https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

Large-scale streamflow regionalization in ungauged West African catchments: How do classical and deep learning approaches compare?

Serigne Bassirou Diop* [1],
Fadilath Kate* [2,3],
Issam Souassi* [2,4],
Ansoumana Bodian [1],
Bastien Dieppois [5],
Joris Guerin [2],
Renaud Hostache [2],
Anne Johannet [6],
Frederik Kratzert [7],
Ludovic Oudin [8],
Vianney Sivelle [6],
Kalil Traore [9],
Yves Tramblay [2]

[1] Laboratoire Leïdi "Dynamique des Territoires et Développement", Univ. Gaston Berger, Saint-Louis, Senegal
[2] Espace-Dev, Univ. Montpellier, IRD, Montpellier, France
[3] RiverLy, INRAE, Villeurbanne, France
[4] Univ. Waterloo, Canada
[5] Centre for Agroecology, Water and Resilience, Coventry University, Coventry, UK
[6] HSM, Univ Montpellier, IMT Mines Ales, CNRS, IRD, Montpellier, France
[7] Google Research, 1010 Vienna, Austria
[8] Sorbonne Université, Université PSL, EPHE, CNRS, Milieux Environnementaux, Transferts et Interactions dans les hydrosystèmes et les Sols, METIS, F-75005 Paris, France
[9] Synapse Informatique, Montpellier, France
* Contributed equally to the study

Correspondence to: Serigne Bassirou Diop (diop.serigne-bassirou@ugb.edu.sn)

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.



Abstract

In West Africa, limited access to hydrometric data remains a major challenge for advancing surface water research and improving water management. Since the early 1980s, many gauging stations have been decommissioned, leaving gaps in reliable streamflow records. Hydrological model regionalization is widely used to estimate runoff and flood quantiles in ungauged catchments. This study presents a region-wide assessment of rainfall–runoff model regionalization across West Africa. We used an unprecedented dataset of 189 near-natural catchments to compare two contrasting approaches: (i) a benchmark conceptual modeling framework using the GR4J (**G**énie **R**ural à **4** paramètres **J**ournalier) model, regionalized with three parameter-transfer techniques (spatial proximity, physiographic similarity, and random forest), and (ii) a data-driven framework based on Long Short-Term Memory (LSTM) neural networks. Using a leave-one-out resampling strategy, the regionalization approaches were evaluated using multiple performance metrics: (i) the Kling–Gupta Efficiency (KGE), computed between simulated and observed streamflow, (ii) the relative bias (rBias) of several hydrological signatures, and (iii) differences between observed and simulated flood quantiles. Results show that the GR4J-based approaches consistently underperform compared to the LSTM, failing to reproduce key hydrological signatures. In contrast, the LSTM model demonstrated better generalization performance, simulating streamflow with a median KGE of 0.67 and reliably capturing hydrological signatures and flood quantiles across West Africa's diverse climates with lower biases. These findings highlight the potential of data-driven approaches to advance hydrological modelling and support flood risk management, water resource planning, and climate adaptation.

Key words: West Africa, regionalization, hydrological signatures, GR4J, LSTM

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

1 Introduction

Reliable estimation of streamflow remains a major global challenge in surface hydrology, as river discharge information is essential for sustainable water resources planning and for modelling hydrological extremes such as floods and droughts. Because it is impossible to explicitly represent all physical processes and interactions occurring within a catchment, hydrological models provide simplified representations of complex catchment behaviour (Devia et al., 2015). These models rely on parameters that are typically estimated through calibration, which compensates for limited process understanding and improves the accuracy of simulated streamflow (Bruneau et al., 1995; Gupta et al., 2009). However, this calibration requires observed flow data, which are often unavailable in data-scarce regions. To overcome this limitation, regionalization techniques have been developed to transfer model parameters from gauged to ungauged basins that are geographically close and/or share similar physiographic and climatic characteristics (Hrachowitz et al., 2013). Regionalization thus provides a means to estimate streamflow in ungauged catchments, one of the most persistent challenges in catchment hydrology (Blöschl et al., 2019; Salinas et al., 2013).

In West Africa, twenty-eight transboundary river basins connect nearly all countries in the region, with most sharing at least one watercourse with neighbouring states (OECD/SWAC, 2009). The region's socioeconomic development, including agriculture, energy production, and livelihoods, depends heavily on the water resources supplied by these shared basins and aquifers (Bonzanigo et al., 2021). Unfortunately, this wealth of water resources is increasingly threatened by climate change (IPCC, 2021). The strong reliance on rainfed agriculture and the close ties of rural communities to the natural environment further exacerbate the region's vulnerability (Diallo et al., 2020; Krishnamurthy et al., 2012; Padgham et al., 2015; van der Land et al., 2018), which is aggravated by a rapid demographic expansion (UN, 2019). The 1970s and 1980s were characterized by significant shifts in precipitation (e.g., Le Barbé and Lebel, 1997; Le Barbé et al., 2002; Mahé and Paturel, 2009; Nicholson, 2018), resulting in notable hydrological impacts, as the region's hydrological regime is largely precipitation driven (Ndehedehe et al., 2017). For instance, the Sahel region experienced an unprecedented drought, with severe rainfall deficits in many areas. The two largest transboundary river systems in the region (i.e., the Niger and Senegal basins) also suffered substantial reductions in streamflow, adversely affecting agriculture and energy production (Wantzen, 2023).

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

Numerous other catchments similarly experienced declines in runoff and diminished groundwater recharge, all of which exacerbated water insecurity (Descroix et al., 2018). Under climate- and human-driven environmental change, effective management of water resources is essential for achieving sustainable development in West Africa. However, as in many developing regions, hydroclimatic data remain difficult to access and are generally very limited (Tramblay et al., 2021). A large proportion of river basins in the region are ungauged or poorly gauged, and only a few streamflow stations provide publicly available data, often covering short or outdated periods (Schuol et al., 2008; Tramblay et al., 2021). This lack of streamflow observations constrains the direct application of rainfall-runoff models. Consequently, despite the strategic importance of the region's hydrosystems, the capacity to monitor and model water dynamics remains limited. Significant knowledge gaps persist regarding large-scale hydrology and the impacts of climate variability, largely due to sparse observational networks and a lack of up-to-date hydrological data (Ndehedehe, 2019).

Several studies have investigated the regionalization of rainfall-runoff models worldwide, highlighting that the effectiveness of different methods strongly depends on catchment characteristics, data availability, and regional context (Arsenault et al., 2019; Golian et al., 2021; Kuana et al., 2024; Kupzig et al., 2024; Oudin et al., 2008; Parajka et al., 2005; Saadi et al., 2019; Yang et al., 2019). Despite these efforts, there remains a lack of comprehensive evaluations of how various regionalization approaches perform in data-scarce regions. In West Africa, although a few studies have explored regionalization, they are generally limited to the catchment scale and a small number of catchments. This limitation is particularly critical given the region's hydroclimatic variability and sparse observational networks. For instance, Ibrahim et al. (2015) applied regionalization in the Volta basin but focused on only three gauged catchments, while Kouakou et al. (2023) assessed regionalization across nineteen major basins across West and Central Africa. This study aims to address this gap by systematically assessing the performance of multiple regionalization techniques for West African catchments, using an unprecedented hydrometric database.

Due to their simplicity and computational efficiency, conceptual hydrological models have been widely applied in West Africa (Bodian et al., 2016; Koubodana et al., 2021; Kwakye and Bárdossy, 2020; Niel et al., 2003). However, regional applications of such models face several challenges, including the limited availability of calibration data, the uncertainty in transferring parameters between catchments (Blöschl et al., 2019), and the difficulty of representing spatial

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

heterogeneity and climate variability (Pokhrel et al., 2008). Recent advances in data-driven methods, such as Recurrent Neural Networks (RNNs) open new avenues for streamflow estimation in ungauged basins. The Long Short-Term Memory (LSTM) network, in particular, has shown strong potential for hydrological time series simulation (Kratzert et al., 2018, 2019c) and streamflow estimation in ungauged basins (Choi et al., 2022; Kratzert et al., 2019b). Nonetheless, these models also have limitations: they require large datasets, and their predictive skill can degrade under data-scarce conditions such as those found in West Africa. To the best of our knowledge, no study has yet applied the LSTM framework specifically in West Africa. This study aims to address existing gaps by evaluating classical, regression-based, and purely data-driven approaches to hydrological model regionalization, comparing the conceptual modelling approach with an LSTM-based regional framework. It is important to clarify that this comparison evaluates two distinct modeling paradigms rather than a simple one-to-one model structure comparison. The first paradigm, represented by the GR4J model, follows a traditional parameter-transfer framework. In this approach, the model is locally calibrated, and its parameters are subsequently transferred to ungauged locations via explicit methods such as physical similarity or regression. The second paradigm, represented by the LSTM, employs a simultaneous regional learning framework. Unlike the two-step process of GR4J, the LSTM learns an implicit regionalization by training a single global model on the pooled data of all available basins, mapping meteorological forcing and static catchment attributes directly to streamflow. Consequently, the performance differences observed in this study reflect the combined impact of both the underlying model architecture (conceptual vs. deep learning) and the inherent nature of the regionalization strategy (explicit transfer vs. implicit learning). Furthermore, appraisal studies of regionalization and deep-learning methodologies in hydrology are frequently carried out based on rich databases characterized by a large number of stations and long, uninterrupted time series (Kratzert et al., 2019b; Wang et al., 2026). However, very little effort has been dedicated to evaluating the robustness of these regional estimation procedures when applied under less-than-ideal conditions, such as a limited number of gauged sites, low spatial density of stations, short or fragmented time series, and incomplete covariates. This study addresses this limitation by conducting a large-sample assessment in West Africa, a region where regionalization is needed but historically constrained by data scarcity. Beyond a local application, the results of this study offer generalizable insights into the limits and capabilities of both conceptual and data-driven learning paradigms when transferred to under-represented, data-sparse regions.

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

The remainder of this paper is organized as follows. Section 2 presents the study area and datasets used. Section 3 describes the materials and methods, including the hydrological models and the regionalization approaches evaluated. The results and their discussion are provided in Section 4, while Section 5 concludes the paper with key findings and perspectives for future research.

2 Study area and data

2.1 Study area description

West Africa covers more than one quarter of the African continent (OECD/SWAC, 2009), extending from the Atlantic coast of Senegal (18° W) to eastern Chad (25° E) and from the Gulf of Guinea (4° N) to the Sahel (25° N). It spans several climatic zones shaped by the seasonal shifts of the Inter-Tropical Convergence Zone (ITCZ; e.g., Thorncroft et al., 2011). These range from the arid Sahel, where annual rainfall rarely exceeds 600 mm, to the humid Guinean domain, characterized by steep topography and annual precipitation reaching up to 2200 mm (Orange, 1990). This climatic gradient results in a large diversity of hydrological regimes and water networks across the region.

