Flow duration curve prediction

A framework integrating regionalization and copula model

Lan, Tian; Zhang, Jiajia; Li, Huanhuan; Zhang, Hongbo; Gong, Xinghui; Sun, Jing; Chen, Yongqin David; Xu, Chong Yu

DOI
10.1016/j.jhydrol.2024.132364

Publication date
2025

Document Version
Final published version

Published in
Journal of Hydrology

Citation (APA)
Lan, T., Zhang, J., Li, H., Zhang, H., Gong, X., Sun, J., Chen, Y. D., & Xu, C. Y. (2025). Flow duration curve prediction: A framework integrating regionalization and copula model. *Journal of Hydrology*, *647*, Article 132364. https://doi.org/10.1016/j.jhydrol.2024.132364

Journal of Hydrology 647 (2025) 132364

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Flow duration curve prediction: A framework integrating regionalization and copula model

Tian Lan [a,b,c], Jiajia Zhang [a], Huanhuan Li [a,b,c,*], Hongbo Zhang [a,b,c], Xinghui Gong [a,b,c], Jing Sun [d], Yongqin David Chen [e], Chong-Yu Xu [f]

[a] School of Water and Environment, Chang'an University, Xi'an 710054, China
[b] Key Laboratory of Subsurface Hydrology and Ecological Effects in Arid Region of the Ministry of Education, Chang'an University, Xi'an 710054, China
[c] Key Laboratory of Eco-Hydrology and Water Security in Arid and Semi-Arid Regions of Ministry of Water Resources, Chang'an University, Xi'an 710054, China
[d] Department of Intelligent Systems, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE Delft, the Netherlands
[e] School of Humanities and Social Science, The Chinese University of Hong Kong, Shenzhen 518172, China
[f] Department of Geosciences, University of Oslo, P.O. Box 1047 Blindern, 0316 Oslo, Norway

ARTICLE INFO

Keywords:
Flow duration curve
Ungauged basin
Vine copula structure
Parameter estimation
Hydrological similarity

ABSTRACT

Flow Duration Curve (FDC) is an essential graphical tool for illustrating the variability of observed historical streamflow. Achieving an advanced understanding of the physical characteristics governing FDCs is crucial for enhancing predictions of FDCs in ungauged basins. However, this remains challenging due to the complex processes that control streamflow components and their interactions. To address this, a novel framework that integrates regionalization and process-based methods is proposed for predicting FDCs in ungauged basins. This framework implements a hydrological similarity-based regionalization method to estimate hydrological model parameters in ungauged basins, enhancing streamflow prediction reliability. It categorizes streamflow into four distinct components based on delayed flow separation: Short-delay, intermediate-delay, long-delay, and baseline-delay. These components are synthesized to construct the FDC, with their interdependencies modeled using a Vine copula structure. A Bayesian-based estimation technique for the copula function parameters is developed to further improve the precision of the predicted FDC. Applied to nine selected MOPEX ungauged basins, the framework demonstrated superior accuracy, especially for low to middle streamflow phases. Moreover, the framework exhibited superior simulation accuracy during the validation period, highlighting its substantial potential for future-oriented water resource management and planning strategies.

1. Introduction

Flow Duration Curve (FDC) represents a graphical depiction of the observed historical flow variation, serving as an instrumental method for characterizing hydrological patterns within a basin (Leong and Yokoo, 2021; Owolabi et al., 2020; Verma et al., 2017; Wolff and Duarte, 2021). By ranking observed streamflow and estimating their corresponding exceedance probabilities through an appropriate plotting position, the FDC can be derived from streamflow records (Leong and Yokoo, 2019). The FDC's application spans a wide array of water resource planning and management scenarios, including flood and low-flow situation assessments (Vorogushyn et al., 2018), hydrologic effects of afforestation (Niemeyer et al., 2020), sedimentation in rivers (Sathya and Srinivas, 2023; Vogel and Fennessey, 1995), water quality management (Asadollah et al., 2021), and determination of environmental flow standards for protecting aquatic habitats and ecosystem health (Suwal et al., 2020). Empirical FDC is pivotal in understanding streamflow behavior within basins, however, its construction is determined by the availability of streamflow data. Due to a lack of data, FDCs are not readily available along ungauged basins, still, these are often the regions that have the greatest need for an understanding of hydrological processes. Recognizing this need, the prediction of FDCs in ungauged basins has been, for years, an active area of research (Qamar et al., 2016). Data interpolation approaches employ streamflow from proximate or analogous basins, along with pertinent geographical and hydrological attributes, to predict the FDC at ungauged basins (Castellarin, 2014; Persiano et al., 2022; Waseem et al., 2015). The distribution of streamflow in mountainous and plain regions may be influenced by

