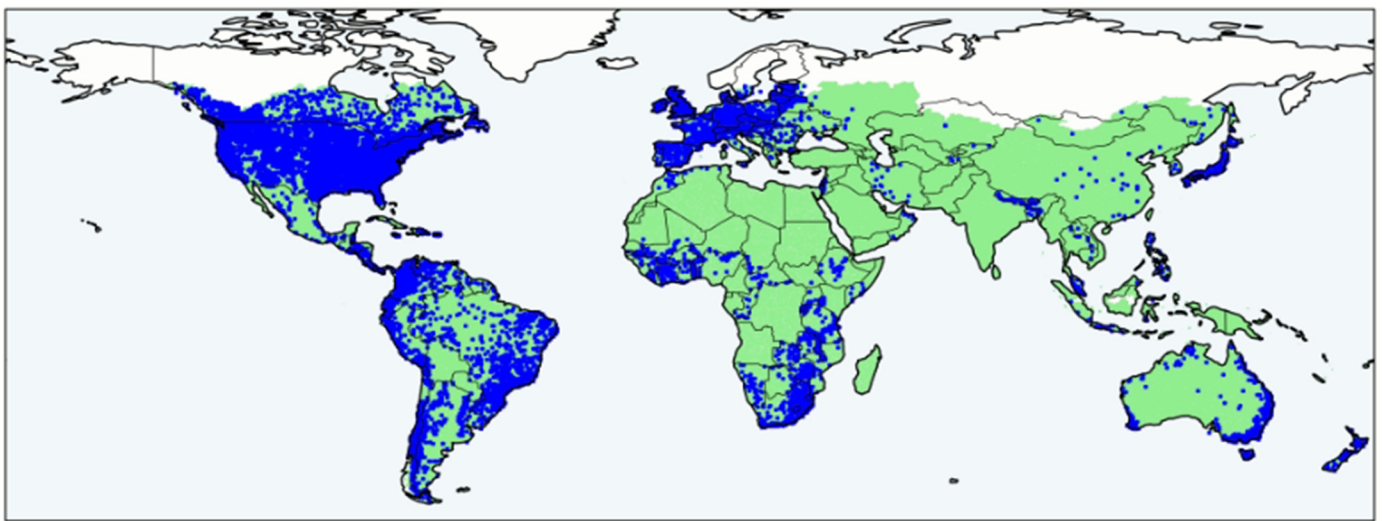


Figure 1. In situ streamflow gauges used in this study (blue dots) and spatial coverage of GEOGLOWS River Forecast System (green).

2.1.2. Gridded Runoff Data

Gridded runoff forcing was extracted from the Copernicus Climate Data Store product “ERA5 hourly data on single levels from 1940 to present”, the ECMWF fifth-generation global reanalysis (ERA5) [38]. It provides runoff data on a regular 0.25° grid, generated by combining model physics with worldwide observations and satellite-based information through data assimilation [17]. The dataset is operationally updated with a latency of about 5 days, enabling consistent long-term analyses and near-real-time applications. In this study, ERA5 runoff data were used from 1979 onwards and aggregated to daily resolution, consistent with the operational configuration of GEOGLOWS RFS Version 1, which provides daily retrospective simulations beginning on 1 January 1979 [33].

2.1.3. Global Streamflow Characteristics Dataset

The Global Streamflow Characteristics Dataset (GSCD), developed by Beck et al. [39], is a Deep Learning-based repository of global maps that describe streamflow characteristics, compiled from observations of thousands of basins obtained from repositories such as GRDC, USGS/GAUGES II, and networks in Australia. Flow duration curve percentiles are derived from these series and then spatialized by using a neural network ensemble trained based on climatic and physiographic variables derived from satellite-based data [39]. The product is delivered at global scale on a regular 0.05° grid and serves as a reference for characterizing ungauged river basins. In this study, we use the GSCD gridded estimates of

the 1st, 5th, 10th, 20th, 50th, 80th, 90th, 95th, and 99th percentiles of daily flow, expressed as runoff equivalents (mm/day).

2.2. GEOGLOWS River Forecast System

The GEOGLOWS River Forecast System (RFS) is an operational, globally scalable streamflow prediction framework that delivers both retrospective and forecast simulations. It was developed by the Group on Earth Observation for Global Water Sustainability (GEOGLOWS), in collaboration with ECMWF, NASA, NOAA, and Brigham Young University (BYU). GEOGLOWS RFS provides accessible hydrological data to support climate adaptation, disaster risk reduction, and water resources management in data-scarce regions [19,40].

Currently, GEOGLOWS RFS operates with two drainage network configurations and model structures. Version 1 employs a HydroSHEDS-derived drainage network comprising approximately one million river reaches [41], partitioned into 13 computational watershed groups. Version 2 uses the higher-resolution TDX-Hydro network with about seven million river reaches [42], partitioned into 125 computational watershed groups. In both versions, GEOGLOWS RFS generates streamflow simulations by (i) combining gridded runoff with the drainage network and basin delineations to compute water volumes (i.e., effective cell area multiplied by runoff depth), and (ii) routing these volumes using the Muskingum method (Figure 2).

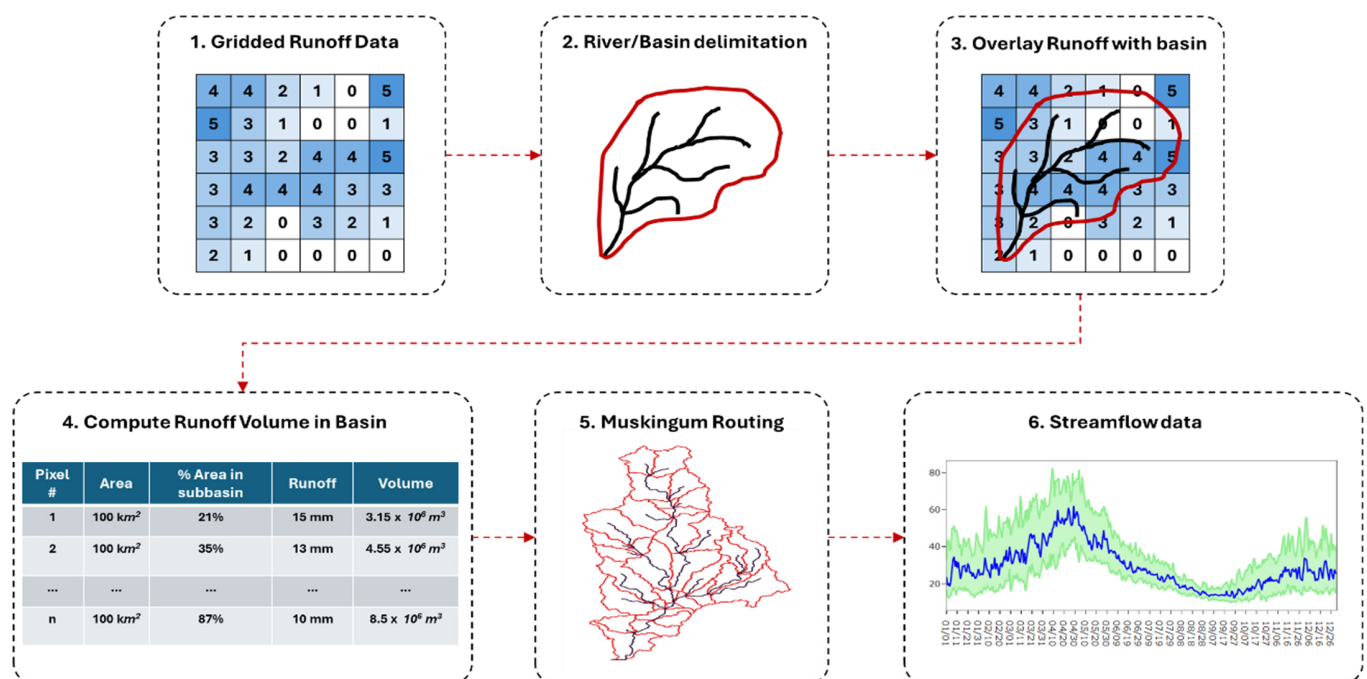


Figure 2. Process and modeling structure of the GEOGLOWS River Forecast System. Modified from Hales et al. [33].

In version 1, Muskingum routing is implemented in the Routing Application for Parallel computation of Discharge (RAPID), with the Muskingum parameters (K, X) derived from reach-scale morphometric characteristics under the assumption of a constant average wave celerity of 0.8 m/s [29,43]. By contrast, in version 2, the river-route Python package is used to perform Muskingum routing, with parameters (K, X) estimated from regression relationships that predict average wave celerity as a function of the natural logarithm of upstream drainage area [19].

In this study, we focused on GEOGLOWS RFS version 1. Although GEOGLOWS RFS version 2 provides a substantially higher-resolution river network and updated routing parameters, previous work has shown that vector-based river routing models such as RAPID generate largely consistent streamflow simulations across different watershed resolutions when forced with the same gridded runoff inputs [44]. Because the runoff bias-correction approach proposed here operates entirely on the gridded hydrological forcing prior to routing, its formulation is not inherently sensitive to river network resolution [44]. Consequently, while differences between GEOGLOWS RFS versions may arise from updated digital elevation models and routing parameterizations, the proposed pre-routing runoff bias-correction framework is expected to be methodologically transferable to GEOGLOWS RFS version 2 and to other hydrological models driven by gridded runoff data. In addition, running the full retrospective simulation and bias-correction workflow using the resolution of GEOGLOWS RFS version 2 exceeds our available computational and storage resources.

GEOGLOWS RFS Version 1 provides daily retrospective simulations beginning on 1 January 1979, with all stream flows initialized at $0 \text{ m}^3/\text{s}$. As a result, early simulation values do not accurately represent true initial conditions, particularly in larger rivers, and the year 1979 effectively serves as a model warm-up period, as mentioned by Sanchez Lozano et al. [32] and Rojas-Lesmes et al. [45]. In this regard, the performance evaluation in this study was therefore conducted over simulations starting on 1 January 1980.

2.3. Runoff Bias-Correction Method

As presented in previous sections, daily gridded runoff simulations from the ERA5 dataset provide the hydrological forcing for GEOGLOWS RFS. To address systematic biases in the runoff data before routing, we implemented a bias-correction scheme that combines: (i) the Monthly Flow Duration Curve Quantile Mapping (MFDC-QM) proposed by Sánchez Lozano et al. [32], and (ii) the Sparse Cumulative Distribution Function (CDF) matching proposed by Lin et al. [37].

For the reference series, we used Flow Duration Curves (FDCs), expressed as runoff equivalents (mm/day), derived from the Global Streamflow Characteristics Dataset (GSCD), developed by Beck et al. [39]. Since GSCD provides only a discrete set of percentiles (1st, 5th, 10th, 20th, 50th, 80th, 90th, 95th, and 99th), we fitted a parametric distribution at each grid cell to derive a continuous reference FDC. We evaluated lognormal, Weibull, and gamma families and estimated their parameters by minimizing the sum of squared differences between observed and modeled quantiles at the available percentiles. The final distribution for each grid cell was then selected as the candidate yielding the lowest sum of squared errors. By contrast, the simulated FDC was obtained from the empirical distribution of the gridded daily runoff data derived from ERA5.

Although these three parametric families (lognormal, Weibull, and gamma) provide a practical way to derive continuous reference FDCs from the discrete GSCD percentiles, the present study does not evaluate non-parametric alternatives such as spline-based or empirical quantile interpolation. Therefore, the resulting correction should be interpreted as conditional on the selected distributional fitting strategy.

Note that ERA5 runoff data and GSCD are provided on different grid resolutions (0.25° and 0.05° , respectively). To preserve the finer GSCD resolution, ERA5 runoff data was resampled to 0.05° using nearest-neighbor mapping, avoiding value modification that can arise from interpolative resampling methods [46].