2.2 Observational hydrometric data

We use daily streamflow data from the unprecedented African Database of Hydrometric Indices (ADHI; Tramblay et al., 2021). The ADHI database provides hydrological indices derived from multiple sources, including 1,466 gauging stations with daily streamflow records spanning at least 10 years between 1950 and 2018. In West Africa, the database comprises 441 gauging stations across several countries, with catchment areas ranging from 95 km² to 2,150,000 km². Following (Diop et al., 2025), we selected catchments meeting three criteria: (i) minimal hydrological regulation, assessed through visual inspection of dam locations and annual hydrograph analysis to evaluate potential infrastructure impacts; (ii) catchment areas up to 150,000 km² to exclude large river systems, where flow dynamics are strongly influenced by spatial heterogeneity and routing effects.; and (iii) daily streamflow records covering at least 10 years during 1975-2018. This resulted in 189 near-natural catchments, covering a substantial portion of West Africa (Figure 1), with catchment areas ranging from 95 km² to 140,000 km²

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

and maximum and mean record lengths of 39 and 18 years, respectively. Missing values were present in a subset of the streamflow records. These gaps were not infilled to avoid introducing additional uncertainty. Instead, time steps with missing observations were excluded during both GR4J calibration and LSTM training. Given the relatively low proportion of missing data, this approach ensures robust model estimation while avoiding artifacts associated with gap-filling procedures.

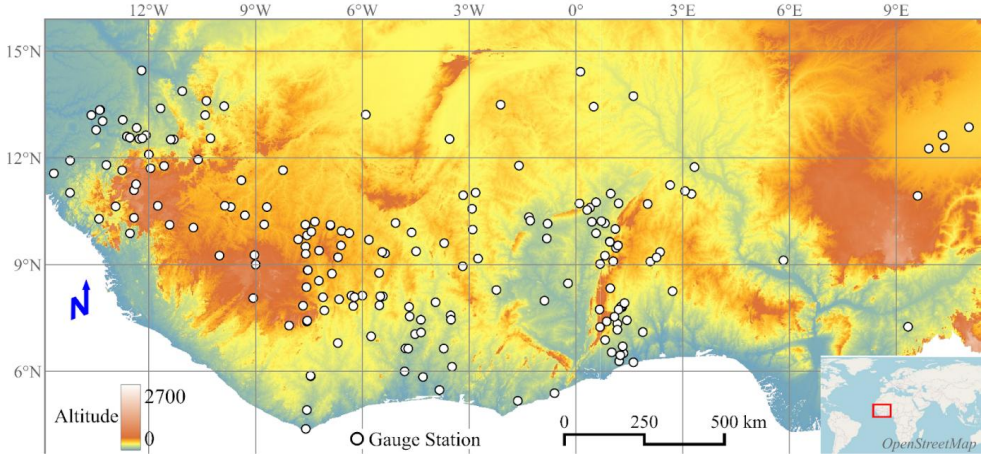


Figure 1: Spatial distribution of the 189 hydrometric stations used in this study. The color shading illustrates topographic elevation, where darker blue tones represent low-lying areas.

2.3 Catchment attributes used for regionalization

Table S1 summarizes the catchment descriptors available in the ADHI database, including their definitions, abbreviations, and sources. From this dataset, we selected a subset of physiographic and climatic attributes to characterize catchment properties and support regionalization analyses. These include catchment area (Area), maximum altitude (MA), precipitation (Prec), potential evapotranspiration (PET), groundwater productivity (GP), groundwater storage (GS), groundwater depth (GD), and available water capacity (AWC). It is important to note that the selection of specific features for the analysis is designed to provide a broad characterization of catchment properties rather than a priori prioritization, ensuring that the dataset encompasses most physiographical factors that could influence catchment behavior. Because some of the methods used in this study, particularly distance-based approaches and gradient-based models such as LSTM, are sensitive to the scale of input variables, all selected attributes were normalized prior to any analysis.

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

2.4 Meteorological forcings

Hydrological models were calibrated and driven using daily precipitation and potential evapotranspiration (PET) derived from the ERA5 reanalysis dataset for the period 1979-2020 (https://doi.org/10.24381/cds.adbb2d47; last access: 13 November 2025). The ERA5 dataset is a global atmospheric reanalysis developed by the Copernicus Climate Change Service at the European Centre for Medium-Range Weather Forecasts (ECMWF). It represents the fifth generation of atmospheric reanalyses, based on four-dimensional variational data assimilation using cycle 41r2 of the ECMWF Integrated Forecasting System (Hersbach et al., 2020). ERA5 supersedes the ERA-Interim reanalysis (Dee et al., 2011), with a spatial resolution of 0.25° × 0.25° and an hourly temporal resolution. PET was extracted directly from ERA5 rather than estimated using empirical equations. Hourly precipitation and PET fields were aggregated to daily values and subsequently spatially averaged over each catchment using the corresponding catchment boundaries. This procedure yielded a lumped daily forcing time series representative of each catchment, consistent with the lumped structure of the GR4J model and the input requirements of the LSTM.

3 Materials and methods

3.1 Description of the hydrological models

3.1.1 GR4J model

The GR4J model is a daily lumped, four-parameter rainfall-runoff model developed by Perrin et al. (2003). It simulates the hydrological behavior of a catchment using a conceptual structure with two main reservoirs: a production store and a routing store. These reservoirs act as filters that partition and process rainfall into two distinct flow components: a fast response, representing rapid surface runoff, and a slow response, corresponding to delayed contributions such as baseflow. The model requires calibration of four parameters: the capacity of the production store (X1), intercatchment exchange coefficient (X2), the capacity of the nonlinear routing store (X3), and the characteristic response time of the basin represented by the unit hydrograph (X4). Figure S1 shows the conceptual scheme of the GR4J model. The GR4J model

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

was selected as the conceptual representative for this study due to its parsimonious structure, which is particularly well-suited for diverse climate and data-scarce regions (Cantoni et al., 2022; Perrin et al., 2003). Its structural simplicity reduces the risks of overfitting, common issues when modeling catchments with limited observational data. The GR4J is commonly used to test parameter transfer methods (Arsenault et al., 2019; Golian et al., 2021; Mihret et al., 2025; Oudin et al., 2025; Yang et al., 2020), allowing for a comparison of our results with the existing literature. Furthermore, as the primary goal of this study is to contrast a classical conceptual approach (GR4J with parameter transfer) against a deep learning approach (LSTM), the GR4J serves as a robust baseline to evaluate the performance gains provided by data-driven regional models in ungauged catchments.

The GR4J parameters optimization was performed at each gauged station using the Michel algorithm (Michel, 1991), and the Kling-Gupta Efficiency (KGE; Gupta et al., 2009) as the objective function. The KGE (Equation 1) is a decomposition of the Nash-Sutcliffe Efficiency which places greater importance on high values (Nash and Sutcliffe, 1970), as it jointly accounts for correlation (r), bias (β), and variability (α). KGE values range from -∞ to 1, with 1 indicating perfect agreement with observations (Knoben et al., 2019). We computed these two metrics between observed daily streamflow series and those simulated by the regionalized hydrological models.

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

From Equations (1), r is the linear Pearson correlation between simulated and observed streamflow, α represents the variability term (ratio of standard deviations), and β is the bias ratio of mean flows.

Calibration was driven by ERA5 reanalysis meteorological forcings. Since the objective of this study is to assess the overall hydrological performance under current conditions rather than the model's robustness to climatic variability, and given the heterogeneous length and temporal coverage of the available streamflow records, performance was assessed using the full set of available discharge observations at each station instead of splitting the data into separate calibration and validation periods (Shen et al., 2022).

It is important to note that, although a temporal split-sample test was not conducted for GR4J model calibration, the regionalization procedure follows a jackknife resampling or leave-one-out cross-validation framework. In this approach, model performance for "ungauged"

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

catchments is systematically evaluated on catchments that are entirely excluded from the parameter-transfer process, providing a robust assessment of the spatial generalizability of the regionalization methods across the diverse hydro-climatic conditions of West Africa. Nevertheless, while spatial transferability is rigorously tested, the robustness of these models under non-stationary climatic conditions or future climate scenarios remains an area for further investigation.

3.1.2 LSTM model

The LSTM model used in this study is based on the equally named recurrent neural network architecture Long Short-Term Memory by Hochreiter and Schmidhuber (1997). The LSTM was established by Kratzert et al. (2018, 2019c) for hydrological applications to simulate streamflow from meteorological inputs at catchment scale. We used NeuralHydrology by (Kratzert et al., 2022) for training and evaluating the model. Unlike conceptual models, the LSTM model does not rely on predefined physical postulate but instead learns temporal dependencies directly from data through memory cells and gates, allowing it to capture both short- and long-term hydrological processes and to construct more flexible transfer functions for simulating the rainfall-runoff relationship.

The LSTM model is trained using both dynamic meteorological forcings and static catchment attributes (Area, MA, Prec, PET, GS, GP, GD, and AWC), enabling it to learn general hydrological behaviors across regions. To account for differences in catchment characteristics, static attributes such as elevation, land use, or soil type are integrated via an embedding layer (Kratzert et al., 2018). This allows the LSTM to learn the rainfall-runoff relationship by integrating both static variables (e.g., catchment attributes) and dynamic variables (e.g., rainfall, evapotranspiration). The LSTM model weights were optimized using gradient-based learning algorithms (e.g., Adam) and trained with a regionalized objective function. To account for the diverse hydrological scales across the 189 catchments, we adopted the catchment-averaged NSE* loss (Equation 2), as introduced by (Kratzert et al., 2019b), which prevents catchments with high absolute discharge from dominating the training process.

$$NSE^{*} = \frac{1}{B}\sum_{b=1}^{B}\sum_{n=1}^{N}\frac{(\hat{y}_n - y_n)^2}{(S(b)+\epsilon)^2} \quad (2)$$

Where B and N are respectively the number of basins and the number of time steps per basin b , $\hat{y}$ and y are the simulated and observed discharge at time step n for basin b, $S(b)$ is the

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.



standard deviation of observed discharge computed over the training period in basin b, and ϵ (with $\epsilon = 0.1$) is a constant term representing the variance of the observations.

For all simulations, the same hyperparameters as in Kratzert et al. (2019b) were adopted to ensure consistency and comparability with previous studies. A preliminary sensitivity analysis to optimize the hyperparameters (Figures S8 and S9) did not significantly improve performance compared to the hyperparameter values of Kratzert et al. (2019b). Table S2 provides the sets of hyperparameters from Kratzert et al. (2019b) along with the optimization search performed using the Optuna framework (https://github.com/optuna/optuna; last access: 13 November 2025). A temporal training/validation split was used to guide the hyperparameter tuning of the LSTM model. The data-driven nature of the LSTM, with a large number of trainable parameters, makes it susceptible to overfitting. Separating the data into training, validation, and testing periods ensures robust hyperparameter tuning and assessment of the model's predictive performance. Within each iteration, the available time series were restricted to the period between the first and last observed daily discharge values for each catchment. The dataset was then divided chronologically, with the last 25% reserved as an independent test set for final evaluation, while the remaining 75% was used for model development. Hyperparameter tuning and model stability assessment were conducted within this training portion using 5-fold cross-validation (K = 5). For a given hyperparameter configuration, five simulations were performed using different training-validation splits to evaluate the consistency of model performance across periods. In addition, the final model was trained using the full training dataset without a validation subset and was subsequently used for testing.