* Corresponding author at: School of Water and Environment, Chang'an University, Xi'an 710054, China.
E-mail address: huanhuanli@chd.edu.cn (H. Li).

https://doi.org/10.1016/j.jhydrol.2024.132364
Received 20 May 2024; Received in revised form 29 October 2024; Accepted 3 November 2024
Available online 17 November 2024

factors such as topography, soil type, and vegetation cover, giving rise to heterogeneity. If these differences are not considered by the interpolation methods, the result of predictions may be inaccurate. In addition, the interpolation results may be affected when observed data with uneven or sparse spatial distribution are utilized for interpolation (Atieh et al., 2017). Regional analysis approaches refer to transfer the of hydrological information from one catchment (location) to another, hydrological information may be either the model parameters or the general structure of models that estimate hydrological responses. Regional analysis approaches are satisfactory if the catchments are similar but error-prone if they are not, the edge effects near the boundaries of a basin have the potential for uncertainty (Booker and Snelder, 2012; Singh and Devi, 2022). Deep learning approaches are explored by researchers, such as neural networks (Yu et al., 2023), to address the FDC prediction problem in ungauged basins based on recent advancements in machine learning and artificial intelligence technologies (Bozchaloei and Vafakhah, 2015; Vafakhah and Bozchaloei, 2020). Traditional machine learning techniques, such as boosting, bootstrapped aggregation, and stacked generalization, demand extensive data, posing a significant challenge for ungauged basins where data scarcity may introduce considerable uncertainty (Xu and Liang, 2021). In these regards, these common approaches do not avoid certain limitations. Hence, the principal aim of this research is to develop a novel framework for improving the accuracy and robustness of FDC prediction for ungauged basins.

The premise for predicting the FDC at ungauged basins is to predict streamflow in ungauged basins. To overcome the lack of streamflow in ungauged basins, methods for streamflow prediction at ungauged basins are widely explored (Costa and Fernandes, 2021; Ridolfi et al., 2020). For instance, performance-weighted methods (Farmer and Vogel, 2013), data-driven hydrological methods (Zhang et al., 2023), and regionalization methods. As existing hydrological data can be effectively utilized to enhance predictions in ungauged basins, through adjusting regionalization strategies, and compared to complex hydrological models that require extensive detailed data, this method still provides effective hydrological predictions even with limited data availability (Singh and Devi, 2022), the regionalization method is the most prevalent in ungauged basins prediction (Müller and Thompson, 2016; Zhang et al., 2015). Different types of regionalization methods such as the regression method, spatial proximity method, hydrological similarity method, and output averaging methods have been recommended and widely used all over the world by different authors (Arsenault et al., 2019; Singh and Devi, 2022). The hydrological similarity method includes the simulation of physical hydrological processes and the transfer of parameters from hydrological models that have been developed and calibrated for gauged basins to ungauged basins, also the hydrological processes and physical characteristics of the basins are considered to ensure alignment with the actual hydrological mechanisms within the basins (Archfield et al., 2015). Based on these conditions, the hydrological similarity method is employed to obtain parameters of hydrological models in ungauged basins (Feng et al., 2022; Huang et al., 2020; Refsgaard et al., 2022), and the common techniques of the hydrological models' parameter transfer involve parameter regression and machine learning. Parameter transfer methods aim to estimate the interrelationship between catchment model parameters and characteristic attributes, deriving the parameters for ungauged basins based on calibration results obtained from gauged basins (Wang et al., 2023). The parameter regression method is sensitive to the effects arising from the multitude of potential parameter combinations, and the subjective nature of selecting basin attributes may pose limitations, particularly when working with datasets characterized by small sample sizes (Jafarzadegan et al., 2020). In this case, state-of-the-art machine learning is regarded as a novel approach to parameter transfer (Yi et al., 2014), including hierarchical cluster analysis and unsupervised neural network, however, the effectiveness of machine learning for parameter transfer is subject to the density of hydrological stations. Therefore, there is a clear requirement for approaches that be used to hydrological model parameters transfer more reliably. A hydrological similarity-based regionalization method integrates a Random Forest model and cluster analysis designed for hydrological model parameters transfer in ungauged basins in this research. The ensemble method presents an enhanced approach for recognizing and utilizing the intrinsic similarities across basins and is better equipped to handle the complex interactions and nonlinear relationships among basin characteristics (Kratzert et al., 2019), thereby, improving the accuracy and robustness of the hydrological model parameters estimation for ungauged basins.