For each grid cell, we used the flow duration curves as a mapping between non-exceedance probability P and runoff. Let $FDC_{ref}(P)$ and $FDC_{sim}(P)$ denote the runoff value from the reference and simulated FDCs, respectively, expressed as a function of P . The correction coefficient at non-exceedance probability P was then defined as shown

in Equation (1). To avoid numerical instability in the correction coefficient, we imposed operational safeguards: when either $FDC_{sim}(P)$ or $FDC_{ref}(P)$ equals zero, the correction coefficient is automatically set to zero.

$$C_{cor}(P) = \frac{FDC_{ref}(P)}{FDC_{sim}(P)} \quad (1)$$

For each time step t , defined at daily resolution, in the gridded runoff time series, we computed the non-exceedance probability $P(t)$ associated with the simulated runoff $R_{sim}(t)$ on the empirical simulated FDC. The bias-corrected runoff $R_{cor}(t)$ was obtained by applying the corresponding correction factor, as shown in Equation (2).

$$R_{cor}(t) = R_{sim}(t) \cdot C_{cor}(P(t)) \quad (2)$$

In this way, each simulated runoff value was replaced by a reference-consistent counterpart with the same non-exceedance probability. However, both the empirical and simulated FDCs are least reliable in the distribution of tails (i.e., below the 1st percentile or above the 99th percentile), where quantiles are computed considering few events. As a result, statistical uncertainty increases, and the quantile-to-quantile transfer function, and its associated correction factor, becomes highly sensitive to outliers. This can amplify noise and lead to unstable, discontinuous, or unrealistically large corrections in the extremes [47]. In this regard, we cap $P(t)$ to this range: if $P(t)$ is below the 1st percentile, the 1st-percentile correction coefficient was used, and if $P(t)$ exceeds the 99th percentile, the 99th-percentile correction coefficient was used. This monotonic, rank-preserving mapping corrects systematic biases in magnitude while preserving the temporal ordering of runoff events [37]. In Figure 3, we present a graphical summary of the proposed runoff bias-correction workflow.

To assess how the proposed bias-correction method modifies runoff data at the grid scale, we compared the original ERA5 runoff with the bias-corrected runoff using the percentage of bias (PBIAS) presented in Equation (3).

$$PBIAS_{i,j} = \frac{100 \cdot \sum_{t=1}^T (S_{t,i,j} - R_{t,i,j})}{\sum_{t=1}^T (R_{t,i,j})} \quad (3)$$

where i and j denote the grid-cell indices, $S_{t,i,j}$ is the simulated runoff (i.e., the original ERA5 runoff), and the $R_{t,i,j}$ is the reference runoff (i.e., the bias-corrected runoff) at day t . The index $t = 1, \dots, T$ runs over all days in the time series (1979 to 2025), and T is the total number of days. This metric allowed us to quantify the magnitude and spatial pattern of systematic errors: positive PBIAS values indicating overestimation and negative values indicating underestimation.

For operational interpretation, we used $|PBIAS| \geq 50\%$ as a diagnostic threshold to identify areas with substantial residual runoff bias. This threshold was selected based on a supplementary threshold-sensitivity analysis that compared alternative $|PBIAS|$ thresholds of 25%, 50%, 75%, and 100% against the resulting station-level ΔKGE responses after pre-routing correction (Figure S1). The analysis showed that $|PBIAS| \geq 50\%$ provides a balanced criterion, capturing a clear separation in performance response while retaining sufficient station coverage for global interpretation.

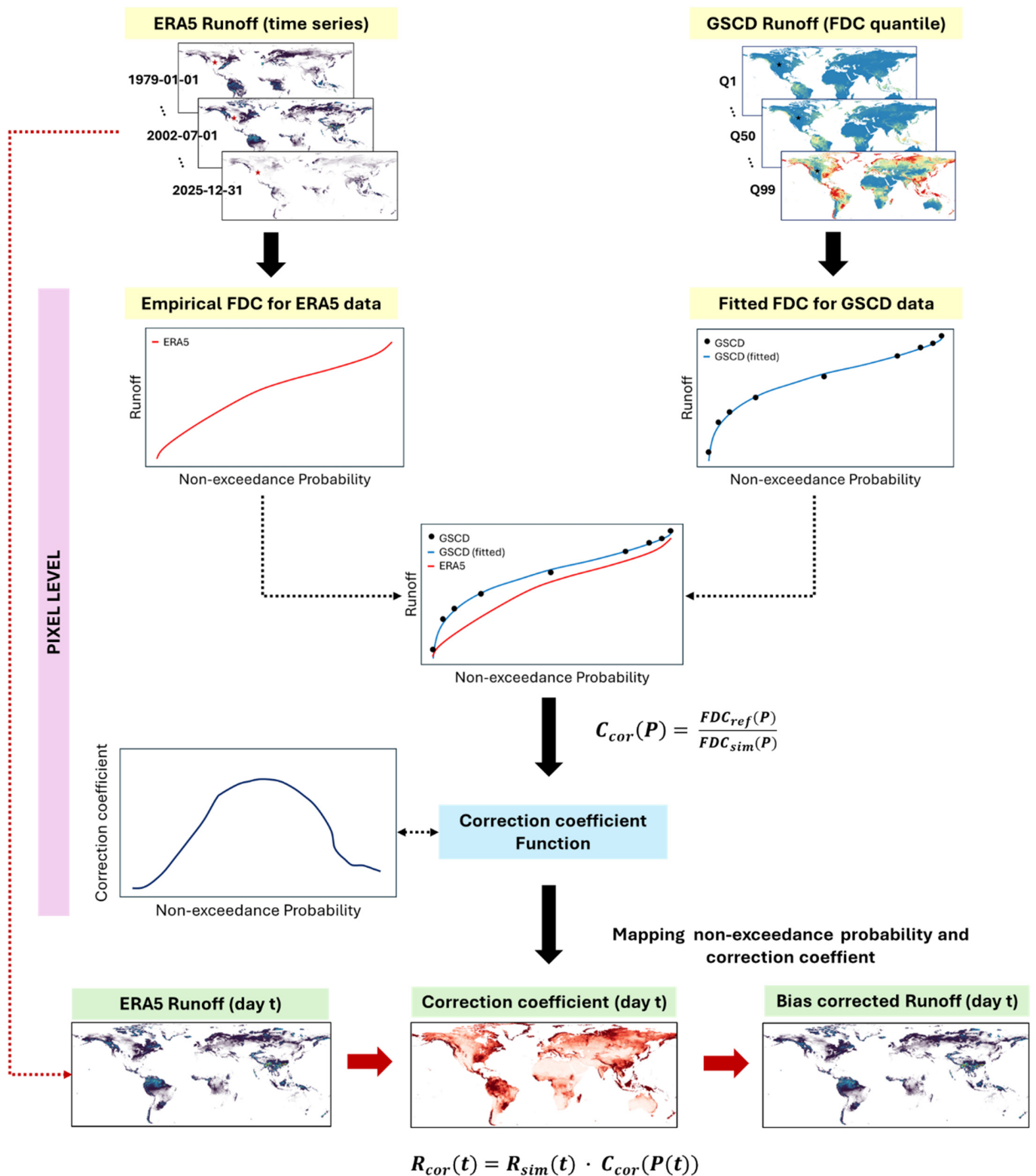


Figure 3. Schematic workflow of the proposed runoff bias-correction method, applied prior to Muskingum routing in the GEOGLOWS River Forecast System. Stars mark the same reference grid cell in the ERA5 and GSCD maps. Red and blue lines represent the empirical ERA5 and fitted GSCD Flow Duration Curves, respectively.

2.4. Model Performance Evaluation Framework

We evaluated the impact of the proposed pre-routing runoff bias-correction on the performance of GEOGLOWS RFS. To this end, we conducted two retrospective simulations

(from 1979 to 2025) with the GEOGLOWS RFS Version 1 configuration coupled with RAPID: (i) a baseline run forced with raw ERA5 runoff, and (ii) a run forced with bias-corrected ERA5 runoff using the proposed method. For comparative purposes, we also applied the post-processing streamflow bias-correction method proposed by Sánchez Lozano et al. [32] to the baseline model: Monthly Flow Duration Curve Quantile Mapping (MFDC-QM). This setup results in three model configurations: (i) the baseline GEOGLOWS RFS simulation, (ii) the GEOGLOWS RFS simulation driven by bias-corrected runoff, and (iii) the bias-corrected GEOGLOWS RFS simulation obtained with MFDC-QM.

These three configurations were evaluated against observed streamflow using the Kling–Gupta efficiency (KGE). We selected KGE because it combines three key aspects of model error (bias, variability, and correlation) into a single performance metric and has become a standard criterion for hydrological model assessment [48,49]. The use of KGE and its components was selected to provide a consistent diagnostic framework for long, continuous streamflow simulations across a large global station network.

The correlation component (r) evaluates the temporal correspondence between simulated and observed hydrographs and therefore captures the degree to which both signals co-vary through time. The bias ratio (β) and variability ratio (γ) separately diagnose errors in mean flow magnitude and flow variability. Together, these components allow changes in overall skill to be interpreted mechanistically, which is central to assessing whether pre-routing runoff correction primarily modifies timing, magnitude, or variability-related errors [32]. We compute the KGE as shown in Equations (4)–(7) [50].

$$KGE = 1 - \sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \quad (4)$$

$$r = \frac{\sigma_{os}}{\sigma_s \cdot \sigma_o} \quad (5)$$

$$\beta = \frac{\mu_s}{\mu_o} \quad (6)$$

$$\gamma = \frac{\sigma_s / \mu_s}{\sigma_o / \mu_o} \quad (7)$$

where r is the Pearson correlation coefficient between simulated (s) and observed (o) data; β is the bias ratio; γ is the variability ratio; μ is the mean streamflow, σ is the standard deviation, and σ_{os} is the covariance between simulated (s) and observed (o) data. The KGE ranges from $-\infty$ to 1. A KGE value of 1 indicates a perfect match between observed and simulated data, while values closer to 0 indicate poorer model performance. KGE values less than -0.41 suggest the average streamflow benchmark is a better predictor than the model [51].

To investigate spatial patterns in model performance, gauged stations were grouped into six regional classes based on their geographic location: North America (A), South America (B), Europe (C), Africa (D), Asia (E), and Oceania (F). This regional stratification enabled the assessment of how regime-dependent performance varies across different hydroclimatic contexts.

2.5. Significance Testing of Model Performance

To assess whether the proposed pre-routing runoff bias-correction leads to a statistically significant change in model performance compared to the baseline GEOGLOWS RFS simulation, we applied the Wilcoxon signed-rank test to station-wise KGE values [52,53]. For each gauging station, we computed the difference in KGE between the runoff bias-corrected and baseline GEOGLOWS RFS simulations and used the Wilcoxon signed-rank test to evaluate whether the median of these paired differences departs significantly from

zero (with a p -value less than 0.05). This non-parametric test is particularly appropriate in this study because KGE is a bounded, often skewed efficiency metric, and the normality assumptions required by parametric tests (e.g., t -test) are unlikely to hold [24,49]. In practice, a positive and statistically significant median KGE difference indicates that the runoff bias-corrected configuration outperforms the baseline, whereas a negative and significant median difference indicates a deterioration in model performance.

2.6. Seasonal Analysis Framework

To assess seasonal model performance for the three model configurations (baseline, pre-routing runoff bias-correction, and post-processing MFDC-QM), we conducted a station-specific, multi-monthly analysis of streamflow patterns. For each gauged station, we first computed the observed multi-monthly mean discharge for each calendar month (i.e., the climatological mean of all Januarys, Februarys, etc.) and then ranked 12 months from the highest to the lowest multi-monthly mean. This approach was adopted because fixed hydrological-year partitions (e.g., Jan–Mar, Apr–Jun, Jul–Sep, Oct–Dec) do not consistently isolate high-flow (wet) and low-flow (dry) conditions across the diverse range of global river regimes. Based on the station-wise ranking, we defined three equally sized seasonal groups: (i) the wet group, comprising the four months with the highest observed multi-monthly mean streamflow; (ii) the dry group, comprising the four months with the lowest values; and (iii) the shoulder group, comprising the four intermediate transition months. For each seasonal group, performance was quantified using the Kling–Gupta Efficiency (KGE) and its components, including the correlation (r), the bias ratio (β), and the variability ratio (γ).