3.2 Regionalization approaches

3.2.1 Regionalization with the GR4J

We evaluated three regionalization approaches to transfer GR4J parameter sets from gauged to ungauged catchments: (i) spatial proximity, (ii) physiographic similarity, and (iii) a regression-based approach. Each method was tested using a jackknife or a leave-one-out resampling strategy, in which each of the 189 gauged stations was alternately treated as ungauged. This cross-validation strategy allowed for a site-specific evaluation of regionalization performance. The leave-one-out validation is commonly used in hydrological regionalization studies (Kim et al., 2026; Martel et al., 2011; Ouarda et al., 2001; Sarhadi and Modarres, 2011; Shu and

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

Ouarda, 2008; Tramblay et al., 2024), as it maximizes the use of available gauged catchments while providing an unbiased estimate of prediction skill for ungauged locations within the study domain. However, this approach may introduce relatively high variability due to its reliance on single-point testing. It may also yield optimistic performance estimates compared with more restrictive validation frameworks, since each prediction benefits from nearly the entire regional dataset. In practical situations where performance is assessed simultaneously at multiple catchments, prediction skill may therefore be lower. Nevertheless, the jackknife procedure remains a useful benchmark for evaluating the transferability of model parameters across the catchment network.

The spatial proximity approach identifies donor catchments based on geographical closeness. For each target site, Euclidean distances were computed between the ungauged catchment and all gauged stations. Parameters from the n nearest neighbors (where n ranged from 1 to 25) were transferred, and streamflows were simulated by averaging outputs across donors. The KGE was calculated for each configuration, and the optimal number of donor catchments was defined as that yielding the highest KGE. The regional optimum was obtained for five neighbors, which maximized the median KGE across all catchments (Figure S2).

The physiographic similarity approach selects donors based on catchment descriptors (Table S1). To reduce dimensionality and avoid redundancy among variables, a Principal Component Analysis (PCA; Naik, 2018) was applied to the physiographic dataset. Euclidean distances were then computed in the PCA-transformed space to identify the n most similar donor catchments for each target site. As with the spatial proximity approach, parameters were transferred from the selected donors and streamflows were simulated by averaging model outputs. Model performance was evaluated using KGE for different numbers of donor catchments, and the optimal configuration corresponded to five donor catchments as for the spatial proximity (Figure S2).

The regression-based approach is implemented using the random forest (RF) algorithm, an ensemble learning method built upon bagged decision trees (Breiman, 2001). RF introduces randomness through (i) bootstrapped sampling of the training data and (ii) random feature selection at each split, which reduces variance and mitigates overfitting. By averaging predictions across multiple decorrelated trees, RF improves both accuracy and generalization, making it particularly well-suited for modeling non-linear relationships between physiographic predictors and hydrological parameters (Schonlau and Zou, 2020). An independent RF model

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

was trained for each target catchment using data from the remaining 188 gauged basins and subsequently applied to predict the GR4J parameters. Model development involved hyperparameter tuning via randomized search cross-validation implemented in the scikit-learn framework (https://scikit-learn.org/stable/; last access: 13 November 2025), such as the number of trees (n_estimators), minimum samples per split (min_samples_split), minimum samples per leaf (min_samples_leaf), and maximum number of leaf nodes (max_leaf_nodes). Predictor variables were ranked by importance, and the five most influential physiographic characteristics were ultimately retained to achieve the most accurate and robust predictions.

3.2.2 Regionalization with the LSTM

A regionalization approach has also been implemented using the LSTM model within the NeuralHydrology framework (Kratzert et al., 2022). Similar to the GR4J model, we applied a jackknife or a leave-one-out resampling strategy, in which each catchment was iteratively fixed as the test set, while the remaining catchments (i.e., 188 gauges) served as the training set. Consequently, a fundamental operational difference arises between the two frameworks: while the GR4J calibration is performed independently for each catchment, the regional nature of the LSTM means that a single catchment effectively contributes to multiple calibration simulations (specifically, up to (N - 1) times within the leave-one-out loop). Unlike the traditional approach to regionalizing rainfall-runoff model parameters, which typically involves calibrating models on a basin-to-basin basis, the LSTM model is trained simultaneously across multiple sites. Consequently, the LSTM configuration used here is inherently regional, enabling runoff simulation at multiple sites simultaneously. While conceptual models can also be calibrated regionally through predefined functional relationships between catchment attributes and model parameters (e.g., Samaniego et al., 2010), the LSTM framework differs in that it learns these transfer relationships directly from data, without requiring a priori functional assumptions. Recent global deep-learning rainfall-runoff models include some West African catchments (e.g., Nearing et al., 2024); however, these approaches rely on globally pooled training datasets and are not tailored to region-specific hydrological dynamics. Here, we conduct a regionally focused assessment of LSTM model transferability across West Africa.

3.3 Assessment of regionalization approaches

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

3.3.1 Evaluation of the accuracy of regional streamflow simulations

We used two dimensionless criteria to evaluate the performance of the regionalization approaches in simulating daily streamflows: (i) the KGE (Equation 1), and (ii) the relative bias (rBias; Equation 3), which quantifies systematic over- or underestimation. These metrics were computed between observed daily streamflow series and those simulated by the regionalized hydrological models. In addition, we evaluated the ability of the regionalization methods to reproduce the seasonal cycle of streamflow, defined as the intra-annual variability of discharge, estimated using a 10-day moving average.

$$rBias = \frac{1}{n}\sum_{i=1}^{n}\frac{P_i - O_i}{O_i} \quad (3)$$

In Equation 3, *rBias* represents the relative bias, n represents the number of observation data, P_i and O_i are respectively the predicted and the actual hydrological signatures at the catchment i.

3.3.2 Evaluation of the performance of the regionalization approaches in reproducing hydrological signatures

The relative bias (rBias) was computed to quantify systematic over- or under-prediction by comparing key hydrological signatures derived from observed and simulated flows, including: (i) baseflow index, (ii) baseflow recession constant, (iii) flashiness index, and (iv) runoff coefficient. Hydrologic signatures (also referred to as hydrologic metrics, indices, or diagnostic signatures) are quantitative indicators that describe key statistical and dynamic characteristics of streamflow variability and behavior (McMillan, 2021). The baseflow index (BFI) represents the proportion of total streamflow derived from groundwater or subsurface sources (Kunkle, 1962; Ladson et al., 2013). The baseflow recession constant (BFC) quantifies the rate at which baseflow declines during rain-free periods (Gnann et al., 2021). The Richards-Baker Flashiness Index (FI) characterizes short-term streamflow variability, reflecting how rapidly runoff responds to rainfall events and returns to baseflow conditions (Baker et al., 2004). The runoff coefficient (RC) is a dimensionless metric indicating the fraction of precipitation converted into direct surface runoff (Mulvaney, 1851).

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

To investigate the potential physical drivers of prediction errors, we applied an unsupervised K-means clustering analysis to the residuals of the best-performing model. Catchments were grouped into three categories (low, moderate, and high error) based on the relative bias computed for the four hydrological signatures (BFI, BFR, FI, and RC). The distribution of key physiographic attributes (e.g., catchment area, aridity index, and groundwater depth), was then examined within each cluster to identify systematic relationships between physical catchment characteristics and model performance. Focusing on the best-performing model allows for a robust interpretation of residual errors, which are more likely to reflect underlying hydrological processes (e.g., small-catchment dynamics or aridity effects) rather than model structural limitations.

3.3.3 Evaluation of the performance of the regionalization approaches for estimating flood quantiles

Once regional simulations were obtained for each station using the three GR4J configurations and the LSTM framework, maximum annual flows were extracted using the block maxima approach (Gumbel, 1958). Subsequently, a local frequency analysis was performed on these series of annual maxima by fitting the Generalized Extreme Value (GEV) distribution. The fitting was performed using the Generalized Maximum Likelihood Estimator (GMLE), which has been identified as the most appropriate method for the West African context in recent studies (Diop et al., 2025). To test the feasibility of estimating flood quantiles at ungauged catchments, the 5-year (Q5) and 20-year (Q20) floods were estimated for each station using the locally fitted GEV distribution and compared with observations using the adimensional rBias metric defined in Equation (2). The GEV distribution is a continuous three-parameter probability distribution characterized by a location, scale, and shape parameter. The location parameter (μ) describes the central tendency of flood peaks, with higher values indicating a shift toward potentially more frequent or intense floods. The scale parameter (σ) reflects the variability of flood magnitudes, where larger values correspond to greater uncertainty and a wider range of extremes. The shape parameter (ξ) controls the tail behavior of the distribution, distinguishing among three types of extreme value distributions (Coles, 2001): (i) $\xi > 0$: Fréchet distribution with a heavy tail (Fréchet, 1927), suggesting a higher probability of extreme floods; (ii) $\xi = 0$: Gumbel distribution with a light tail (Gumbel, 1958); and (iii) $\xi < 0$: Weibull distribution with a bounded or short tail (Weibull, 1951), implying the existence of an upper bound and asymptotically zero probability beyond this upper bound. The shape parameter is

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

particularly important for evaluating the risk of rare events and designing infrastructure capable of withstanding extreme floods. The cumulative distribution function of the GEV is provided in Equation 4.

$$F(x;u,\alpha,\xi) = exp\left\{-\left[1-\xi\frac{x-u}{\alpha}\right]^{1-\xi}\right\} \quad \xi \neq 0$$
$$F(x;u,\alpha) = exp\left\{-exp\left[-\frac{x-u}{\alpha}\right]\right\} \quad \xi = 0 \tag{4}$$

Where x, u, α and ξ are the data, location, scale and shape parameters respectively, and $(u + \alpha/\xi) \leq x < \infty$ if $\xi < 0$; $-\infty < x < \infty$ if $\xi = 0$; $-\infty < x \leq (u + \alpha/\xi)$ if $\xi > 0$

The GMLE method was developed to overcome the limitations of the standard Maximum Likelihood Estimator when applied to small sample sizes. To this end, Martins & Stedinger (2000) proposed using a beta distribution with shape parameters p = 6 and q = 9 as a prior, which constrains the GEV shape parameter to the interval [-0.5,0.5] and thus prevents excessively negative values. To construct a more informative and context-specific prior, the same study further recommends refining this prior when regional information is available. In the present study, we adopt a normal distribution as the prior, which is best suited to the West African context and provides a more informative constraint on the GEV shape parameter (Diop et al., 2025).

4 Results and discussions

4.1 Accuracy of regional streamflow estimates compared to observations

The comparative evaluation of regionalization approaches for streamflow simulation in West African catchments reveals significant variability in model performance. As shown in Figure 2, physiographic similarity (Figure 2b) and the RF algorithm (Figure 2c) exhibited the weakest performance, with median KGE values of 0.26 and 0.25, respectively. Such a physiographic approach's inadequacy could stem from (i) its failure to comprehensively account for the combined effects of climatic variability and basin heterogeneity, and (ii) the limited relevance of used physical descriptors in explaining catchment hydrological behavior (Oudin et al., 2008). Similarly, despite its capacity to capture non-linear relationships, RF demonstrates limited generalizability across catchments (Figure 2c), which could be due to overfitting to local data and insufficient consideration of spatiotemporal dependencies (Luo and Su, 2025;

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

Saha et al., 2023). The spatial proximity-based regionalization approach yielded moderate improvements (Figure 2a), with a median KGE of 0.44. However, its performance remains uneven across basins, reflecting the limitation of relying solely on geographical closeness as neighboring catchments do not necessarily have comparable hydrological behavior (Wagener et al., 2004). Thus, this oversimplification may not be suitable in large regions with strong climatic gradients and heterogeneous catchment characteristics (Kupzig et al., 2024). It is also important to note that these findings are specific to the GR4J hydrological model: different model structures might potentially lead to different performance rankings (Hrachowitz et al., 2013).