Statistical techniques play a key role in constructing the FDC, relying on fitting statistical distribution functions to observe FDCs from multiple gauged locations within a basin (Dey et al., 2024). Statistical techniques do not rely on specific assumptions about hydrological processes, which allows them to exhibit considerable flexibility and adaptability when addressing diverse climatic and geomorphological conditions. In non-stationary climatic conditions, the inherent flexibility of statistical techniques enables them to better adapt to predict future streamflow variations in certain scenarios (Müller and Thompson, 2016). A regional regression relationship between the fitted distribution parameters and basin characteristics is then established (Leong and Yokoo, 2021; Sauquet and Catalogne, 2011). However, such statistical approaches may not fully capture the underlying hydrological processes influencing the FDC, affecting the accuracy of FDC predictions for basins beyond the specific study regions and periods for which the statistical relationships have been developed (Cislaghi et al., 2020). As Fig. 1 indicates, an array of physical factors—including precipitation, evaporation, infiltration, runoff, topography, soil properties, and vegetation cover—play a crucial role in shaping the FDC. Among these, precipitation is determined as the primary factor affecting the FDC's shape, where its volume, intensity, form, and temporal distribution determine the curve's shape (Coopersmith et al., 2012; Ghotbi et al., 2020b). Topography and soil permeability stand as essential natural determinants of the FDC shape by governing runoff and infiltration, thereby shaping the basin's hydrological response. Highly permeable soils enhance precipitation infiltration, boosting groundwater levels and leading to a more gradual FDC decline. Low-permeability soils, however, restrict infiltration, causing a steeper FDC indicative of rapid precipitation response and limited baseflow support (Chouaib et al., 2018). Hence, considering the influence of these physical factors is crucial for accurately constructing FDCs. Previous studies have adopted a derived distribution approach by integrating stochastic representations of precipitation inputs with deterministic models of the precipitation-runoff transformation process (Costa and Fernandes, 2017). It is essential to highlight that the derived distribution approach accounts for the key processes affecting temporal streamflow variability, including climate forcing and catchment characteristics. However, rainfall-runoff processes in actual basins display a greater level of complexity and regional heterogeneity than previously assumed in initial studies (Leong and Yokoo, 2021). A clear need exists for approaches or conceptual models that enhance the general study of the process controls of the FDC. Ghotbi et al. (2020a) proposed a new stochastic model based on process-based methods for constructing FDCs. The stochastic model involves partitioning streamflow into its fast and slow components (Duncan, 2019), the distribution of each component is fitted separately before being integrated to model the FDC of streamflow. In this way, the FDC of streamflow is computed as a statistical sum of a fast flow duration curve and a slow flow duration curve, and accounting for the dependence structure between fast and slow flows, the dependence is addressed through the utilization of a bivariate copula function (Ghotbi et al., 2020a). The stochastic model attributes the shape of the FDC to physical processes and provides an effective pathway for predicting FDC in ungauged basins. This study advances an extended version of the stochastic model proposed by Ghotbi et al. (2020a), including dividing the streamflow into multiple delayed flows to model their interdependence separately explore their process controls independently, and later combine them to model the FDC.