2.7. Operational Interpretation

For operational interpretation, we defined a region as suitable for pre-routing runoff bias correction when three decision-support criteria were simultaneously met: (i) the regional median change in KGE relative to the baseline was at least 0.05 (median $\Delta\text{KGE} \geq 0.05$), (ii) the paired station-wise KGE differences were statistically significant according to the Wilcoxon signed-rank test ($p < 0.05$), and (iii) the proportion of stations with very poor performance ($\text{KGE} < -0.41$) decreased by at least 5 percentage points relative to the baseline. Regions that did not meet these criteria were not considered suitable for the uniform operational application of the correction, and retaining the baseline ERA5 forcing was preferred unless local validation indicated otherwise.

3. Results

3.1. Grid-Scale Differences Between Original and Bias-Corrected Runoff Data

The grid-scale PBIAS between the original and bias-corrected runoff data is shown in Figure 4. Negative values indicate that the original ERA5 runoff underestimates the bias-corrected runoff, whereas positive values denote overestimation. Using $|\text{PBIAS}| < 50\%$ as an operational consistency threshold, 46% of the global land surface falls within this interval, forming the most extensive contiguous patches across low-relief, temperate regions including much of Canada and the eastern–central United States, western and central Europe, northern Eurasia, and parts of the Amazon and La Plata lowlands. Strong negative departures ($\text{PBIAS} < -50\%$) account for 40% of land area and are concentrated in arid subtropical areas and continental interiors, notably North Africa, the Arabian Peninsula, and Central Asia. In contrast, strong positive departures ($\text{PBIAS} > 50\%$) cover 14% of land area and cluster along humid tropical and mountainous regions, particularly the Andes and the Himalayan arc, as well as parts of equatorial Africa (including the Congo sector and East African highlands) and northern Australia.

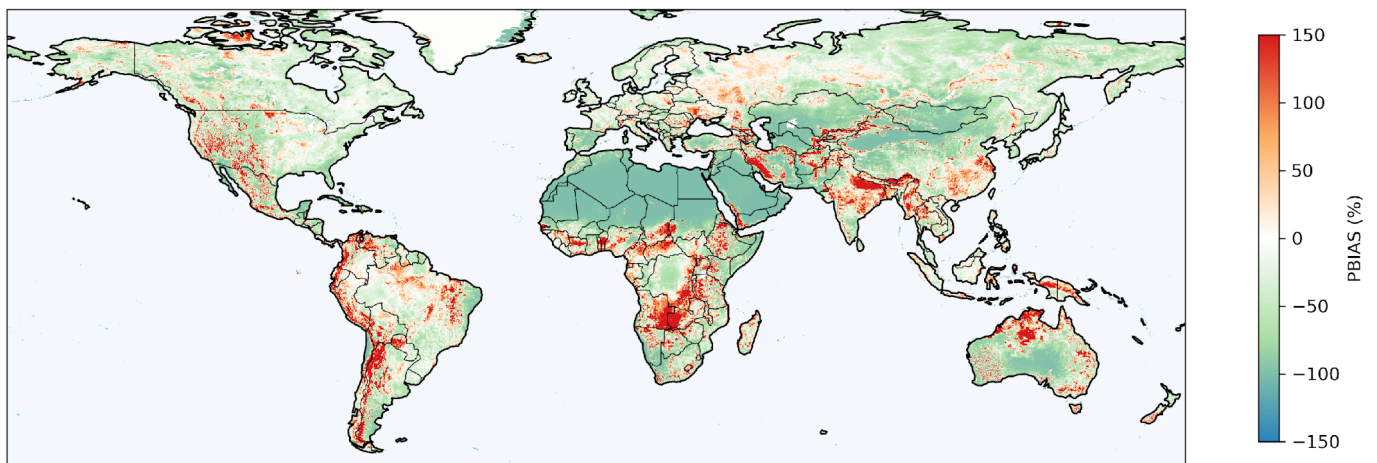


Figure 4. Percent bias (PBIAS) between original and bias-corrected ERA5 runoff for the period 1979–2025.

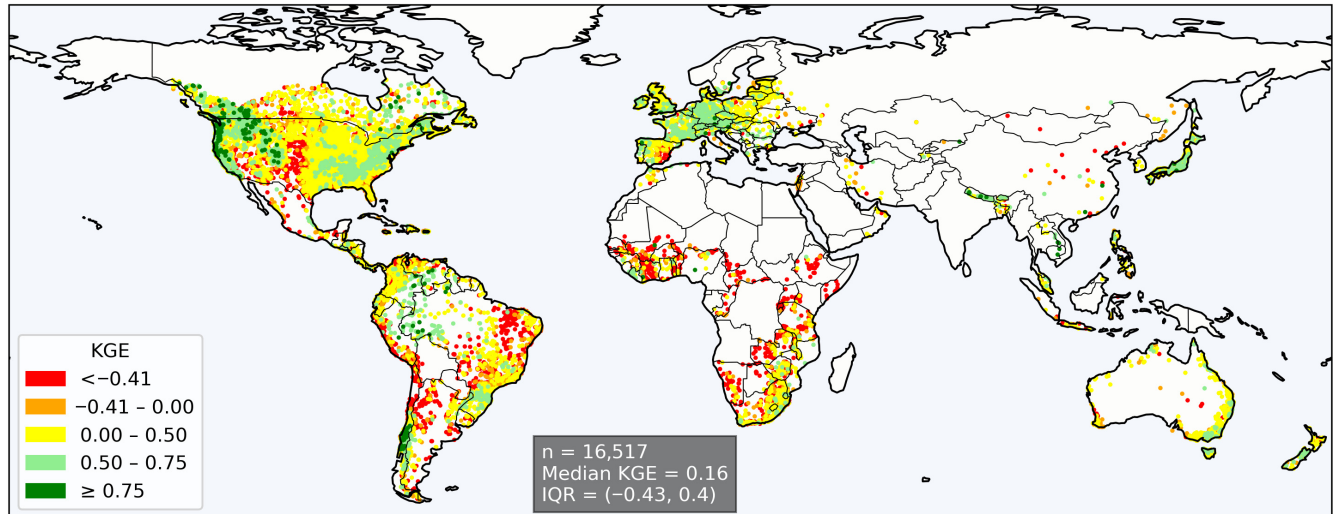
3.2. Model Performance Assessment

The spatial distribution of station-wise performance, expressed as KGE for the three model configurations, is shown in Figure 5. In the baseline GEOGLOWS RFS simulation (Figure 5a), out of 16,517 gauging stations, 30% present very poor performance ($KGE < -0.41$) and 14% show poor performance ($-0.41 \leq KGE < 0$). These low-performance stations are concentrated mainly in Africa, South America (Andes and Brazilian east coast), and central North America. Stations with satisfactory performance ($0 \leq KGE < 0.50$) account for 46% of the stations, whereas 10% exhibit good performance ($KGE \geq 0.50$). These better-performing stations are primarily located in western and central Europe, as well as in both eastern and western North America. Overall, the baseline GEOGLOWS RFS simulation yields a global median KGE of 0.16, with an interquartile range from -0.43 to 0.40 .

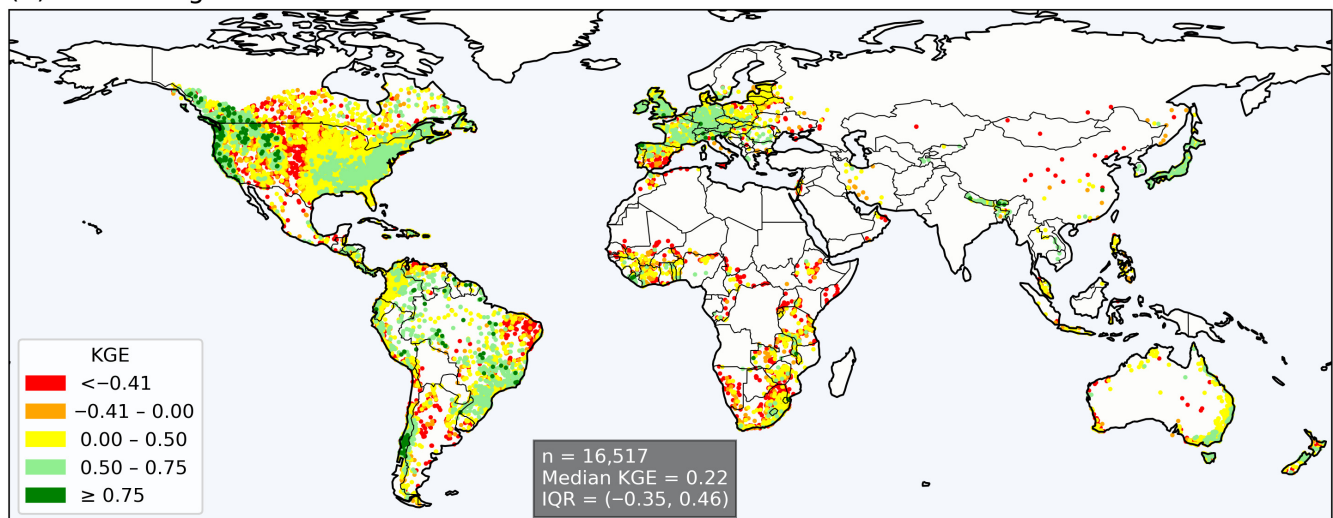
The GEOGLOWS RFS simulation driven by the proposed bias-corrected runoff (Figure 5b) leads to modest global improvement. The median KGE increases to 0.22 (IQR: -0.36 to 0.46), and the proportion of very poor ($KGE < -0.41$) and poor stations ($-0.41 \leq KGE < 0$) remains similar, with 29% and 14% of stations, respectively. Stations with satisfactory performance ($0 \leq KGE < 0.50$) represent 44%, while the fraction of gauges with good performance ($KGE \geq 0.50$) expands to 13%, with noticeable improvements in parts of South America and western North America.

The bias-corrected GEOGLOWS RFS simulation using MFDC-QM (Figure 5c) increases median performance from 0.16 to 0.48 (IQR: 0.26 to 0.65), corresponding to a ~ 2.5 -fold rise in median KGE compared to the GEOGLOWS RFS simulation driven by the proposed bias-corrected runoff. The fraction of stations in the very poor class ($KGE < -0.41$) drops to 9% and the poor class ($-0.41 \leq KGE < 0$) to 10%, while 47% of stations exhibit satisfactory performance ($0 \leq KGE < 0.50$). Notably, 44% of stations achieved good performance ($KGE \geq 0.50$), indicating a substantial shift toward higher performance levels under the MFDC-QM configuration.