In contrast, the LSTM model achieves the best performance (Figure 2d), yielding a median KGE of 0.67. Negative KGE values were rare (only 8%), and 67% of sites had KGE values greater than 0.5. Moreover, the LSTM exhibits a narrower interquartile range of KGE values (Figure 2e), reflecting both greater accuracy and more consistent performance across diverse hydrological regimes. This underscores the effectiveness of combining an architectural capacity for non-linear temporal dependencies with a simultaneous multi-site learning strategy, which provides an efficient framework for regional transferability compared to the tested GR4J parameter-transfer approaches. Nevertheless, despite its superior overall performance, the LSTM does not uniformly outperform traditional approaches across all catchments. As shown in Figure 2d, few localized cases of performance degradation are observed, where LSTM simulations are inferior to those obtained with GR4J regionalization methods. These discrepancies may arise from site-specific characteristics such as data scarcity or insufficient representation of local processes in the training data.

The superior performance of the LSTM model reflects the integrated strengths of its flexible framework and regional data pooling, allowing it to construct effective transfer functions across the network. These results align with recent advances in hydrological modeling, which consistently demonstrate the superior predictive capability and generalization of LSTM architectures over conventional regionalization approaches (Arsenault et al., 2023; Kratzert et al., 2019b; Martel et al., 2025). In addition, although LSTM models were traditionally viewed as black-box tools, a growing number of recent studies now demonstrate that their internal states, parameters, and attention mechanisms can be meaningfully interpreted, offering hydrologically consistent insights into the processes they represent (De la Fuente et al., 2024; Ding et al., 2020; Lees et al., 2022).

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

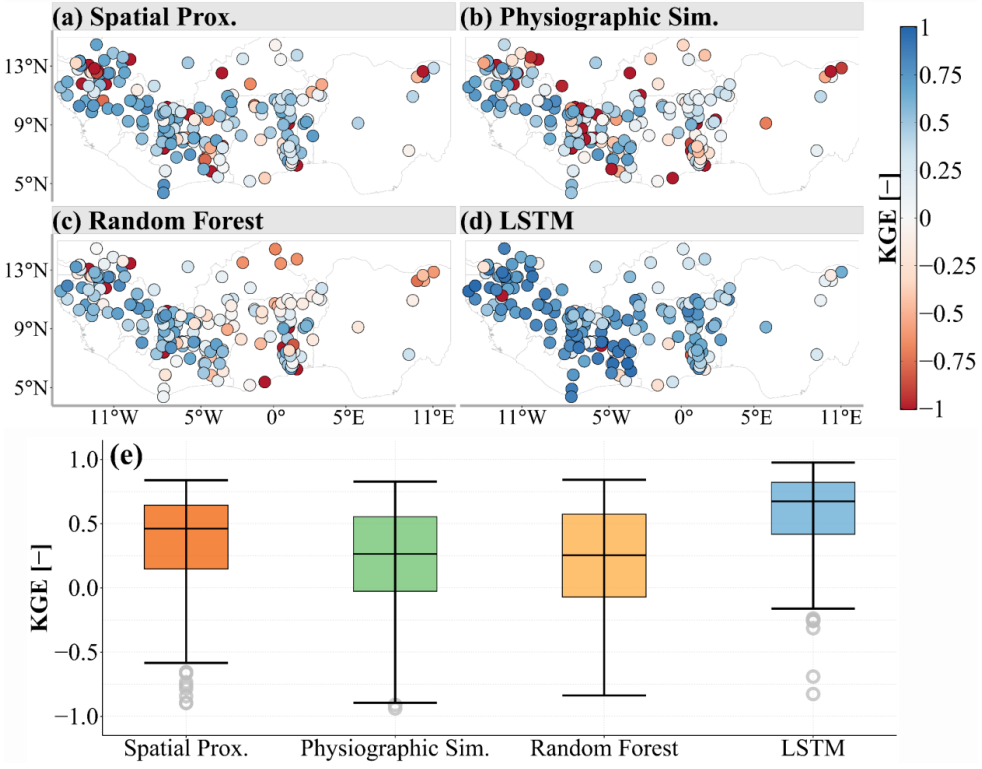


Figure 2: Spatial and statistical distribution of KGE for regional streamflow simulations across West African catchments using four regionalization approaches: (a) spatial proximity, (b) physiographic similarity (based on PCA), (c) random forest (RF), and (d) Long Short-Term Memory (LSTM). Each point represents a discharge station. The diverging color scale is centered at 0 (white), with blue tones indicating KGE > 0 and red tones indicating KGE < 0. Panel (e) shows the statistical distribution of the KGE values for regional streamflow simulations.

Figure 3 compares the observed and simulated seasonal streamflow cycles, derived using a 10-day moving average for the four regional estimation methods (Figure 3a), together with their associated 10-day running relative biases (Figure 3b). All approaches reproduce the shape and timing of the seasonal hydrograph. However, substantial differences emerge when examining the magnitude of flows and the distribution of biases throughout the annual cycle. The physiographic similarity method systematically overestimates the rising and falling limbs as well as the seasonal peak. In contrast, the GR4J-RF persistently underestimates streamflow across the entire hydrograph, resulting in a pronounced negative bias. The spatial proximity

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

method provides a close match to the observed seasonal cycle yielding overall lowest bias. The LSTM model also reproduces the general seasonal dynamics but exhibits more pronounced seasonal biases, including a slight overestimation during the rising limb and an underestimation during peak flow and the recession period.

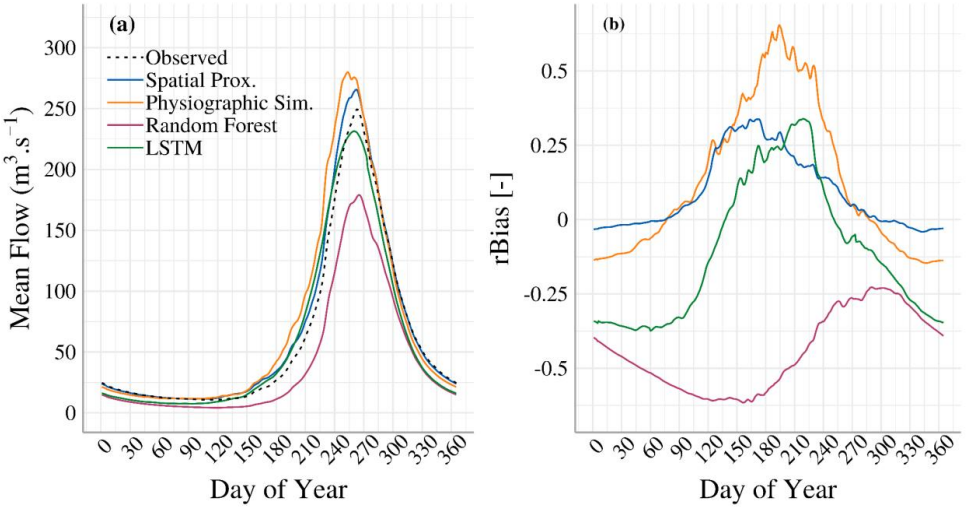


Figure 3: Average seasonal cycles of streamflow (a) computed using a 10-day moving average for observed and simulated data, and corresponding relative bias (b). Streamflow are simulated using different regionalization approaches of the GR4J model: (a) spatial proximity, (b) Physiographical similarity (PCA-based), (c) random forest (RF); and (d) an approach based on the Long Short-Term Memory (LSTM) structure.

While Figure 3 suggests that the RF-based approach substantially underestimates streamflow, Figure 2 shows that the same RF model and physiographic similarity yield comparable overall KGE distributions. To investigate this apparent inconsistency between the KGE distributions (Figure 2) and the differences observed in the seasonal hydrographs (Figure 3), we decomposed the KGE into its individual components: correlation (r), variability ratio (α), and bias ratio (β) (Figure S7). The results show that the RF approach achieves correlation (r) values comparable to those of the physiographic similarity method (Figure S7), indicating that temporal dynamics are generally well reproduced. However, RF exhibits a systematic negative bias (Figure S7), with median β value below unity ($\beta < 1$), reflecting a persistent underestimation of streamflow magnitude. In contrast, the physiographic similarity approach shows a bias ratio closer to one

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

(Figure S7), indicating a more accurate representation of flow magnitude on average. This analysis highlights how similar KGE values can arise from different combinations of error components. In the case of RF, the negative bias is partially compensated by strong correlation and moderate variability errors, resulting in KGE values comparable to those of other regionalization approaches. These findings highlight the importance of examining KGE components in addition to the aggregated metric

4.2 Modelling performance with respect to hydrological signatures

Traditional evaluations of hydrological models, which only rely on performance metrics computed from observed and simulated streamflow series, provide a limited understanding of why a model performs well or not. To overcome this limitation, (Gupta et al. (2008) proposed the use of hydrological signatures, which are quantitative indicators that characterize key aspects of catchment behavior and reflect underlying hydrological processes. These signatures offer a process-oriented perspective on model performance. Consequently, comparing observed and simulated signatures enables a more diagnostic evaluation of model behavior, revealing which hydrological processes are well represented and which are inadequately captured (Horner, 2020). Figure 4 compares observed and simulated streamflow-based signatures derived from regional simulations using the GR4J-based regionalization approaches (i.e., spatial proximity, physiographic similarity, and RF) and the data-driven LSTM model. While boxplots summarize the regional distributions, the per-catchment performance was further verified using observed-versus-simulated scatter plots for all evaluated hydrological signatures and models (Figure S5). Figure S3 shows the spatial distribution of the relative bias (rBias) in the estimation of these hydrological signatures. Overall, the LSTM model exhibits the highest consistency and accuracy across all signatures, closely reproducing observed distributions and producing spatially coherent estimates with lower biases (Figure S3d). In contrast, the GR4J-based methods display larger biases and variability (Figure 4a to 4c and Figure S3a to S3c),reflecting the operational challenges inherent to explicit parameter-transfer workflows when capturing complex hydrological variations across highly heterogeneous hydroclimatic gradients.

All methods tend to overestimate BFI (Figure 4a), although the magnitude of this bias varies across approaches. The LSTM provides the closest agreement with the observed distributions of both BFI and BFR, highlighting the effectiveness of its multi-site learning framework in

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

characterizing subsurface flow dynamics and groundwater-surface water interactions. The GR4J-based regionalization methods exhibit more pronounced structural biases, with a stronger overestimation of BFI and a systematic underestimation of BFR. (Figure 4a-b). This may reflect the known structural limitations of conceptual rainfall-runoff models forced only by atmospheric inputs in representing slow subsurface processes (Le Moine et al., 2007), limitations that regionalization tends to further propagate. Pronounced discrepancies also appear for FI (Figure 4c). Although the LSTM model underestimates the observed FI distribution, this underestimation is much lower than that of other approaches. Indeed GR4J-based approaches markedly underestimate it, highlighting a reduced capacity to capture the short-term flow variability driven by intense convective rainfall characteristic of West African climates. These systematic biases may, however, arise from a smoothing side-effect of averaging GR4J transfer-based simulations, which dampens high-frequency variability. Regarding RC (Figure 4d), while spatial proximity approaches show a median value close to observations, the LSTM provides a more robust representation of the overall distribution. As shown in the scatter plots (Figure S5, row 4), the LSTM simulations exhibit a tighter alignment with the 1:1 line and lower overall dispersion compared to spatial proximity, which shows higher variance (Figure S5, row 4). In contrast, the GR4J-RF generally underestimates RC, while the physiographic similarity method tends to overestimate it.