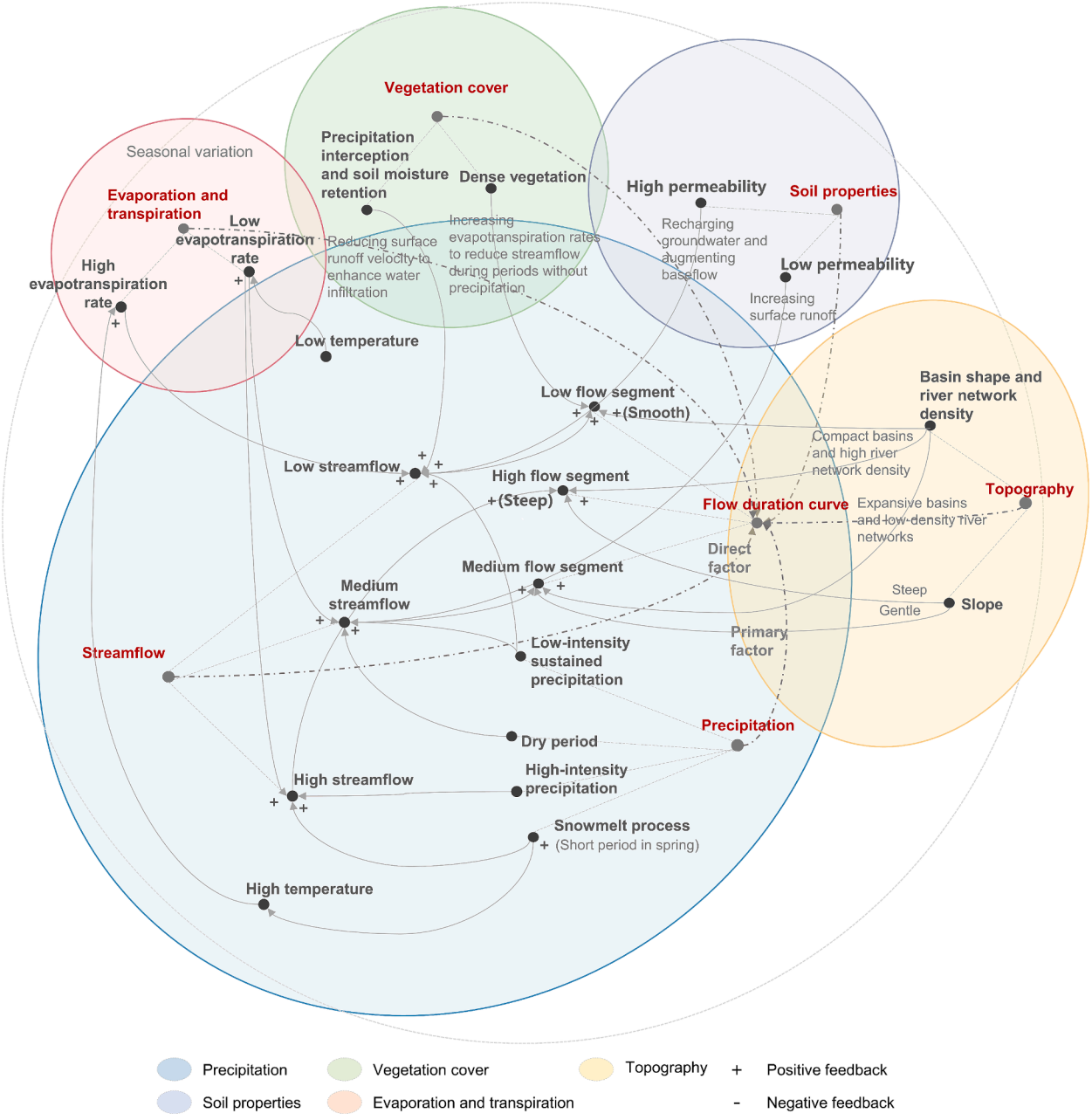

Fig. 1. Exploring the impact of various physical factors on the shape of FDCs through causal loop diagrams.