(a) Baseline model



(b) Pre-routing bias-corrected model



(c) MFDC-QM model

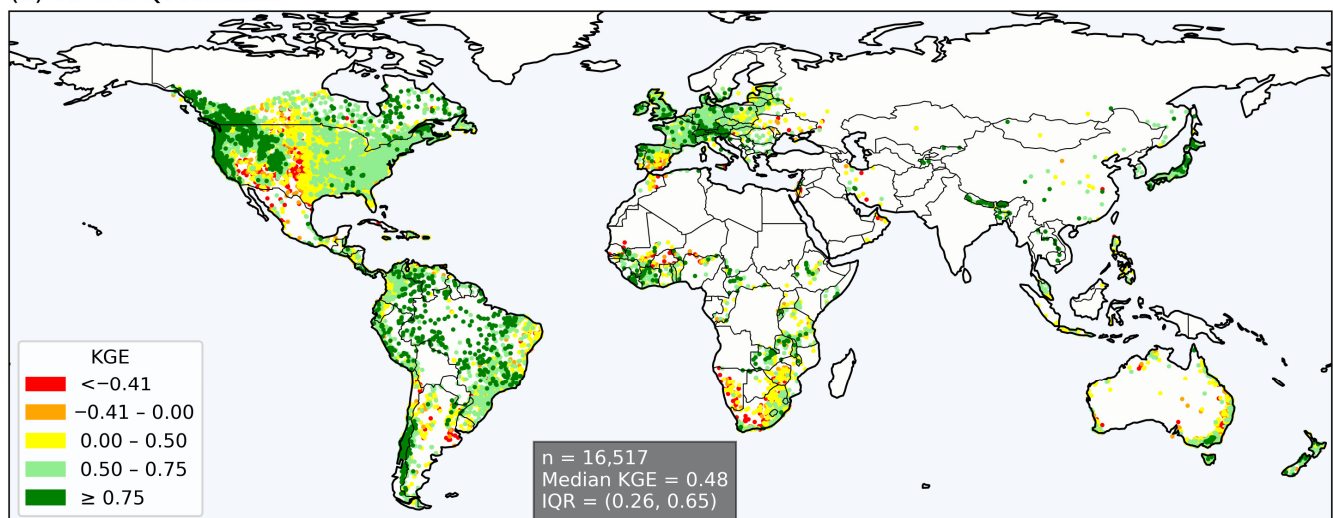


Figure 5. Model performance. (a) Baseline GEOGLOWS RFS simulation, (b) GEOGLOWS RFS simulation driven by bias-corrected runoff, and (c) bias-corrected GEOGLOWS RFS simulation obtained with MFDC-QM.

3.3. Cumulative Distribution Function (CDF) and Significance Testing of Model Performance

Figure 6 presents the cumulative distribution functions (CDFs) of station-wise KGE for the three GEOGLOWS RFS simulations. At the global scale, the baseline and runoff bias-corrected CDFs overlap, and the Wilcoxon signed-rank test indicated no statistically significant change in KGE ($p > 0.05$). If MFDC-QM provided no added performance, its CDF would similarly overlap the other curves and yield a comparable fraction of stations above the KGE = -0.41 benchmark. Instead, the MFDC-QM CDF was consistently shifted toward higher KGE values and showed a larger proportion of stations exceeding -0.41 benchmark, indicating improved performance relative to the baseline and runoff bias-corrected GEOGLOWS RFS configurations.

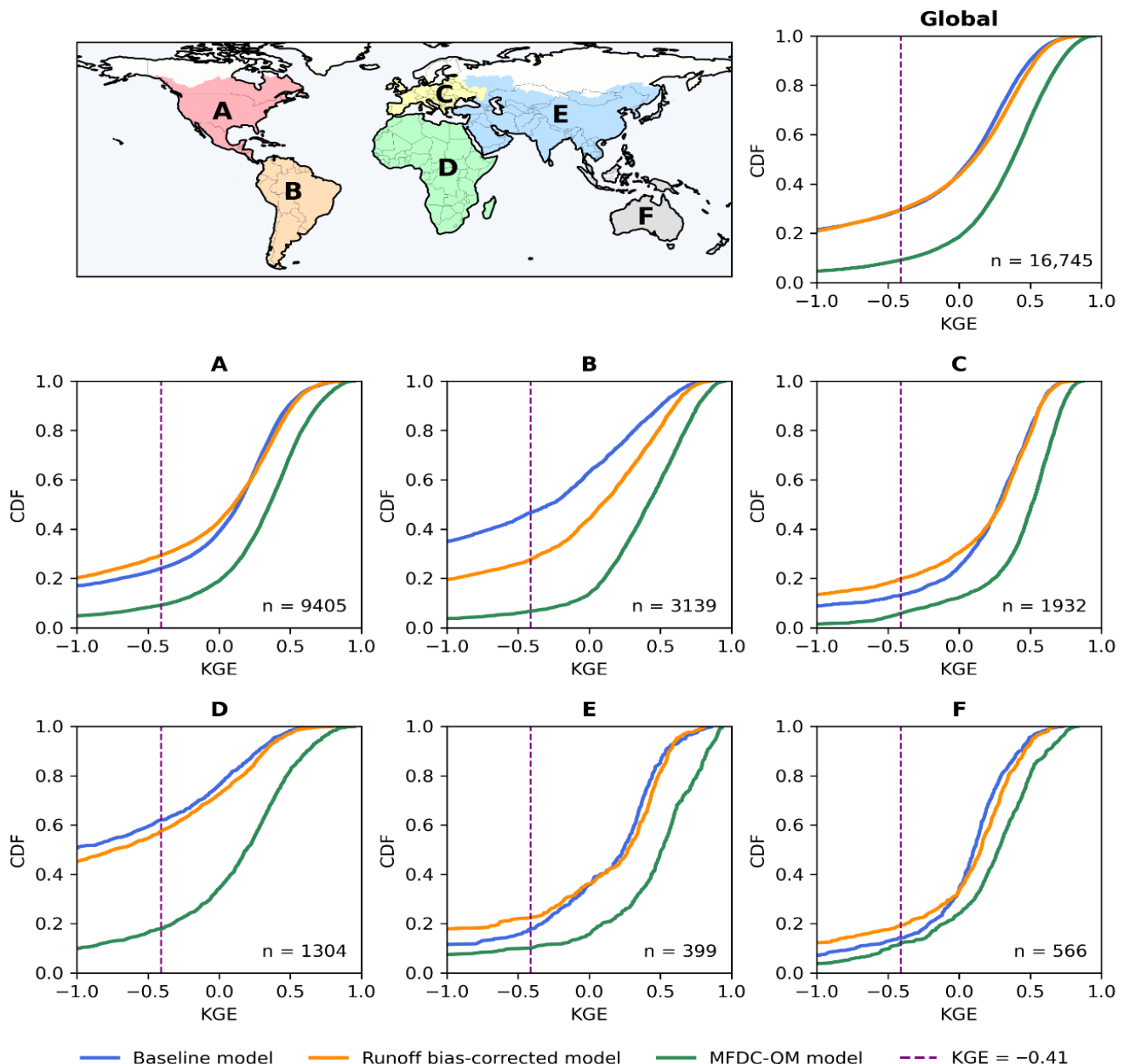


Figure 6. Cumulative distribution functions (CDF) of model performance for the three GEOGLOWS RFS simulations, shown globally and for six continental regions: (A) North and Central America, (B) South America, (C) Europe, (D) Africa, (E) Asia, and (F) Oceania.

Regionally, the impact of the runoff bias-correction was heterogeneous. In North America (Region A) and Europe (Region C), the bias-corrected GEOGLOWS RFS simulation

slightly degraded performance relative to the baseline model, particularly for stations with initially low performance ($KGE < 0$), whose CDF shifts toward more negative values. The Wilcoxon signed-rank test indicated that these changes were statistically significant in both regions ($p < 0.05$), implying a small but consistent deterioration in KGE performance of stations with $KGE < 0$. In contrast, over South America (Region B), the runoff bias-correction led to a statistically significant improvement in KGE performance ($p < 0.005$). Specifically, the proportion of stations with $KGE < -0.41$ decreased from about 44% in the baseline simulation to 23% by using bias-corrected runoff, and the CDF showed a consistent rightward shift towards higher KGE values across the distribution.

In Africa (Region D), the CDF for the bias-corrected simulation lay systematically to the right of the baseline curve, indicating improved performance across the KGE range. The Wilcoxon signed-rank test confirmed that this improvement was statistically significant ($p < 0.05$), consistent with a robust regional gain in median KGE. In Asia (Region E) and Oceania (Region F), the effect of the bias correction was mixed: stations with $KGE < 0$ tended to show a slight deterioration, whereas stations with $KGE > 0$ exhibited a modest improvement. In general, these shifts remained small, and the Wilcoxon test did not detect a statistically significant change in median KGE in either region ($p > 0.05$).

To further examine station-count-based changes among performance classes, we conducted a supplementary KGE category transition analysis between the baseline simulation and each bias-correction strategy (Figures S2 and S3). This analysis tracks whether each station moved to a higher KGE class, remained in the same class, or moved to a lower class after correction. Regionally, the largest categorical improvement under the pre-routing correction occurred in South America (Region B), where 43.9% of stations improved and only 6.6% degraded. By contrast, Europe (Region C) showed more stations degrading (19.8%) than improving (15.8%), consistent with the limited response observed in Figures 5 and 6.

3.4. Seasonal Analysis

Figures 7 and 8 summarize the seasonal performance of the three model configurations across all stations using the KGE components (correlation r , the bias ratio β , and the variability ratio γ). The pre-routing runoff bias-corrected model showed a season-dependent response: relative to the baseline, it produced marginal improvements during the wet season, as indicated by the increase of median KGE from 0.11 to 0.18. However, it degrades performance during the dry season, where mean KGE shifts from -0.10 to -0.15 (Figure 8). Across all seasonal groups, MFDC-QM remains consistently superior, yielding the highest KGE distributions and thus setting an upper bound on achievable improvement where streamflow records are available.

The correlation (r) remained similar between the baseline and the pre-routing runoff bias-corrected simulation across the dry (~ 0.41), shoulder (~ 0.45), and wet (~ 0.47) groups, while MFDC-QM exhibited a slightly higher central tendency, with an average increase of approximately 0.05 among groups (Figure 8). The bias ratio (β) exhibited a consistently less favorable response under the pre-routing runoff bias-correction. Across the dry, shoulder, and wet groups, the mean β increased and shifted further away from 1, indicating a systematic amplification of bias relative to the baseline. MFDC-QM maintained β values closer to 1 across seasons, with tighter interquartile ranges. By contrast, the variability ratio (γ) improved more consistently under the pre-routing runoff bias-correction, with γ distributions moving closer to 1 across all seasonal groups, and the largest shifts observed in the wet and shoulder seasons. MFDC-QM showed the most compact distributions around unity.

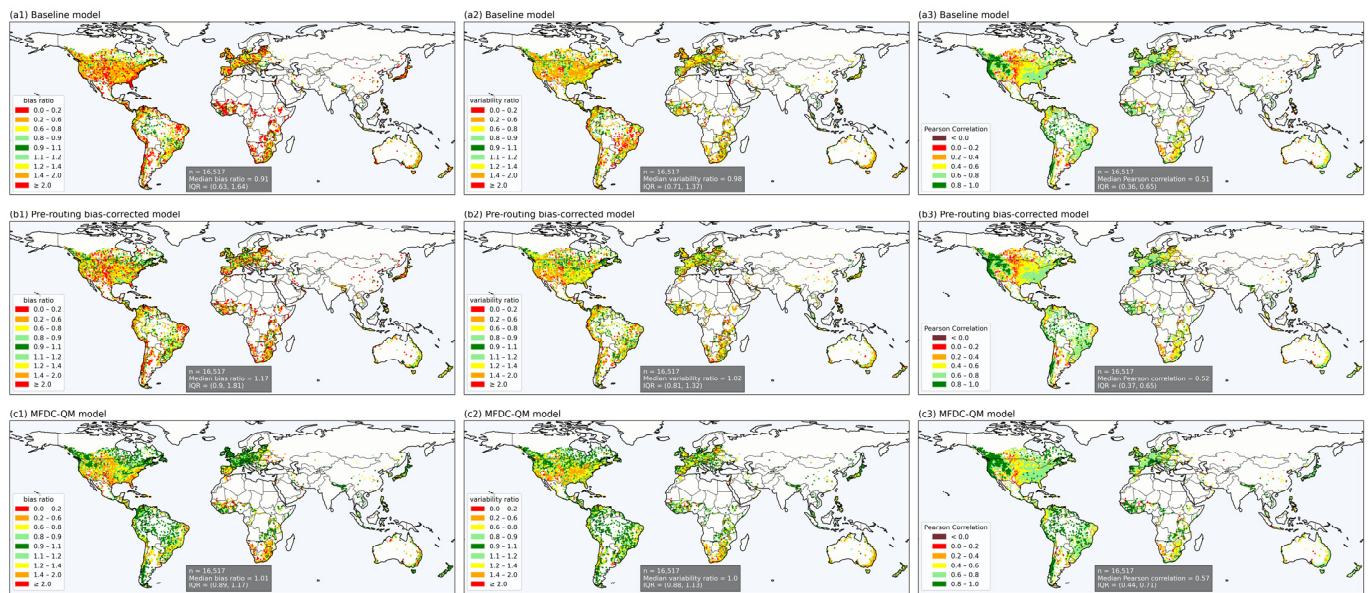


Figure 7. Spatial distribution of KGE components, bias ratio (β), variability ratio (γ), and Pearson correlation (r), for the three modeling configurations. (a1–a3) Baseline GEOGLOWS RFS simulation, (b1–b3) GEOGLOWS RFS simulation driven by bias-corrected runoff, and (c1–c3) bias-corrected GEOGLOWS RFS simulation obtained with MFDC-QM.