The performance of the LSTM model across the evaluated signatures highlights its capacity to capture baseflow dynamics and recession characteristics by learning complex, nonlinear relationships and temporal dependencies between rainfall, evapotranspiration, and catchment attributes, thereby capturing the memory and dynamics inherent in hydrological processes. Its multi-site training strategy enables spatial transfer learning, allowing the model to approximate shared hydrological patterns across catchments while also adapting to local variability. This approach provides greater flexibility compared to traditional regionalization methods, which rely on explicit parameter-transfer assumptions or predefined spatial and distance-based relationships (Oudin et al., 2008; Pagliero et al., 2019). These results align with recent state-of-the-art studies demonstrating the predictive capability and consistency of the LSTM model in climatically diverse regions (Arsenault et al., 2023; Kratzert et al., 2019b; Lees et al., 2021; Martel et al., 2025; Ye et al., 2024). In contrast, the lower performance of the tested GR4J-based regionalization approaches illustrates the practical challenges of parameter-transfer approaches in regions characterized by sharp climatic and physiographic contrasts (Yang et al., 2020).

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

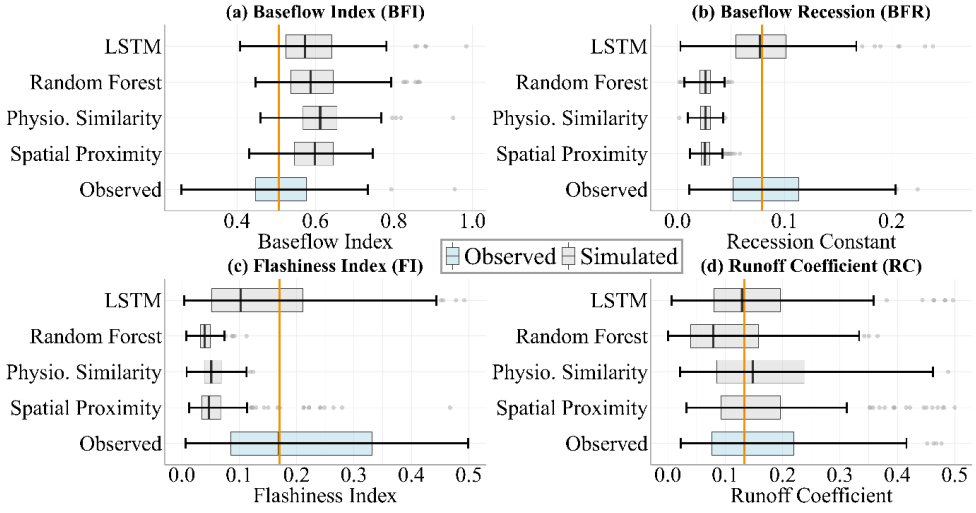


Figure 4: Distribution of observed (light blue) and simulated (light gray) values of four hydrological signatures: (a) Baseflow Index (BFI), (b) Baseflow Recession Constant (BFR), (c) Flashiness Index (FI), and (d) Runoff Coefficient (RC), obtained from regional simulations using the GR4J-based regionalization methods (spatial proximity, physiographic similarity, random forest) and the regional LSTM model. The vertical orange line indicates the median of the observed data for each signature.

4.3 Relationship between catchment attributes and prediction errors

To explore potential patterns in the errors associated with the estimation of hydrological signatures, we applied a K-means clustering analysis based on the results of the LSTM model, which produces the lowest overall errors among the tested models. This unsupervised classification approach aims to determine whether specific physiographic characteristics are systematically linked to different levels of model error. Figure 5 presents the distribution of catchment area (Area) and aridity index (AI) across the three identified error-based clusters (Low, Moderate, and High). This figure is complemented by Figure S4, which displays the corresponding distributions for groundwater depth (GD), available water capacity (AWC), and maximum altitude (MA). Overall, catchments exhibiting higher prediction errors tend to have smaller catchment areas and shallower water tables. This could be explained by the daily time step, which is likely too large to accurately represent surface water dynamics in small

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

catchments. In contrast, catchments with low to moderate prediction errors show larger catchment areas, higher elevations, and higher aridity index values (i.e., wetter conditions), indicating that model performance improves under more humid conditions. This pattern suggests that the model captures hydrological dynamics more accurately in large, spatially integrated catchments, where flow responses are smoother and less influenced by localized variability. Consistent with Salinas et al. (2013), humid catchments often display more linear hydrological behavior, which enhances predictability. In contrast, smaller and drier catchments are more prone to prediction errors, likely reflecting the difficulty of accurately representing evapotranspiration and soil moisture processes in these hydrosystems (Atkinson et al., 2002; Massmann, 2020).

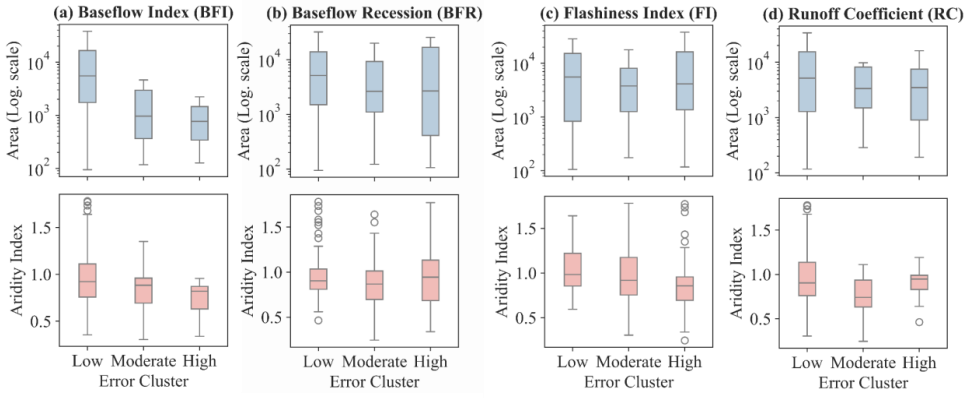


Figure 5: Distribution of catchment area and aridity index across model error clusters (Low, Moderate, High) derived from the relative bias of the LSTM model. Clustering was performed using k-means on four hydrological signatures, including (a) Baseflow Index (BFI), (b) Baseflow Recession Constant (BFR), (c) Flashiness Index (FI), and (d) Runoff Coefficient (RC).

4.4 Evaluation of the accuracy of regionally estimated flood quantiles against observations

Figure 6 presents a comparison between observed flood quantiles for 5-year (Q5) and 20-year (Q20) return periods and those estimated regionally using the GEV distribution. Figure S6 shows the mean relative bias for the regional estimation of 5- and 20-year flood quantiles across the four regionalization methods. For each regionalization method, the Spearman's rank

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

correlation coefficient (ρ) is computed from the paired observed and simulated flood quantiles. The statistical significance of ρ coefficients was determined using a two-tailed test. All reported correlations were found to be statistically significant at the 95% confidence level ($p < 0.05$). For the GR4J model the physiographic similarity method (Figure 6b) shows the lowest performance for frequent floods (Q5), with a ρ value of 0.85, and a systematic positive bias. The RF algorithm shows a scale-dependent bias (Figure S6), underestimating frequent floods (rBias = -0.34 for Q5) and overestimating rare events (rBias = 0.35 for Q20). While the spatial proximity approach (Figure 6a) provides an improvement in Q5 estimation accuracy ($\rho = 0.91$), overestimation and high variability remain (Figure 6a). For Q20, although ρ values remain similar, very high biases with rBias ranging from 0.35 to 0.79 (Figure S6) indicate a sharp decline in predictive performance of these three parameter-transfer methods, highlighting the challenges of applying these methods for rare events. In contrast,the regional LSTM model (Figure 6d) provides a closer alignment with observations. The Spearman's ρ between observed and estimated flood quantiles reaches 0.94 for Q5 and 0.93 for Q20, both significant at the 0.05 level (Figure 6d). Moreover, the LSTM model achieves the lowest estimation errors, with rBias values of -0.32 and -0.33 for the Q5 and Q20, respectively. Despite slight underestimation (i.e., the regression line deviates slightly from the 1:1 line), these results demonstrate that the implemented regional LSTM framework provides more robust statistical predictions of both frequent and rare flood quantiles across the West African network compared to the tested parameter-transfer approaches.

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

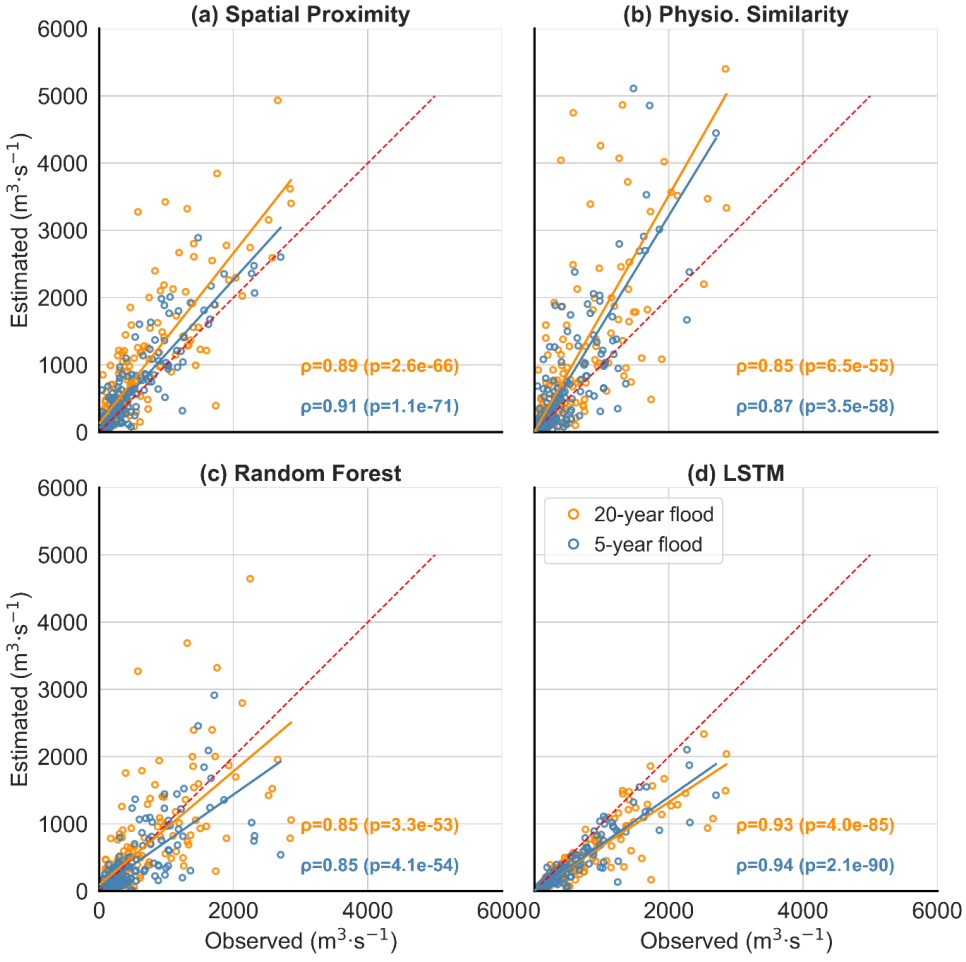


Figure 6: Regional estimation of 5- and 20-year flood quantiles based on streamflows simulated through regionalization of the GR4J hydrological model using three approaches: (a) spatial proximity, (b) physiographic similarity, and (c) random forest (RF), as well as (d) the regional LSTM model. The red dashed line represents the 1:1 identity line, while solid colored lines represent linear regressions for each return period. Spearman's rank correlation coefficients () and their corresponding p-values (in brackets) are indicated within each panel for each method and return period.