The challenge of modeling FDCs and understanding the process controls of FDC's shape from the fact that streamflow represents a synthesis of catchment hydrological responses functioning across multiple time scales (Cheng et al., 2012; Leong and Yokoo, 2021). For simplicity, as a first step, streamflow is divided into various physical processes. Through numerical simulations utilizing a physically based hydrological model, Chouaib et al. (2018) proposed separating streamflow into two components: Fast flow indicative of surface runoff, and slow flow, representative of subsurface streamflow and groundwater flow, to generate deeper insights into the shape of the FDCs, including their process controls. Binary or two-component separation of streamflow into a fast and a slow component are often based on arbitrary choices of separation parameters and also merge different delayed components into one baseflow component and one Baseflow Index (BFI) (Aboelnour et al., 2021). The generation of streamflow during dry weather frequently is derived from the drainage of multiple sources. To deal with this, Stoelzle et al. (2020) proposed to extend the BFI by a Delayed Flow Index (DFI) to account for the dynamics of multiple delayed contributions to streamflow. The DFI relies on Characteristic Delay Curves (CDCs), wherein the identification of breakpoint estimates serves to avoid rather subjective separation parameters and allows for distinguishing four delayed classes: Short-delay class indicates near-surface runoff or transient groundwater contributions, intermediate-delay class reflects certain groundwater or snowmelt contributions, whose response times lie between those of short-delay and long-delay flows. The long-delay class represents deep groundwater flows, which have a longer response time to precipitation events and the baseline-delay class represents the slowest water source contributions, which maintain the stability of streamflow during dry periods. To construct the FDC with greater accuracy, this method was employed to separate the streamflow. Copulas are functions that link the marginal distribution functions of random variables to their joint distribution function, thereby describing the dependence structure between these variables (Sklar, 1973). The great advantage of copulas lies in constructing the

joint distribution function through two distinct processes: Modeling the dependence and the modeling of the marginal distribution functions (Vernieuwe et al., 2015). As this property allows for modeling a large variety of joint probability functions, copulas have gained popularity among the hydrological community. Liu et al. (2020) developed a Bayesian-copula-based frequency analysis method for snowmelt flood risk assessment and provided a scientific basis for flood management and disaster prevention. Tahroudi et al. (2020) used copula functions to analyze the joint signatures of groundwater level deficiency and rainfall deficiency in the Naqadeh sub-basin in the Lake Urmia Basin. In the study of Nazeri Tahroudi et al. (2021), the application of the conditional density of bivariate copulas was investigated in flood routing via a copula-based approach. However, their use is generally restricted to the description of bivariate dependence. Recent advancements have introduced Vine copulas, enabling the description of multi-dimensional dependence through a sequential combination of two-dimensional copulas (Vernieuwe et al., 2015). Pouliasis et al. (2021) developed a methodology for generating a synthetic time series of hydro-climatic variables using the Vine copula model, and Wu et al. (2021) proposed a novel drought prediction model via conditional Vine copulas for China. This study explores the application of Vine copulas to integrate all pertinent dependencies among the four delayed flows, aiming to improve the accuracy and reliability of constructed FDCs by capturing the complex interdependencies among various streamflow patterns.