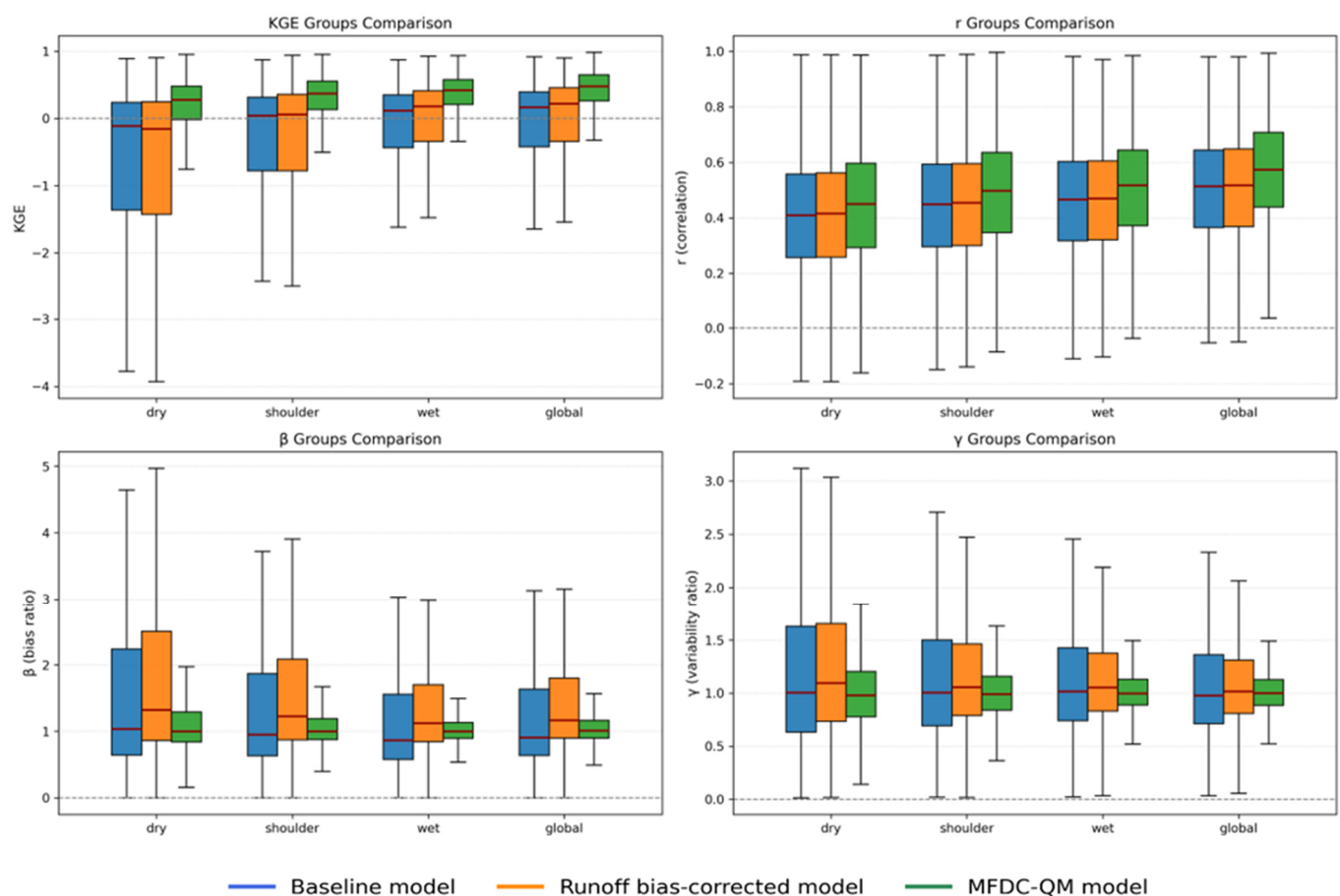


Figure 8. Seasonal performance of the baseline, pre-routing runoff bias-corrected, and MFDC-QM model configurations across all stations using KGE and its components (KGE, r , β , γ). The red line in each boxplot indicates the median.

Figure 9 extends the seasonal evaluation by stratifying stations by region, revealing marked continental contrasts in both KGE and its components. In North America and

Europe, the pre-routing runoff bias correction does not yield a consistent improvement in KGE across the dry, shoulder, and wet groups. Instead, the bias ratio (β) and variability ratio (γ) systematically depart from 1 and their interquartile ranges widen relative to the baseline, indicating a larger dispersion of residual errors across stations.

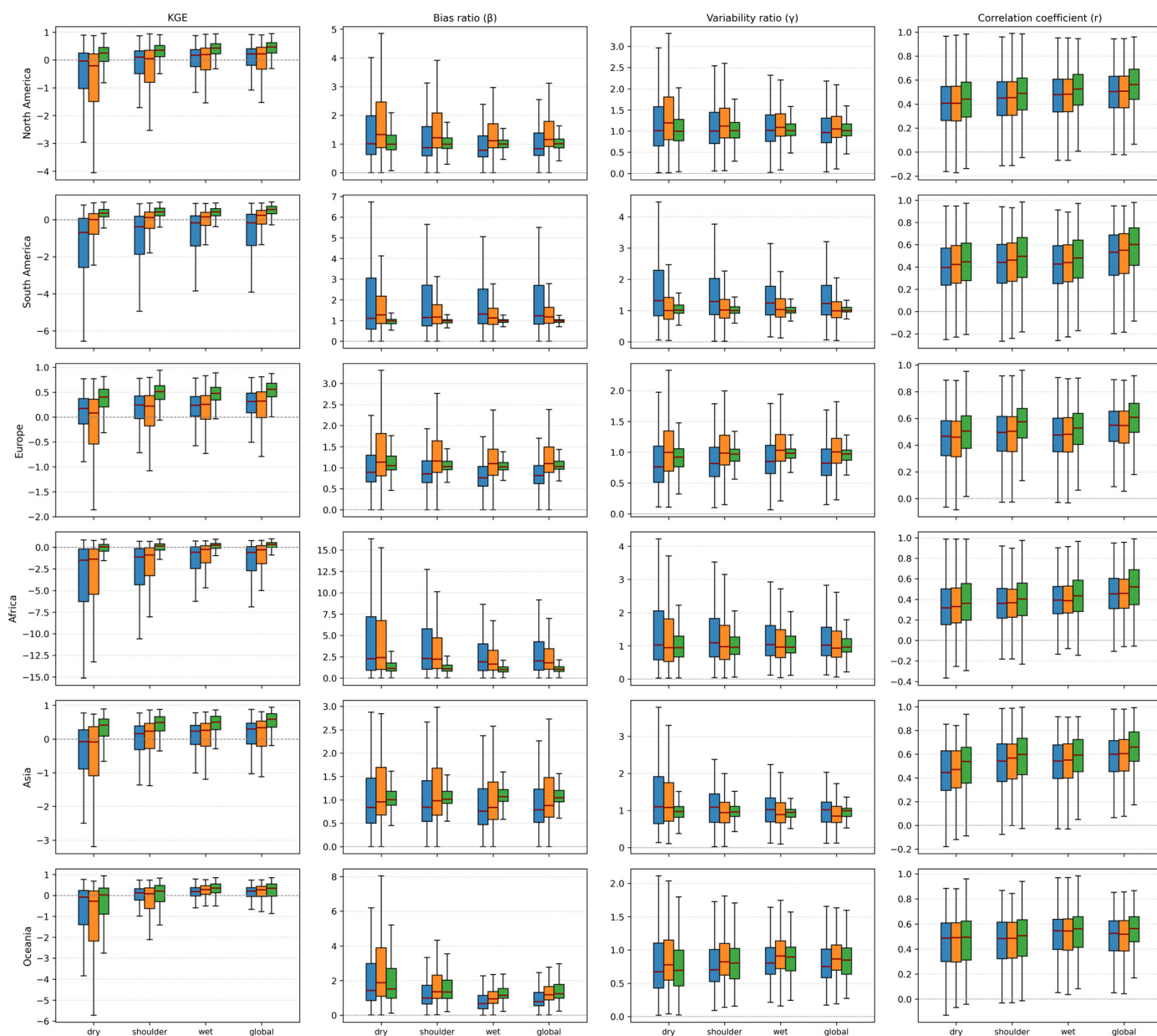


Figure 9. Regional analysis of seasonal performance for the baseline (blue), pre-routing runoff bias-corrected (orange), and MFDC-QM (green) model configurations using KGE and its components (KGE, r , β , γ). The red line in each boxplot indicates the median.

In South America, the response is notably different and consistently positive. Across seasons (dry, wet, and shoulder), the station-wise median KGE under the pre-routing correction shifts upward, with seasonal values typically spanning from approximately -0.7 to -0.1 in the baseline configuration to -0.05 to 0.15 after correction. The largest improvements occur during the wet season. This improvement is accompanied by coherent changes in all KGE components: both β and γ move closer to 1 and their distributions become more compact, while the mean correlation (r) increases modestly (about $+0.05$ across seasonal groups).

Africa shows a comparable behavior to South America, with clear increases in KGE and consistent component-level improvements, particularly through γ shifting toward 1 and a reduction in the spread of β , again with the strongest benefits during the wet season. In Asia and Oceania, changes are more heterogeneous and generally weaker, with small and inconsistent shifts in KGE across seasons, limited changes in r , and no systematic improvement in β and γ . Across all regions and seasonal groups, MFDC-QM remains the best-performing configuration, yielding the highest KGE distributions and the most compact component ratios around unity where discharge observations are available.

4. Discussion

4.1. Effectiveness of the Pre-Routing Runoff Bias Correction

At the global scale, the GEOGLOWS RFS simulation driven by the proposed pre-routing bias correction produced marginal improvements compared to the baseline GEOGLOWS RFS simulation. Results from the Wilcoxon signed-rank test indicated that changes in station-wise Kling–Gupta Efficiency (KGE) were not statistically significant ($p > 0.05$, Figure 6). In contrast, at the regional scale, the effect of runoff bias correction on simulated streamflow performance was heterogeneous. This pattern suggests that the impact of the proposed correction is mainly determined by the available scope for improvement in ERA5 runoff, which depends on the extent of reanalysis constraint by observations within each hydroclimatic domain.