5 Conclusions

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

Reliable estimation of streamflow remains a major challenge in ungauged catchments, notably in West Africa, where a large proportion of river basins remain ungauged or poorly gauged. Sparse observational networks, limited hydrometric data, and pronounced hydroclimatic variability hamper the modelling of surface water resources in the region. To address this gap, this study provides a comprehensive evaluation of hydrological regionalization approaches for streamflow and flood quantile estimation at ungauged catchments in West Africa. Historically, regional estimation procedures have been benchmarked within data-abundant frameworks, leaving a substantial knowledge gap regarding their robustness under severe data and monitoring constraints. The present study addresses this gap by demonstrating how alternative modeling paradigms, ranging from explicit conceptual parameter transfer to simultaneous deep learning, perform under the hydroclimatic and data-sparse conditions of West Africa. We compared classical parameter-transfer methods, including spatial proximity, physiographic similarity, and RF-based regionalization of the GR4J model, with a data-driven LSTM framework. Using a jackknife sampling strategy applied to 189 near-natural catchments, we found that the tested traditional parameter-transfer approaches consistently underperform, exhibiting low KGE values and failing to reproduce key hydrological signatures in West Africa. These results are consistent with previous studies in other regions, where classical regionalization methods have also shown limited predictive skill in data-scarce or highly heterogeneous catchments (Arsenault et al., 2019, 2023; Ghaneei et al., 2024; Lees et al., 2021). This limitation reflects the operational challenges of transferring calibrated parameters across highly heterogeneous environments (Chouaib et al., 2018). In contrast, the implemented regional model demonstrated better generalization performance, achieving a median KGE of 0.67 and robustly simulating streamflow, hydrological signatures, and flood quantiles across a broad climatic and physiographic gradient. These results highlight the ability of deep learning approaches, such as LSTM, to capture regional complex, non-linear hydrological dynamics even in data-scarce regions (Kratzert et al., 2019b). It is worth noting that the respective effects of model structure and regionalization strategy cannot fully disentangle the current experimental design. Because the LSTM combines an inherently flexible data-driven architecture with a simultaneous regional learning process, its superior performance cannot be uniquely attributed to either factor alone. To isolate these mechanisms, future research should incorporate alternative controlled configurations. These include evaluating regional learning architectures against localized versions of the same deep-learning model class, testing hybrid conceptual-deep-learning frameworks, or systematically varying the regional pooling density. In addition, while LSTM-based regionalization offers substantial promise for operational

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

hydrological modeling and water resource management in West Africa, challenges remain. High-quality, extensive datasets are essential for model training, and enhancing the interpretability of the LSTM predictions remains a key priority (De la Fuente et al., 2024; Kratzert et al., 2019a). Furthermore, recent studies on alternative architectures such as attention-based mechanisms used in Temporal Fusion Transformers (TFT) and doubly residual mechanisms used in N-BEATS have shown promise in general time-series forecasting and streamflow prediction (Koya and Roy, 2024; Makridakis et al., 2022; Park et al., 2025). While the LSTM remains the current state-of-the-art benchmark in large-sample hydrology, future research could explore whether these alternative architectures provide additional performance gains in the specific context of West African hydroclimatic variability. Overall, this work demonstrates the transformative potential of data-driven approaches to advance hydrological modelling in the climatically complex and data-limited West African context. By bridging the gap between local observations and regional-scale modelling, the LSTM framework can significantly improve flood risk assessment, water resource planning, and climate adaptation strategies in West Africa.

Code availability

All modelling and statistical analyses rely on publicly available open-source software: the R-airGR library (https://hydrogr.github.io/airGR/), the Python scikit-learn library (https://scikit-learn.org/stable/), the NeuralHydrology Python framework (neuralhydrology/neuralhydrology), and the R-extRemes package for extreme value analysis (Gilleland & Katz, 2016).

Data availability

The daily observed streamflow records used in this study are not experimental or privately collected data produced by the authors. The hydrometric stations considered in this study are located in West African river basins whose monitoring networks have historically been operated by national hydrological services and, for several transboundary river systems, coordinated in conjunction with regional basin organizations. The original discharge observations therefore remain under the custodianship of the national institutions responsible for hydrometric monitoring and are subject to their respective data-access and redistribution

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

policies. Consequently, the authors are not authorized to redistribute all original daily streamflow time series in a public repository. To facilitate access by the scientific community, Table S4 (Supplementary Material) provides the national institutions responsible for water-resource and hydrometric monitoring in West African countries, together with publicly available institutional contact information. Relevant transboundary basin organizations are also identified where applicable. Researchers interested in obtaining the underlying daily observations are encouraged to contact the appropriate national hydrological service or basin organization directly.

Author contributions

SBD, FKate, IS, and YT conceived and designed the study. YT provided the ADHI dataset. IS, JG, RH, VS, and KT implemented the regional LSTM simulations, while FKate, SBD, and LO implemented the GR4J regionalization framework. The remaining co-authors contributed to the analysis and interpretation of the results. YT supervised the study. All authors contributed to the writing and revision of the manuscript and approved the final version.

Competing interest

The authors declare that they have no conflict of interest.

Financial support

This research has been funded by the HYDRAFIA project funded by the Carnot Institute for Water & Environment and the French National Research Agency (ANR). Anne Johannet, Kalil Traore and Yves Tramblay are funded by the French ANR (ANR-21-LCV1-0001) support on the Laboratory of Hydrological forecasting by Artificial Intelligence (Hydr.IA) project. Serigne Bassirou Diop, Fadilath Kate, Ansoumana Bodian, and Yves Tramblay are funded by the AFD/IRD project CECC.

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

References

Arsenault, R., Breton-Dufour, M., Poulin, A., Dallaire, G., and Romero-Lopez, R.: Streamflow prediction in ungauged basins: analysis of regionalization methods in a hydrologically heterogeneous region of Mexico, Hydrological Sciences Journal, 64, 1297–1311, https://doi.org/10.1080/02626667.2019.1639716, 2019.

Arsenault, R., Martel, J.-L., Brunet, F., Brissette, F., and Mai, J.: Continuous streamflow prediction in ungauged basins: long short-term memory neural networks clearly outperform traditional hydrological models, Hydrology and Earth System Sciences, 27, 139–157, https://doi.org/10.5194/hess-27-139-2023, 2023.

Atkinson, S. E., Woods, R. A., and Sivapalan, M.: Climate and landscape controls on water balance model complexity over changing timescales, Water Resources Research, 38, 50-1-50–17, https://doi.org/10.1029/2002WR001487, 2002.

Baker, D. B., Richards, R. P., Loftus, T. T., and Kramer, J. W.: A New Flashiness Index: Characteristics and Applications to Midwestern Rivers and Streams, JAWRA Journal of the American Water Resources Association, 40, 503–522, https://doi.org/10.1111/j.1752-1688.2004.tb01046.x, 2004.

Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., McDonnell, J. J., Savenije, H. H. G., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Széles, B., Viglione, A., Aksoy, H., Allen, S. T., Amin, A., Andréassian, V., Arheimer, B., Aryal, S. K., Baker, V., Bardsley, E., Barendrecht, M. H., Bartosova, A., Batelaan, O., Berghuijs, W. R., Beven, K., Blume, T., Bogaard, T., Amorim, P. B. de, Böttcher, M. E., Boulet, G., Breinl, K., Brilly, M., Brocca, L., Buytaert, W., Castellarin, A., Castelletti, A., Chen, X., Chen, Y., Chen, Y., Chifflard, P., Claps, P., Clark, M. P., Collins, A. L., Croke, B., Dathe, A., David, P. C., Barros, F. P. J. de, Rooij, G. de, Baldassarre, G. D., Driscoll, J. M., Duethmann, D., Dwivedi, R., Eris, E., Farmer, W. H., Feiccabrino, J., Ferguson, G., Ferrari, E., Ferraris, S., Fersch, B., Finger, D., Foglia, L., Fowler, K., Gartsman, B., Gascoin, S., Gaume, E., Gelfan, A., Geris, J., Gharari, S., Gleeson, T., Glendell, M., Bevacqua, A. G., González-Dugo, M. P., Grimaldi, S., Gupta, A. B., Guse, B., Han, D., Hannah, D., Harpold, A., Haun, S., Heal, K., Helfricht, K., Herrnegger, M., Hipsey, M., Hlaváčiková, H., Hohmann, C., Holko, L., Hopkinson, C., Hrachowitz, M., Illangasekare, T. H., Inam, A., Innocente, C., Istanbulluoglu, E., Jarihani, B., Kalantari, Z., Kalvans, A., Khanal, S., Khatami, S., Kiesel, J., Kirkby, M., Knoben, W., Kochanek, K., Kohnová, S., Kolechkina, A., Krause, S., Kreamer, D., Kreibich, H., Kunstmann, H., Lange, H., Liberato, M. L. R., Lindquist, E., Link, T., Liu, J., Loucks, D. P., Luce, C., Mahé, G., Makarieva, O., Malard, J., Mashtayeva, S., Maskey, S., Mas-Pla, J., Mavrova-Guirguinova, M., Mazzoleni, M., Mernild, S., Misstear, B. D., Montanari, A., Müller-Thomy, H., Nabizadeh, A., Nardi, F., Neale, C., Nesterova, N., Nurtaev, B., Odongo, V. O., Panda, S., Pande, S., Pang, Z., Papacharalampous, G., Perrin, C., Pfister, L., Pimentel, R., Polo, M. J., Post, D., Sierra, C. P., Ramos, M.-H., Renner, M., Reynolds, J. E., Ridolfi, E., Rigon, R., Riva, M., Robertson, D. E., Rosso, R., Roy, T., Sá, J. H. M., Salvadori, G., Sandells, M., Schaefli, B., Schumann, A., Scolobig, A., Seibert, J., Servat, E., Shafiei, M., Sharma, A., Sidibe, M., Sidle, R. C., Skaugen, T., Smith, H., Spiessl, S. M., Stein, L., Steinsland, I., Strasser, U., Su, B., Szolgay, J., Tarboton, D., Tauro, F., Thirel, G., Tian, F., Tong, R., Tussupova, K., Tyralis, H., Uijlenhoet, R., Beek, R. van, Ent, R. J. van der, Ploeg, M. van der, Loon, A. F. V., Meerveld, I. van, Nooijen, R. van, Oel, P. R. van, Vidal, J.-P., Freyberg, J. von, Vorogushyn, S., Wachniew, P., Wade, A. J., Ward, P., Westerberg, I. K., White, C., Wood, E. F., Woods, R., Xu, Z., Yilmaz, K. K., and Zhang, Y.: Twenty-three unsolved problems in hydrology (UPH) – a community perspective, Hydrological Sciences Journal, 2019.

Bodian, A., Dezetter, A., Deme, A., and Diop, L.: Hydrological Evaluation of TRMM Rainfall over the Upper Senegal River Basin, Hydrology, 3, 15, https://doi.org/10.3390/hydrology3020015, 2016.

Bonzanigo, L., Fraval, P., Lajaunie, M.-L., and Scheierling, S. M.: World bank engagement in transboundary waters in west africa: Retrospective and lessons learned, World Bank Group, Washington, D.C., http://documents.worldbank.org/curated/en/652141636371253908 (last access: 17 October 2025), 2021.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

Bruneau, P., Gascuel-Odoux, C., Robin, P., Merot, Ph., and Beven, K.: Sensitivity to space and time resolution of a hydrological model using digital elevation data, Hydrological Processes, 9, 69–81, https://doi.org/10.1002/hyp.3360090107, 1995.

Cantoni, E., Tramblay, Y., Grimaldi, S., Salamon, P., Dakhlaoui, H., Dezetter, A., and Thiemig, V.: Hydrological performance of the ERA5 reanalysis for flood modeling in Tunisia with the LISFLOOD and GR4J models, Journal of Hydrology: Regional Studies, 42, 101169, https://doi.org/10.1016/j.ejrh.2022.101169, 2022.

Choi, J., Lee, J., and Kim, S.: Utilization of the Long Short-Term Memory network for predicting streamflow in ungauged basins in Korea, Ecological Engineering, 182, 106699, https://doi.org/10.1016/j.ecoleng.2022.106699, 2022.

Chouaib, W., Alila, Y., and Caldwell, P. V.: Parameter transferability within homogeneous regions and comparisons with predictions from a priori parameters in the eastern United States, Journal of Hydrology, 560, 24–38, https://doi.org/10.1016/j.jhydrol.2018.03.018, 2018.

Coles, S.: An introduction to statistical modeling of extreme values, 1st ed., Springer, London Berlin Heidelberg, 208 pp., https://doi.org/10.1007/978-1-4471-3675-0, 2001.

De la Fuente, L. A., Ehsani, M. R., Gupta, H. V., and Condon, L. E.: Toward interpretable LSTM-based modeling of hydrological systems, Hydrology and Earth System Sciences, 28, 945–971, https://doi.org/10.5194/hess-28-945-2024, 2024.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Descroix, L., Guichard, F., Grippa, M., Lambert, L. A., Panthou, G., Mahé, G., Gal, L., Dardel, C., Quantin, G., Kergoat, L., Bouaïta, Y., Hiernaux, P., Vischel, T., Pellarin, T., Faty, B., Wilcox, C., Malam Abdou, M., Mamadou, I., Vandervaere, J.-P., Diongue-Niang, A., Ndiaye, O., Sané, Y., Dacosta, H., Gosset, M., Cassé, C., Sultan, B., Barry, A., Amogu, O., Nka Nnomo, B., Barry, A., and Paturel, J.-E.: Evolution of Surface Hydrology in the Sahelo-Sudanian Strip: An Updated Review, Water, 10, 748, https://doi.org/10.3390/w10060748, 2018.

Devia, G. K., Ganasri, B. P., and Dwarakish, G. S.: A Review on Hydrological Models, Aquatic Procedia, 4, 1001–1007, https://doi.org/10.1016/j.aqpro.2015.02.126, 2015.

Diallo, A., Donkor, E., and Owusu, V.: Climate change adaptation strategies, productivity and sustainable food security in southern Mali, Climatic Change, 159, 309–327, https://doi.org/10.1007/s10584-020-02684-8, 2020.

Ding, Y., Zhu, Y., Feng, J., Zhang, P., and Cheng, Z.: Interpretable spatio-temporal attention LSTM model for flood forecasting, Neurocomputing, 403, 348–359, https://doi.org/10.1016/j.neucom.2020.04.110, 2020.

Diop, S. B., Tramblay, Y., Bodian, A., Ekolu, J., Rouché, N., and Dieppois, B.: Flood Frequency Analysis in West Africa, J Flood Risk Management, 18, e70001, https://doi.org/10.1111/jfr3.70001, 2025.

Fréchet, M.: Sur la loi de probabilité de l'écart maximum. Annales de la Société Polonaise de Mathématique, 6, 1927.

Ghaneei, P., Foroumandi, E., and Moradkhani, H.: Enhancing Streamflow Prediction in Ungauged Basins Using a Nonlinear Knowledge-Based Framework and Deep Learning, Water Resources Research, 60, e2024WR037152, https://doi.org/10.1029/2024WR037152, 2024.

Gnann, S. J., Coxon, G., Woods, R. A., Howden, N. J. K., and McMillan, H. K.: TOSSH: A Toolbox for Streamflow Signatures in Hydrology, Environmental Modelling & Software, 138, 104983, https://doi.org/10.1016/j.envsoft.2021.104983, 2021.

Golian, S., Murphy, C., and Meresa, H.: Regionalization of hydrological models for flow estimation in ungauged catchments in Ireland, Journal of Hydrology: Regional Studies, 36, 100859, https://doi.org/10.1016/j.ejrh.2021.100859, 2021.

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

Gumbel, E. J.: Statistics of Extremes, Columbia University Press, https://doi.org/10.7312/gumb92958, 1958.

Gupta, H. V., Wagener, T., and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrological Processes, 22, 3802–3813, https://doi.org/10.1002/hyp.6989, 2008.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, Neural Computation, 9, 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735, 1997.

Horner, I.: Design and evaluation of hydrological signatures for the diagnostic and improvement of a process-based distributed hydrological model, phd, Université Grenoble Alpes, Grenoble, France, 2020.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E., and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB)—a review, Hydrological Sciences Journal, 58, 1198–1255, https://doi.org/10.1080/02626667.2013.803183, 2013.

Ibrahim, B., Wisser, D., Barry, B., Fowe, T., and Aduna, A.: Hydrological predictions for small ungauged watersheds in the Sudanian zone of the Volta basin in West Africa, Journal of Hydrology: Regional Studies, 4, 386–397, https://doi.org/10.1016/j.ejrh.2015.07.007, 2015.

IPCC: Sixth Assessment Report (AR6): Climate Change 2021: The Physical Science Basis., 2021.

Kim, J., Shin, S., Park, D., and Jung, K.: Ensemble artificial neural network and generalized additive model for data-scarce regional frequency analysis in design flood estimation, Journal of Hydrology, 667, 134886, https://doi.org/10.1016/j.jhydrol.2025.134886, 2026.

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

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

Koubodana, H. D., Atchonouglo, K., Adounkpe, J. G., Amoussou, E., Kodja, D. J., Koungbanane, D., Afoudji, K. Y., Lombo, Y., and Kpemoua, K. E.: Surface runoff prediction and comparison using IHACRES and GR4J lumped models in the Mono catchment, West Africa, in: Proceedings of IAHS, Hydrology of Large River Basins of Africa - 4th International Conference on the "'Hydrology of the Great Rivers of Africa,'" 63–68, https://doi.org/10.5194/piahs-384-63-2021, 2021.

Koya, S. R. and Roy, T.: Temporal Fusion Transformers for streamflow Prediction: Value of combining attention with recurrence, Journal of Hydrology, 637, 131301, https://doi.org/10.1016/j.jhydrol.2024.131301, 2024.

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

Kratzert, F., Herrnegger, M., Klotz, D., Hochreiter, S., and Klambauer, G.: NeuralHydrology – Interpreting LSTMs in Hydrology, in: Explainable AI: Interpreting, Explaining and Visualizing

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

Deep Learning, edited by: Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., and Müller, K.-R., Springer International Publishing, Cham, 347–362, https://doi.org/10.1007/978-3-030-28954-6_19, 2019a.

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., and Nearing, G. S.: Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning, Water Resources Research, 55, 11344–11354, https://doi.org/10.1029/2019WR026065, 2019b.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., and Nearing, G.: Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets, Hydrology and Earth System Sciences, 23, 5089–5110, https://doi.org/10.5194/hess-23-5089-2019, 2019c.

Kratzert, F., Gauch, M., Nearing, G., and Klotz, D.: NeuralHydrology --- A Python library for Deep Learning research in hydrology, Journal of Open Source Software, 7, 4050, https://doi.org/10.21105/joss.04050, 2022.

Krishnamurthy, P. K., Lewis, K., and Choularton, R. K.: Climate impacts on food security and nutrition—a review of existing knowledge., 2012.

Kuana, L. A., Almeida, A. S., Mercuri, E. G. F., and Noe, S. M.: Regionalization of GR4J model parameters for river flow prediction in Paraná, Brazil, Hydrology and Earth System Sciences, 28, 3367–3390, https://doi.org/10.5194/hess-28-3367-2024, 2024.

Kunkle, G. R.: The baseflow-duration curve, A technique for the study of groundwater discharge from a drainage basin, Journal of Geophysical Research (1896-1977), 67, 1543–1554, https://doi.org/10.1029/JZ067i004p01543, 1962.

Kupzig, J., Kupzig, N., and Flörke, M.: Regionalization in global hydrological models and its impact on runoff simulations: a case study using WaterGAP3 (v 1.0.0), Geoscientific Model Development, 17, 6819–6846, https://doi.org/10.5194/gmd-17-6819-2024, 2024.

Kwakye, S. O. and Bárdossy, A.: Hydrological modelling in data-scarce catchments: Black Volta basin in West Africa, SN Appl. Sci., 2, 628, https://doi.org/10.1007/s42452-020-2454-4, 2020.

Ladson, A. R., Brown, R., Neal, B., and Nathan, R.: A Standard Approach to Baseflow Separation Using The Lyne and Hollick Filter, Australasian Journal of Water Resources, 17, 25–34, https://doi.org/10.7158/13241583.2013.11465417, 2013.

van der Land, V., Romankiewicz, C., and Geest, K. van der: Environmental change and migration: A review of West African case studies, in: Routledge Handbook of Environmental Displacement and Migration, Routledge, London, UK, 464, 2018.

Le Barbé, L. and Lebel, T.: Rainfall climatology of the HAPEX-Sahel region during the years 1950–1990, Journal of Hydrology, 188–189, 43–73, https://doi.org/10.1016/S0022-1694(96)03154-X, 1997.

Le Barbé, L., Lebel, T., and Tapsoba, D.: Rainfall Variability in West Africa during the Years 1950–90, Journal of Climate, 15, 187–202, https://doi.org/10.1175/1520-0442(2002)015%253C0187:RVIWAD%253E2.0.CO;2, 2002.

Le Moine, N., Andréassian, V., Perrin, C., and Michel, C.: How can rainfall-runoff models handle intercatchment groundwater flows? Theoretical study based on 1040 French catchments, Water Resources Research, 43, https://doi.org/10.1029/2006WR005608, 2007.

Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G., and Dadson, S. J.: Benchmarking data-driven rainfall–runoff models in Great Britain: a comparison of long short-term memory (LSTM)-based models with four lumped conceptual models, Hydrology and Earth System Sciences, 25, 5517–5534, https://doi.org/10.5194/hess-25-5517-2021, 2021.

Lees, T., Reece, S., Kratzert, F., Klotz, D., Gauch, M., De Bruijn, J., Kumar Sahu, R., Greve, P., Slater, L., and Dadson, S. J.: Hydrological concept formation inside long short-term memory (LSTM) networks, Hydrology and Earth System Sciences, 26, 3079–3101, https://doi.org/10.5194/hess-26-3079-2022, 2022.