Performance improvements were most evident in South America and Africa. These regions are characterized by sparser in situ observational coverage, which constrains the reanalysis of ERA5 data and leads to more pronounced systematic biases in the runoff forcing. In addition, challenges associated with tropical atmospheric processes and strong topography further complicate runoff estimation [54–56]. In areas with strong topographic gradients, such as the Andes, the Congo basin, and the East African highlands, the bias correction effectively reduced runoff overestimation in the original ERA5 data (regions with PBIAS > 50%; Figure 4), thereby improving streamflow simulations after routing, particularly during the wet season, when improvements were more consistent and stronger than during the dry season (Figure 9).

In contrast, in well-instrumented regions such as Europe and North America, ERA5 benefits from the assimilation of dense hydrometeorological networks, including ground-based stations, weather radars, and satellite observations [17,57,58]. As a result, the runoff derived in these regions is already adjusted to the hydroclimatic conditions, leaving limited scope for additional improvements through the proposed bias correction. In these well-instrumented regions, the proposed bias correction showed minimal changes for stations with positive model skill ($KGE > 0$), while performance degradation was observed mainly among stations with poor baseline performance ($KGE < 0$, Figure 6). This result suggests that, under conditions where the forcing is already of relatively high quality, additional correction based on GSCD may transfer residual uncertainties from the reference dataset into the routed simulation (Figures 8 and 9). Moreover, in most areas of these regions, differences between the original and corrected runoff remained within the operational consistency threshold ($|PBIAS| < 50\%$; Figure 4), further explaining the limited impact on overall model performance.

Results in Asia and Oceania showed no statistically significant changes in model performance. In Asia, this outcome was strongly influenced by limitations in the evaluation dataset rather than by the correction method. Although Asia encompasses a wide range of hydroclimatic regimes [59], the assessment relied on a relatively small number of streamflow gauges, most of which are concentrated in Japan. This spatially unrepresentative station distribution limits the robustness of regional-scale inferences and may mask

localized improvements or degradations in performance across the continent. A similar situation was observed in Oceania. While streamflow stations exist across several island regions, most evaluation gauges are concentrated in Australia and New Zealand. In these areas, ERA5 reanalysis benefits from extensive data assimilation, resulting in runoff with relatively small residual biases [60]. Consequently, the proposed runoff bias-correction approach produces limited changes in routed streamflow performance, consistent with the patterns observed in other well-instrumented regions such as Europe and North America.

4.2. Decomposing Seasonal Performance: Variability-Driven Improvements and Bias Trade-Offs

Seasonal and component-wise analysis (Figure 9) indicate that the impact of the pre-routing runoff bias correction is primarily expressed through changes in the KGE variability component. Across all seasonal groups (dry, shoulder, wet), the variability ratio (γ) of the runoff bias-corrected simulations shifts systematically toward unity relative to the baseline model simulations, particularly in South America and Africa, indicating a more suitable reproduction of routed flow variability after correction. By contrast, the correlation component (r) shows only marginal and spatially inconsistent changes, in line with Lin et al. [37], who previously discussed that grid-scale forcing adjustments do not materially modify event sequencing or phase coherence between simulated and observed hydrographs.

The bias ratio (β) exhibits the most heterogeneous response: in data-scarce regions such as South America and Africa, β generally shifts closer to unity, indicating that the pre-routing correction substantially reduces mean-flow bias, whereas in well-instrumented regions (e.g., North America and Europe) β more often departs from unity after correction, consistent with the propagation of residual distributional uncertainties from the GSCD reference into the mean-flow component of the routed simulation (Figures 8 and 9). These component-level trade-offs provide a mechanistic explanation for the regional dependence of net KGE changes: in data-scarce regions where baseline variability errors are dominant, improvements in γ tend to outweigh β deviations, yielding positive KGE responses, whereas in regions with already well-constrained forcing, limited gains in γ combined with β shifts result in neutral to slightly negative outcomes. Seasonal stratification further reinforces this interpretation, with more robust improvements during wet and shoulder periods, when variability exerts a stronger control on overall skill, and weaker or mixed responses during dry months, when low-flow conditions increase the leverage of mean-bias errors on KGE.

4.3. Benchmarking of the Proposed Runoff Bias-Correction Against MFDC-QM: Local Performance Improvements Versus Global Transferability

Results confirm that the bias-correction method developed by Sanchez Lozano et al. [32] remains the most effective approach when local observations are available, improving model performance relative to the reference simulation and the forced simulation with bias-corrected runoff (global mean KGE = 0.19 vs. KGE = 0.49). This is consistent with its methodological formulation: MFDC-QM directly corrects simulated streamflow at the station level by using the observed signal to adjust the streamflow distribution and its seasonality [29]. Therefore, MFDC-QM implicitly accounts for the combined effects of forcing, runoff generation, and routing errors [32]. In addition, its potential for improvement is greater because it acts on the final operational variable (streamflow) and not only on the forcing factor.

However, the strength of MFDC-QM also limits its scope. MFDC-QM is inherently local and data-dependent, as it can only be applied to reaches with sufficiently long, high-quality historical series. In contrast, the proposed bias-correction method in this study prioritizes spatial portability by correcting runoff at the grid scale before routing, allowing

its applicability in both gauged and ungauged basins. Operationally, this positions the proposed approach as a first-order correction of the hydrological forcing, especially relevant for regions with low hydrological gauge density, where methods based on observed flow are not systematically implementable.

An additional methodological consideration is the independence between information used to define the correction and the data used for evaluation. In MFDC-QM, the correction is defined from the same observed series used to evaluate performance. Although this is common in operational post-processing schemes, it can lead to an optimistic performance estimate if explicit temporal separation strategies are not implemented (e.g., out-of-sample validation). In this study, the correction relies on an independent global product (i.e., GSCD), and performance was evaluated using a set of stations compiled for this study, including not only the GRDC dataset but also additional stations and an extended period up to 2025. In general, these introduce an additional degree of independence between the data used for correction and the data used for evaluation.

4.4. Role and Limitations of GSCD as a Priori Hydrological Information

The Global Streamflow Characteristics Dataset (GSCD) plays a central role in the proposed runoff bias-correction method by providing spatially distributed reference information on runoff magnitude by using Flow Duration Curves. It is important to emphasize that GSCD does not represent direct runoff observations, but rather synthesized hydrological characteristics derived from available streamflow records, satellite-based information, and modeling assumptions [39]. Consequently, the information it provides is inherently distributional rather than temporal, making it well-suited for FDC-based correction approaches that aim to adjust runoff magnitudes while preserving event ranking [37].

Despite these limitations, the results indicated that GSCD contains sufficient hydrological signals to improve runoff forcing in data-scarce regions [39]. This finding highlights the potential for integrating heterogeneous hydrological information sources into global bias-correction frameworks. At the same time, the effectiveness of the correction is bounded by the accuracy and representativeness of GSCD, linking future improvements in global runoff correction to advances in hydrological data availability and coverage. Longer streamflow records, expanded monitoring networks, satellite-based discharge observations (e.g., SWOT), and machine learning techniques for spatial generalization all represent promising avenues for enhancing future versions of GSCD and, by extension, the effectiveness of pre-routing correction strategies.

A further limitation is that this study does not include a formal sensitivity analysis of the runoff correction to alternative FDC fitting strategies, resampling methods, or uncertainty in the GSCD percentiles. The correction is therefore conditional on the selected parametric distribution at each grid cell, the nearest-neighbor resampling of ERA5 runoff to the GSCD grid, and the GSCD-derived percentile estimates used as reference information. Future work should explicitly quantify this uncertainty by testing non-parametric spline or empirical quantile fits, alternative spatial remapping approaches such as bilinear and conservative resampling, and perturbation or ensemble experiments applied to key GSCD percentiles. Such analyses would allow the robustness of ΔKGE and the wet/dry seasonal responses to be quantified more directly.

The treatment of distribution tails represents an additional limitation of the proposed framework. Because GSCD provides a finite set of percentiles, the correction does not extrapolate beyond the 1st and 99th percentile bounds. Instead, correction factors are capped at these limits to avoid unstable tail behavior. While this choice improves numerical robustness, it also means that the present study does not explicitly quantify how the correction affects the most extreme flood or drought conditions. Future work should

therefore report the frequency of capping by region and season and evaluate tail-sensitive diagnostics, including Q95/Q5 bias, peak-over-threshold hit rates, and event-based flood and drought detection metrics.

4.5. Operational Implications for GEOGLOWS and Global Forecasting Systems

From an operational perspective, the proposed pre-routing runoff bias correction offers significant advantages for global forecasting systems, such as the GEOGLOWS River Forecast System. Applying the adjustment directly to gridded runoff forcing enables consistent deployment across river networks without the need for local discharge observations. This approach supports scalability, facilitates streamlined maintenance, and ensures spatially coherent outputs in large-domain applications.

The decision-support criteria defined in this study indicate that uniform global deployment is not recommended. Instead, operational implementation should follow a region-aware configuration. South America and Africa satisfy the proposed criteria, combining statistically significant performance improvements (median $\Delta\text{KGE} \geq 0.05$, $p < 0.05$) and reductions in the proportion of stations with $\text{KGE} < -0.41$. These regions also coincide with areas where ERA5 runoff exhibits larger residual biases, making them the strongest candidates for operational pre-routing correction. In contrast, North America and Europe do not meet the proposed criteria, because the correction produced limited or slightly negative changes in model performance. In these well-instrumented regions, retaining the baseline ERA5 forcing is advisable unless local validation demonstrates added value from the correction.

It is important to note that pre-routing correction does not resolve uncertainties associated with river geometry, routing parameterization, or streamflow regulation. Additionally, it does not substitute for the improvements attainable through station-based bias correction when observational data are available. Within an operational framework, its most appropriate role is complementary: enhancing baseline performance in ungauged and data-limited regions, while coexisting with discharge-based post-processing methods (such as MFDC-QM) in well-instrumented basins. The findings of this study focused on retrospective simulations rather than real-time forecast verification, and the implementation used GEOGLOWS RFS version 1, although with a methodology designed to be transferable across model resolutions. Finally, the effectiveness of the correction is intrinsically linked to the quality of GSCD, which remains an evolving dataset.

5. Conclusions

This study evaluated a grid-scale, pre-routing runoff bias-correction framework for global hydrological modelling within GEOGLOWS RFS version 1. The method combines Flow Duration Curve mapping and sparse Cumulative Distribution Function matching to adjust ERA5 runoff using GSCD-derived runoff distributions before river routing. Its main methodological contribution is demonstrating that globally available distributional hydrological information can be used to correct runoff forcing prior to routing, providing a scalable alternative to discharge-based post-processing in ungauged and data-limited basins. Across 16,517 gauging stations from 1980 to 2025, the correction produced a modest global increase in median KGE from 0.16 to 0.22, compared with 0.48 for MFDC-QM where discharge observations were available. However, the response was strongly regional: statistically significant improvements occurred in South America and Africa, where residual ERA5 runoff biases were larger and observational constraints are more limited, whereas North America and Europe showed limited or slightly negative responses, indicating that a uniform global correction is not appropriate.

Seasonal and component-wise analyses showed that the main benefit of pre-routing correction was an improved representation of streamflow variability. The variability ratio γ shifted closer to unity, particularly during wet and transitional seasons, while the correlation component r changed only marginally and the bias ratio β showed mixed behavior. These results indicate that pre-routing runoff correction is most effective where baseline errors are dominated by runoff magnitude and variability biases, but less effective where streamflow errors arise from routing, river geometry, regulation, or local-scale processes that cannot be corrected through runoff forcing alone. This also explains the continued superiority of MFDC-QM at gauged sites, as station-based post-processing corrects the final routed streamflow and can implicitly compensate for multiple sources of error beyond the runoff forcing.

Overall, the results support a region-specific use of bias-correction strategies in global hydrological forecasting systems. Based on the operational decision-support criteria defined in this study—median $\Delta KGE \geq 0.05$, Wilcoxon $p < 0.05$, and at least a 5-percentage-point reduction in the proportion of stations with $KGE < -0.41$ —South America and Africa are the regions where pre-routing correction is currently most justified, while observation-based post-processing should remain preferred where reliable discharge records exist. The main limitations are the use of GEOGLOWS RFS version 1 rather than version 2, reliance on retrospective rather than real-time forecast verification, and dependence on the quality and representativeness of GSCD. Although the method is expected to be methodologically transferable to GEOGLOWS RFS version 2 and other gridded-runoff-driven routing frameworks, its performance with higher-resolution river networks and updated routing parameterizations requires explicit evaluation. Future work should also test the sensitivity of the correction to alternative FDC fitting approaches, spatial resampling methods, GSCD percentile uncertainty, and tail-specific diagnostics relevant to flood and drought applications.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/hydrology13050128/s1>. Figure S1: PBIAS threshold sensitivity analysis comparing ΔKGE_{pre} distributions and fraction of improving stations for |PBIAS| thresholds of 25%, 50%, 75%, and 100%. Figure S2: KGE category transition matrices for Baseline $\rightarrow$ Pre-routing correction and Baseline $\rightarrow$ MFDC-QM. Figure S3: Regional distribution of KGE category transitions for Baseline $\rightarrow$ Pre-routing correction and Baseline $\rightarrow$ MFDC-QM, expressed as percentage of stations classified as Improved, No change, or Degraded. Table S1: Summary of PBIAS threshold sensitivity analysis including station counts, median ΔKGE_{pre} , fraction of improving stations, and Mann–Whitney U test results for each threshold. Tables S2–S6: Global and regional KGE category transition summaries for both bias-correction configurations relative to the baseline.

Author Contributions: Conceptualization, J.E.C., J.L.S.-L. and B.G.V.; methodology, J.E.C., J.L.S.-L. and B.G.V.; software, J.E.C.; validation, J.E.C. and J.L.S.-L.; formal analysis, J.E.C., J.L.S.-L. and B.G.V.; investigation, J.E.C. and J.L.S.-L.; data curation, J.E.C. and J.L.S.-L.; writing—original draft preparation, J.E.C., J.L.S.-L. and B.G.V.; writing—review and editing, B.G.V., M.G.T.-V., E.J.N., R.C.H. and A.L.G.; supervision, J.L.S.-L. and B.G.V. All authors have read and agreed to the published version of the manuscript.

Funding: This work was conducted without dedicated funding. All contributions were made voluntarily by the authors during their personal time.

Data Availability Statement: The datasets and software used in this study are available as follows: 1. **GEOGLOWS Stream Network:** The stream network data used for model simulations is publicly available at <https://training.geogloss.org/rfs/datasets/catalog/> (accessed on 12 February 2026). 2. **Observed Station Network and Data:** Observed discharge station network, and the observed discharge data are stored on Hydro-Server: <https://hydroserver.geogloss.org> (accessed on 12 February 2026). For some regions included in this study, the observed discharge data are not publicly available. Access must be requested from the respective national hydrometeorological institutions. 3. **Gridded**

Runoff data: The gridded runoff data are available in the following OneDrive repository. The repository includes the following gridded daily time series: (i) ERA5 Runoff Data; (ii) Bias-corrected ERA5 Runoff Data; and (iii) Flow duration curves computed for simulated and reference runoff data. 4. **Global Streamflow Characteristics Dataset:** The GSCD is available on this official website: <https://www.gloh2o.org/gscd/> (accessed on 12 February 2026). 5. **GEOGLOWS Historical Simulations:** Historical simulations (1979–2025) of the GEOGLOWS River Forecast System are available in the following OneDrive repository. The repository contains both the original runs and the runoff bias-corrected runs. 6. **Codes.** The Python and R codes used for this study are available in the following GitHub repository: <https://github.com/Fundacion-Ecociencia/runoff-bias-correction-geogloWS> (accessed on 12 February 2026).

Acknowledgments: We express gratitude to many collaborators at hydrometeorological agencies worldwide, including CEMADEN, INAMHI, SENAMHI, IDEAM, INDRHI, USGS, Australian Government Department of Meteorology, and the Water Survey of Canada, among others, who, along with GRDC, contributed gauge data and shared local expertise with the authors over several years. We also thank the NASA GEOGLOWS and NASA SERVIR (SERVIR-Hindu Kush Himalaya, SERVIR-Mekong, SERVIR-East and Southern Africa, and SERVIR-Amazonia), and CIIFEN, who helped us obtain some of the observed data used in this study. During the preparation of this manuscript, the authors used Grammarly and Chat-GPT 5.2 for the purposes of fixing typos, spelling, grammar, and improve clarity in sentences and paragraphs. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: Author Jorge Luis Sánchez-Lozano was employed by the company Aquaveo LLC. Aquaveo LLC had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Nearing, G.; Cohen, D.; Dube, V.; Gauch, M.; Gilon, O.; Harrigan, S.; Hassidim, A.; Klotz, D.; Kratzert, F.; Metzger, A.; et al. Global Prediction of Extreme Floods in Ungauged Watersheds. *Nature* **2024**, *627*, 559–563. [\[CrossRef\]](#)
2. Okacha, A.; Salhi, A.; Bouchouou, M.; Fattasse, H. Enhancing Flood Forecasting Accuracy in Data-Scarce Regions through Advanced Modeling Approaches. *J. Hydrol.* **2024**, *645*, 132283. [\[CrossRef\]](#)
3. Tu, T.; Wang, J.; Zhao, G.; Zhao, T.; Dong, X. Scaling from Global to Regional River Flow with Global Hydrological Models: Choice Matters. *J. Hydrol.* **2024**, *633*, 130960. [\[CrossRef\]](#)
4. Brunner, M.I.; Slater, L.; Tallaksen, L.M.; Clark, M. Challenges in Modeling and Predicting Floods and Droughts: A Review. *WIREs Water* **2021**, *8*, e1520. [\[CrossRef\]](#)
5. CRED. *2019 Disasters in Numbers*; Centre for Research on the Epidemiology of Disasters: Brussels, Belgium, 2020.
6. CRED. *2024 Disasters in Numbers*; Centre for Research on the Epidemiology of Disasters: Brussels, Belgium, 2025.
7. Rentschler, J.; Salhab, M.; Jafino, B.A. Flood Exposure and Poverty in 188 Countries. *Nat. Commun.* **2022**, *13*, 3527. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Wyns, A. Disaster Diplomacy in the Wake of the 2022 Pakistan Floods. *Lancet Planet. Health* **2022**, *6*, e855–e856. [\[CrossRef\]](#)
9. Clarke, B.; Barnes, C.; Rodrigues, R.; Zachariah, M.; Alves, L.M.; Haarsma, R.; Pinto, I.; Yang, W.; Vahlberg, M.; Vecchi, G.; et al. Climate Change and El Niño behind Extreme Precipitation Leading to Major Floods in Southern Brazil in 2024. *npj Nat. Hazards* **2026**, *3*, 4. [\[CrossRef\]](#)
10. Anderson, W.; Taylor, C.; McDermid, S.; Ilboudo-Nébié, E.; Seager, R.; Schlenker, W.; Cottier, F.; De Sherbinin, A.; Mendeloff, D.; Markey, K. Violent Conflict Exacerbated Drought-Related Food Insecurity between 2009 and 2019 in Sub-Saharan Africa. *Nat. Food* **2021**, *2*, 603–615. [\[CrossRef\]](#)
11. Alfieri, L.; Bucherie, A.; Libertino, A.; Campo, L.; D’Andrea, M.; Ghizzoni, T.; Gabellani, S.; Massabò, M.; Rossi, L.; Rudari, R.; et al. Operational Impact-Based Flood Early Warning in Lao PDR and Cambodia. *Int. J. Disaster Risk Reduct.* **2025**, *128*, 105718. [\[CrossRef\]](#)
12. Busker, T.; Rodriguez Castro, D.; Vorogushyn, S.; Kwadijk, J.; Zoccatelli, D.; Loureiro, R.; Murdock, H.J.; Pfister, L.; Dewals, B.; Slager, K.; et al. Comparing Flood Forecasting and Early Warning Systems in Transboundary River Basins. *EGUosphere* **2025**. [\[CrossRef\]](#)
13. Sutanto, S.J.; Syaehuddin, W.A.; De Graaf, I. Hydrological Drought Forecasts Using Precipitation Data Depend on Catchment Properties and Human Activities. *Commun. Earth Environ.* **2024**, *5*, 118. [\[CrossRef\]](#)