Luo, Y. and Su, S.: SpatioTemporal Random Forest and SpatioTemporal Stacking Tree: A novel spatially explicit ensemble learning approach to modeling non-linearity in spatiotemporal non-stationarity, International Journal of Applied Earth Observation and Geoinformation, 136, 104315, https://doi.org/10.1016/j.jag.2024.104315, 2025.

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

Mahé, G. and Paturel, J.-E.: 1896–2006 Sahelian annual rainfall variability and runoff increase of Sahelian Rivers, Comptes Rendus Geoscience, 341, 538–546, https://doi.org/10.1016/j.crte.2009.05.002, 2009.

Makridakis, S., Spiliotis, E., Assimakopoulos, V., Chen, Z., Gaba, A., Tsetlin, I., and Winkler, R. L.: The M5 uncertainty competition: Results, findings and conclusions, International Journal of Forecasting, 38, 1365–1385, https://doi.org/10.1016/j.ijforecast.2021.10.009, 2022.

Martel, B., Ouarda, T. B. M. J., Barbet, M., Bruneau, P., Latraverse, M., and Kamali Nezhad, M.: Regional frequency analysis of autumnal floods in the province of Quebec, Canada, Nat Hazards, 59, 681–698, https://doi.org/10.1007/s11069-011-9789-5, 2011.

Martel, J.-L., Brissette, F., Arsenault, R., Turcotte, R., Castañeda-Gonzalez, M., Armstrong, W., Mailhot, E., Pelletier-Dumont, J., Rondeau-Genesse, G., and Caron, L.-P.: Assessing the adequacy of traditional hydrological models for climate change impact studies: a case for long short-term memory (LSTM) neural networks, Hydrology and Earth System Sciences, 29, 2811–2836, https://doi.org/10.5194/hess-29-2811-2025, 2025.

Massmann, C.: Identification of factors influencing hydrologic model performance using a top-down approach in a large number of U.S. catchments, Hydrological Processes, 34, 4–20, https://doi.org/10.1002/hyp.13566, 2020.

McMillan, H. K.: A review of hydrologic signatures and their applications, WIREs Water, 8, e1499, https://doi.org/10.1002/wat2.1499, 2021.

Michel, C.: Hydrologie appliquée aux petits bassins ruraux: Hydrology handbook (in French), Cemagref, Antony, France, Antony, France, 1991.

Mihret, T. T., Zemale, F. A., Worqlul, A. W., Ayalew, A. D., and Fohrer, N.: Unlocking watershed mysteries: Innovative regionalization of hydrological model parameters in data-scarce regions, Journal of Hydrology: Regional Studies, 57, 102163, https://doi.org/10.1016/j.ejrh.2024.102163, 2025.

Mulvaney, T. J.: On the use of self-registering rain and flood gauges in making observations of the relations of rainfall and flood discharges in a catchment, Transactions of the Institution of Civil Engineers of Ireland, 4, 18–31, 1851.

Naik, G. R. (Ed.): Advances in principal component analysis: Research and development, 1st ed., Springer Singapore, Singapore, https://doi.org/10.1007/978-981-10-6704-4, 2018.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Ndehedehe, C. E.: The water resources of tropical West Africa: problems, progress, and prospects, Acta Geophys., 67, 621–649, https://doi.org/10.1007/s11600-019-00260-y, 2019.

Ndehedehe, C. E., Awange, J. L., Kuhn, M., Agutu, N. O., and Fukuda, Y.: Climate teleconnections influence on West Africa's terrestrial water storage, Hydrological Processes, 31, 3206–3224, https://doi.org/10.1002/hyp.11237, 2017.

Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., Hassidim, A., Klotz, D., Kratzert, F., Metzger, A., Nevo, S., Pappenberger, F., Prudhomme, C., Shalev, G., Shenzis, S., Tekalign, T. Y., Weitzner, D., and Matias, Y.: Global prediction of extreme floods in ungauged watersheds, Nature, 627, 559–563, https://doi.org/10.1038/s41586-024-07145-1, 2024.

Nicholson, S. E.: Climate of the Sahel and West Africa, in: Oxford Research Encyclopedia of Climate Science, Oxford University Press, https://doi.org/10.1093/acrefore/9780190228620.013.510, 2018.

Niel, H., Paturel, J.-E., and Servat, E.: Study of parameter stability of a lumped hydrologic model in a context of climatic variability, Journal of Hydrology, 278, 213–230, https://doi.org/10.1016/S0022-1694(03)00158-6, 2003.

OECD/SWAC: West African Studies: Regional Atlas on West Africa, OECD/Publishing, Paris, 2009.

Orange, D.: Hydroclimatologie du Fouta Djalon et dynamique actuelle d'un vieux paysage latéritique (Afrique de l'Ouest), Université Louis Paster, CNRS, ORSTOM, 1990.

Ouarda, T. B. M. J., Girard, C., Cavadias, G. S., and Bobée, B.: Regional flood frequency estimation with canonical correlation analysis, Journal of Hydrology, 254, 157–173, https://doi.org/10.1016/S0022-1694(01)00488-7, 2001.

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.

Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments, Water Resources Research, 44, https://doi.org/10.1029/2007WR006240, 2008.

Oudin, L., Caupin, R., and Brigode, P.: Improved predictions in ungauged basins using flood marks, Journal of Hydrology, 657, 133145, https://doi.org/10.1016/j.jhydrol.2025.133145, 2025.

Padgham, J., Abubakari, A., Ayivor, J., Dietrich, K., Fosu-Mensah, B., Gordon, C., Habtezion, S., Lawson, E., Mensah, A., Nukpezah, D., Ofori, B., Piltz, S., Sidibé, A., Sissoko, M., Totin, E., and Traoré, S.: Vulnerability and adaptation to climate change in the semi-arid regions of West Africa, 2015.

Pagliero, L., Bouraoui, F., Diels, J., Willems, P., and McIntyre, N.: Investigating regionalization techniques for large-scale hydrological modelling, Journal of Hydrology, 570, 220–235, https://doi.org/10.1016/j.jhydrol.2018.12.071, 2019.

Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters, Hydrol. Earth Syst. Sci., 9, 157–171, https://doi.org/10.5194/hess-9-157-2005, 2005.

Park, J., Joh, J. S., Choi, M., Kim, T., Cho, J., and Lee, Y.: AI-Based Time-Series Ensemble Approach Coupled with a Hydrological Model for Reservoir Storage Prediction in Korea, Water, 17, 3296, https://doi.org/10.3390/w17223296, 2025.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, Journal of Hydrology, 279, 275–289, https://doi.org/10.1016/S0022-1694(03)00225-7, 2003.

Pokhrel, P., Gupta, H. V., and Wagener, T.: A spatial regularization approach to parameter estimation for a distributed watershed model, Water Resources Research, 44, 2007WR006615, https://doi.org/10.1029/2007WR006615, 2008.

Saadi, M., Oudin, L., and Ribstein, P.: Random Forest Ability in Regionalizing Hourly Hydrological Model Parameters, Water, 11, 1540, https://doi.org/10.3390/w11081540, 2019.

Saha, A., Basu, S., and Datta, A.: Random Forests for Spatially Dependent Data, Journal of the American Statistical Association, 118, 665–683, https://doi.org/10.1080/01621459.2021.1950003, 2023.

Salinas, J. L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies, Hydrology and Earth System Sciences, 17, 2637–2652, https://doi.org/10.5194/hess-17-2637-2013, 2013.

Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resources Research, 46, https://doi.org/10.1029/2008WR007327, 2010.

Sarhadi, A. and Modarres, R.: Flood seasonality-based regionalization methods: a data-based comparison, Hydrological Processes, 25, 3613–3624, https://doi.org/10.1002/hyp.8088, 2011.

Schonlau, M. and Zou, R. Y.: The random forest algorithm for statistical learning, The Stata Journal, 20, 3–29, https://doi.org/10.1177/1536867X20909688, 2020.

Schuol, J., Abbaspour, K. C., Srinivasan, R., and Yang, H.: Estimation of freshwater availability in the West African sub-continent using the SWAT hydrologic model, Journal of Hydrology, 352, 30–49, https://doi.org/10.1016/j.jhydrol.2007.12.025, 2008.

Shen, H., Tolson, B. A., and Mai, J.: Time to Update the Split-Sample Approach in Hydrological Model Calibration, Water Resources Research, 58, e2021WR031523, https://doi.org/10.1029/2021WR031523, 2022.

Shu, C. and Ouarda, T. B. M. J.: Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system, Journal of Hydrology, 349, 31–43, https://doi.org/10.1016/j.jhydrol.2007.10.050, 2008.

Thorncroft, C. D., Nguyen, H., Zhang, C., and Peyrillé, P.: Annual cycle of the West African monsoon: regional circulations and associated water vapour transport, Quarterly Journal of the Royal Meteorological Society, 137, 129–147, https://doi.org/10.1002/qj.728, 2011.

Tramblay, Y., Rouché, N., Paturel, J.-E., Mahé, G., Boyer, J.-F., Amoussou, E., Bodian, A., Dacosta, H., Dakhlaoui, H., Dezetter, A., Hughes, D., Hanich, L., Peugeot, C., Tshimanga, R., and

https://doi.org/10.5194/egusphere-2026-7029
Preprint. Discussion started: 20 November 2026
© Author(s) 2026. CC BY 4.0 License.



Lachassagne, P.: ADHI: the African Database of Hydrometric Indices (1950–2018), Earth Syst. Sci. Data, 13, 1547–1560, https://doi.org/10.5194/essd-13-1547-2021, 2021.

Tramblay, Y., El Khalki, E. M., Khedimallah, A., Sadaoui, M., Benaabidate, L., Boulmaiz, T., Boutaghane, H., Dakhlaoui, H., Hanich, L., Ludwig, W., Meddi, M., Elmehdi Saidi, M., and Mahé, G.: Regional flood frequency analysis in North Africa, Journal of Hydrology, 630, 130678, https://doi.org/10.1016/j.jhydrol.2024.130678, 2024.

Wagener, T., Wheater, H. S., and Gupta, H. V.: Rainfall-Runoff Modelling in Gauged and Ungauged Catchments, Imperial College Press, London, UK, https://doi.org/10.1142/p335, 2004.

Wang, Y., Hu, Y., Zhang, C., and Wang, X.: Advancing Deep Learning for Regionalized Streamflow Modeling: A Comparative Study of Basin Similarity Approaches, Water, 18, 1693, https://doi.org/10.3390/w18141693, 2026.

Wantzen, K. M. (Ed.): River culture: life as a dance to the rhythm of the waters, UNESCO, Beijing, China, 894 pp., https://doi.org/10.54677/HHMI3947, 2023.

Yang, X., Magnusson, J., and Xu, C.-Y.: Transferability of regionalization methods under changing climate, Journal of Hydrology, 568, 67–81, https://doi.org/10.1016/j.jhydrol.2018.10.030, 2019.

Yang, X., Magnusson, J., Huang, S., Beldring, S., and Xu, C.-Y.: Dependence of regionalization methods on the complexity of hydrological models in multiple climatic regions, Journal of Hydrology, 582, 124357, https://doi.org/10.1016/j.jhydrol.2019.124357, 2020.

Ye, K., Liang, Z., Chen, H., Qian, M., Hu, Y., Bi, C., Wang, J., and Li, B.: Regionalization Strategy Guided Long Short-Term Memory Model for Improving Flood Forecasting, Hydrological Processes, 38, e15296, https://doi.org/10.1002/hyp.15296, 2024.