14. Jing, X.; Yang, X.; Luo, J.; Zuo, G. A Flexible, Differentiable Framework for Neural-Enhanced Hydrological Modeling: Design, Implementation, and Applications with HydroModels.Jl. *Environ. Model. Softw.* **2026**, *197*, 106802. [\[CrossRef\]](#)
15. Singh, V.P. Hydrologic Modeling: Progress and Future Directions. *Geosci. Lett.* **2018**, *5*, 15. [\[CrossRef\]](#)
16. Sanchez Lozano, J.L.; Chancay, J.E.; Rojas Lesmes, D.J.; Romero Bustamante, E.G.; Nelson, E.J.; Trujillo-Vela, M.G.; Larco, K.; Hales, R.C.; Souffront, M.A.; Gutierrez, A.L. Global Water Level Forecast System: Integrating EO Data for Hydrological Hazard Monitoring. *GeoHorizons* **2026**, *1*, gh2025-14. [\[CrossRef\]](#)
17. Thépaut, J.-N. Satellite Data Assimilation in Numerical Weather Prediction: An Overview. In *Proceedings of the ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean, Shinfield Park, Reading, UK, 8–12 September 2003*; ECMWF: Reading, UK, 2003.
18. Alfieri, L.; Burek, P.; Dutra, E.; Krzeminski, B.; Muraro, D.; Thielen, J.; Pappenberger, F. GloFAS—Global Ensemble Streamflow Forecasting and Flood Early Warning. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 1161–1175. [\[CrossRef\]](#)
19. Hales, R.C.; Nelson, E.J.; Kopp, S.; Levin, H.K.; Morris, A.L.; Souffront Alcantara, M.A.; Gutiérrez, A.; Magoffin, R.H.; Rosas, L.; Baaniya, Y. The Second Generation Geoglows River Forecast System. *SSRN* **2025**. [\[CrossRef\]](#)
20. Souffront Alcantara, M.A.; Nelson, E.J.; Shakya, K.; Edwards, C.; Roberts, W.; Krewson, C.; Ames, D.P.; Jones, N.L.; Gutierrez, A. Hydrologic Modeling as a Service (HMaaS): A New Approach to Address Hydroinformatic Challenges in Developing Countries. *Front. Environ. Sci.* **2019**, *7*, 158. [\[CrossRef\]](#)
21. Troin, M.; Arsenault, R.; Wood, A.W.; Brissette, F.; Martel, J. Generating Ensemble Streamflow Forecasts: A Review of Methods and Approaches Over the Past 40 Years. *Water Resour. Res.* **2021**, *57*, e2020WR028392. [\[CrossRef\]](#)
22. Wang, W.; Liu, J.; Xu, B.; Li, C.; Liu, Y.; Yu, F. A WRF/WRF-Hydro Coupling System with an Improved Structure for Rainfall-Runoff Simulation with Mixed Runoff Generation Mechanism. *J. Hydrol.* **2022**, *612*, 128049. [\[CrossRef\]](#)
23. Rojas-Lesmes, D.; Sanchez-Lozano, J.; Nelson, E.J.; Chancay-Sanchez, J.; Rodriguez-Chaves, J.; Larco-Erazo, K.; Romero, E.G.; Trujillo-Vela, M.; Hales, R.C.; Ames, D.P.; et al. National Water Level Forecast (NWLF): An Open-Source Customizable Web Application for the GEOGLOWS ECMWF Global Hydrological Model. *Inf. Geogr.* **2025**, *1*, 100008. [\[CrossRef\]](#)
24. Kupzig, J.; Kupzig, N.; Flörke, M. Regionalization in Global Hydrological Models and Its Impact on Runoff Simulations: A Case Study Using WaterGAP3 (v 1.0.0). *Geosci. Model Dev.* **2024**, *17*, 6819–6846. [\[CrossRef\]](#)
25. Sperna Weiland, F.C.; Vrugt, J.A.; Van Beek, R.P.H.; Weerts, A.H.; Bierkens, M.F.P. Significant Uncertainty in Global Scale Hydrological Modeling from Precipitation Data Errors. *J. Hydrol.* **2015**, *529*, 1095–1115. [\[CrossRef\]](#)
26. Sivelle, V.; Massari, C.; Trambly, Y.; Filippucci, P.; Dakhlaoui, H.; Quintana-Seguí, P.; Clavera-Gispert, R.; Boutaghane, H.; Boulmaiz, T.; Quast, R.; et al. Investigating Hydrological Modeling Uncertainties in the Mediterranean Region by Combining Precipitation and Soil Moisture Products. *J. Hydrol. Reg. Stud.* **2025**, *62*, 103015. [\[CrossRef\]](#)
27. Guo, H.; Hou, Y.; Yang, Y.; McVicar, T.R. Global Evaluation of Simulated High and Low Flows from 23 Macroscale Models. *J. Hydrometeorol.* **2024**, *25*, 425–443. [\[CrossRef\]](#)
28. Heinicke, S.; Volkholz, J.; Schewe, J.; Gosling, S.N.; Müller Schmied, H.; Zimmermann, S.; Mengel, M.; Sauer, I.J.; Burek, P.; Chang, J.; et al. Global Hydrological Models Continue to Overestimate River Discharge. *Environ. Res. Lett.* **2024**, *19*, 074005. [\[CrossRef\]](#)
29. Sanchez Lozano, J.L.; Romero Bustamante, G.; Hales, R.C.; Nelson, E.J.; Williams, G.P.; Ames, D.P.; Jones, N.L. A Streamflow Bias Correction and Performance Evaluation Web Application for GEOGloWS ECMWF Streamflow Services. *Hydrology* **2021**, *8*, 71. [\[CrossRef\]](#)
30. Farmer, W.H.; Over, T.M.; Kiang, J.E. Bias Correction of Simulated Historical Daily Streamflow at Ungauged Locations by Using Independently Estimated Flow Duration Curves. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 5741–5758. [\[CrossRef\]](#)
31. Tanguy, M.; Eastman, M.; Chevuturi, A.; Magee, E.; Cooper, E.; Johnson, R.H.B.; Facer-Childs, K.; Hannaford, J. Optimising Ensemble Streamflow Predictions with Bias Correction and Data Assimilation Techniques. *Hydrol. Earth Syst. Sci.* **2025**, *29*, 1587–1614. [\[CrossRef\]](#)
32. Sanchez Lozano, J.L.; Rojas Lesmes, D.J.; Romero Bustamante, E.G.; Hales, R.C.; Nelson, E.J.; Williams, G.P.; Ames, D.P.; Jones, N.L.; Gutierrez, A.L.; Cardona Almeida, C. Historical Simulation Performance Evaluation and Monthly Flow Duration Curve Quantile-Mapping (MFDC-QM) of the GEOGLOWS ECMWF Streamflow Hydrologic Model. *Environ. Model. Softw.* **2025**, *183*, 106235. [\[CrossRef\]](#)
33. Hales, R.C.; Nelson, E.J.; Souffront, M.; Gutierrez, A.L.; Prudhomme, C.; Kopp, S.; Ames, D.P.; Williams, G.P.; Jones, N.L. Advancing Global Hydrologic Modeling with the GEOGloWS ECMWF Streamflow Service. *J. Flood Risk Manag.* **2025**, *18*, e12859. [\[CrossRef\]](#)
34. Jian, J.; Ryu, D.; Costelloe, J.F.; Su, C.-H. Towards Hydrological Model Calibration Using River Level Measurements. *J. Hydrol. Reg. Stud.* **2017**, *10*, 95–109. [\[CrossRef\]](#)
35. Mohamoud, Y.M. Prediction of Daily Flow Duration Curves and Streamflow for Ungauged Catchments Using Regional Flow Duration Curves. *Hydrol. Sci. J.* **2008**, *53*, 706–724. [\[CrossRef\]](#)
36. Baaniya, Y. Empowering Local Water Resources Managers with Global Hydrological Model Information Through Bias Correction of Their In-Situ Data Using SABER. Master's Thesis, Brigham Young University, Provo, UT, USA, 2025.

37. Lin, P.; Pan, M.; Beck, H.E.; Yang, Y.; Yamazaki, D.; Frasson, R.; David, C.H.; Durand, M.; Pavelsky, T.M.; Allen, G.H.; et al. Global Reconstruction of Naturalized River Flows at 2.94 Million Reaches. *Water Resour. Res.* **2019**, *55*, 6499–6516. [\[CrossRef\]](#)
38. Hersbach, H.; Bell, B.; Berrisford, P.; Biavati, G.; Horányi, A.; Muñoz Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Rozum, I.; et al. *ERA5 Hourly Data on Single Levels from 1940 to Present*; Copernicus Climate Change Service (C3S) Climate Data Store (CDS): Reading, UK, 2018. [\[CrossRef\]](#)
39. Beck, H.E.; De Roo, A.; Van Dijk, A.I.J.M. Global Maps of Streamflow Characteristics Based on Observations from Several Thousand Catchments. *J. Hydrometeorol.* **2015**, *16*, 1478–1501. [\[CrossRef\]](#)
40. Hales, R.C.; Sowby, R.B.; Williams, G.P.; Nelson, E.J.; Ames, D.P.; Dundas, J.B.; Ogden, J. SABER: A Model-Agnostic Postprocessor for Bias Correcting Discharge from Large Hydrologic Models. *Hydrology* **2022**, *9*, 113. [\[CrossRef\]](#)
41. Lehner, B.; Verdin, K.; Jarvis, A. New Global Hydrography Derived from Spaceborne Elevation Data. *EoS Trans.* **2008**, *89*, 93–94. [\[CrossRef\]](#)
42. Carlson, K.A.; Levin, H.K.; Morris, A.L.; Candela, S.G.; Rivera, A.M.M.; Huening, V.G.; Fredericks, J.G. TDX-Hydro: Global High-Resolution Hydrography Derived from TanDEM-X. *Authorea Preprints* **2024**. [\[CrossRef\]](#)
43. Snow, A.D.; Christensen, S.D.; Swain, N.R.; Nelson, E.J.; Ames, D.P.; Jones, N.L.; Ding, D.; Noman, N.S.; David, C.H.; Pappenberger, F.; et al. A High-Resolution National-Scale Hydrologic Forecast System from a Global Ensemble Land Surface Model. *J. Am. Water Resour. Assoc.* **2016**, *52*, 950–964. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Qiao, X.; Nelson, E.J.; Ames, D.P.; Li, Z.; David, C.H.; Williams, G.P.; Roberts, W.; Sánchez Lozano, J.L.; Edwards, C.; Souffront, M.; et al. A Systems Approach to Routing Global Gridded Runoff through Local High-Resolution Stream Networks for Flood Early Warning Systems. *Environ. Model. Softw.* **2019**, *120*, 104501. [\[CrossRef\]](#)
45. Rojas Lesmes, D.J.; Sanchez Lozano, J.L.; Nelson, J.; Trujillo Vela, M.G.; Williams, G.; Ames, D.P.; Jones, N.; Gutierrez, A.L.; Romero Bustamante, E.G.; Hales, R. Discharge-to-Water Level Transformation (DWLT) Using Monthly Duration Curves: Enhancing the Utility of the GEOGLOWS ECMWF Hydrological Model. *ESS Open Arch.* **2024**. [\[CrossRef\]](#)
46. Chancay, J.E.; Espitia-Sarmiento, E.F. Improving Hourly Precipitation Estimates for Flash Flood Modeling in Data-Scarce Andean-Amazon Basins: An Integrative Framework Based on Machine Learning and Multiple Remotely Sensed Data. *Remote Sens.* **2021**, *13*, 4446. [\[CrossRef\]](#)
47. Berg, P.; Bosshard, T.; Bozhinova, D.; Bähring, L.; Löw, J.; Nilsson, C.; Strandberg, G.; Södling, J.; Thuresson, J.; Wilcke, R.; et al. Robust Handling of Extremes in Quantile Mapping—“Murder Your Darlings”. *Geosci. Model Dev.* **2024**, *17*, 8173–8179. [\[CrossRef\]](#)
48. Liu, D. A Rational Performance Criterion for Hydrological Model. *J. Hydrol.* **2020**, *590*, 125488. [\[CrossRef\]](#)
49. Pool, S.; Vis, M.; Seibert, J. Evaluating Model Performance: Towards a Non-Parametric Variant of the Kling-Gupta Efficiency. *Hydrol. Sci. J.* **2018**, *63*, 1941–1953. [\[CrossRef\]](#)
50. Kling, H.; Fuchs, M.; Paulin, M. Runoff Conditions in the Upper Danube Basin under an Ensemble of Climate Change Scenarios. *J. Hydrol.* **2012**, *424–425*, 264–277. [\[CrossRef\]](#)
51. Knoben, W.J.M.; Freer, J.E.; Woods, R.A. Technical Note: Inherent Benchmark or Not? Comparing Nash–Sutcliffe and Kling–Gupta Efficiency Scores. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 4323–4331. [\[CrossRef\]](#)
52. Rey, D.; Neuhauser, M. Wilcoxon-Signed-Rank Test. In *International Encyclopedia of Statistical Science*; Lovric, M., Ed.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 1658–1659. ISBN 978-3-642-04897-5.
53. Wilcoxon, F. Individual Comparisons by Ranking Methods. *Biom. Bull.* **1945**, *1*, 80. [\[CrossRef\]](#)
54. Zhao, W.; Li, A. A Review on Land Surface Processes Modelling over Complex Terrain. *Adv. Meteorol.* **2015**, *2015*, 607181. [\[CrossRef\]](#)
55. Nogueira, M. Inter-Comparison of ERA-5, ERA-Interim and GPCP Rainfall over the Last 40 Years: Process-Based Analysis of Systematic and Random Differences. *J. Hydrol.* **2020**, *583*, 124632. [\[CrossRef\]](#)
56. Robaina, L.; Campoazano, L.; Villacís, M.; Rehbein, A. Mesoscale Convective Systems over Ecuador: Climatology, Trends and Teleconnections. *Atmosphere* **2025**, *16*, 1157. [\[CrossRef\]](#)
57. Lavers, D.A.; Simmons, A.; Vamborg, F.; Rodwell, M.J. An Evaluation of ERA5 Precipitation for Climate Monitoring. *Q. J. R. Meteorol. Soc.* **2022**, *148*, 3152–3165. [\[CrossRef\]](#)
58. Tarek, M.; Brissette, F.P.; Arsenault, R. Evaluation of the ERA5 Reanalysis as a Potential Reference Dataset for Hydrological Modelling over North America. *Hydrol. Earth Syst. Sci.* **2020**, *24*, 2527–2544. [\[CrossRef\]](#)
59. Fang, K.; Chen, F.; Sen, A.K.; Davi, N.; Huang, W.; Li, J.; Seppä, H. Hydroclimate Variations in Central and Monsoonal Asia over the Past 700 Years. *PLoS ONE* **2014**, *9*, e102751. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Teng, J.; Bennett, J.C.; Charles, S.; Chiew, F.; Ji, F.; Potter, N.; Fu, G.; Thatcher, M.; Remenyi, T. Trend and Inter-Annual Variability in Regional Climate Models—Validation and Hydrological Implications in Southeast Australia. *J. Hydrol.* **2024**, *642*, 131817. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.