Through the separation of streamflow into delayed flows, four distinct types corresponding to specific hydrological processes are identified. The simulation of intrinsic connections among these processes improves the accuracy of FDC predictions and the robustness based on the understanding of their physical controls (Janssen and Ameli, 2021). In this context, precise estimation of copula function parameters is essential for capturing the dependence between various delayed flows through the Vine copula structure. Copula parameters, reflecting the strength of mutual dependence between two variables (Hao and Singh, 2016), are typically estimated through a theoretical relationship (if exists) with empirical dependence measures, including Kendall's τ and Spearman's ρ, or through inference from empirical multivariate probability distributions of data (Sadegh et al., 2017). The most common practice for parameter estimation is using local optimization algorithms, such as the Newton-Raphson method (Gräler et al., 2013; Ribatet and Sedki, 2013). Local optimization approaches benefit from efficient (mostly gradient-based) search algorithms but suffer from susceptibility to getting trapped in local optima and are not capable of characterizing the underlying uncertainties (Sadegh et al., 2017). Global optimization approaches and the Bayesian approach have also been explored for inferring copula parameters (Kwon and Lall, 2016; Parent et al., 2014; Smith et al., 2012). Global optimization approaches are essential in parameter estimation, particularly for complex models susceptible to local minima (Bangyal et al., 2021). For example, the differential evolution algorithm is particularly effective in noisy environments where gradient-based methods fail. It handles complex objective functions by evolving a solution population through mutation and crossover, thus accommodating diverse problem landscapes (Poovathingal and Gunawan, 2010). The Bayesian approach provides a robust framework to account for the uncertainty that originates from model selection, parameter uncertainty, and shortage of data, the latter one being the major source of uncertainty in the result (Kaplan, 2021). Furthermore, the Bayesian approach allows for the use of prior distributions that represent previous knowledge and beliefs on the distribution of model parameters. Prior distributions help to improve model convergence and parameter estimation in the Bayesian approach (Van de Schoot et al., 2021). Sadegh et al. (2017) presented a developed Multivariate Copula Analysis Toolbox (MvCAT) which includes a wide range of copula families with different levels of complexity. MvCAT employs a Bayesian approach with a residual-based Gaussian likelihood function for inferring copula parameters and estimating the underlying uncertainties. The approximation of the posterior distribution through random sampling is achieved with a developed hybrid-evolution Markov Chain Monte Carlo (MCMC) algorithm. This algorithm is capable of not only finding a good estimate of the global optimum but also providing an approximation of the underlying uncertainties within a Bayesian approach (Mazdiyasni et al., 2018; Sadegh et al., 2017). Thereby, the MvCAT is applied to estimate copula parameters in this study and is compared with the copula parameters obtained through the method based on Kendall's tau to enhance the precision of predicted FDCs in ungauged basins.

The objective of this study is to propose a framework, integrating regionalization and process-based approaches for predicting FDCs at ungauged basins. This framework aims to improve the accuracy of the FDC predictions in ungauged basins and deepen our comprehension of the process controls influencing the shape of FDCs, which focuses on the following three aspects.

- A mixture copula model is designed to explore process controls of FDCs, delineating hydrological processes from diverse delayed flow components.
- Bayesian-based parameter estimation for the construction of Vine copula structures to quantify the dependence of diverse delayed components is developed to further improve the accuracy of the predicted FDC.
- A regionalization method based on hydrological similarity is developed to estimate the hydrological model parameters in ungauged basins.

The rest of the study is organized as follows: Section 2 presents the data description and analysis as study cases. Section 3 introduces the methodology for proposing the framework for predicting FDC at ungauged basins. Section 4 outlines the results of applying the framework to the MOPEX basins in the United States. Section 5 discusses the evaluation results of the framework for FDC prediction in another study case. Section 6 presents the key conclusions and offers suggestions for future research endeavors.

2. Study cases

The MOPEX database is a product of a collaborative international initiative involving academic institutions, governmental agencies, and non-governmental organizations, with the primary objective of advancing atmospheric and hydrological modeling, also the MOPEX database includes historical hydrometeorological data and land surface characteristics for 438 hydrologic basins in the United States (http://hydrology.nws.gov/pub/gcip/mopex/US_Data) (Duan et al., 2006). The developed framework is illustrated on the selected 201 MOPEX basins (124°03′W–71°05′W, 27°13′N-48°29′N) ranging from 52 km^2 to 3988 km^2 as shown in Fig. 2a. These basins have been selected for encompassing a range of geographical features and climates, from arid to humid, and the generalization capability of the developed framework under various conditions can be enhanced across diverse environmental conditions (Chouaib et al., 2018). For constructing the developed

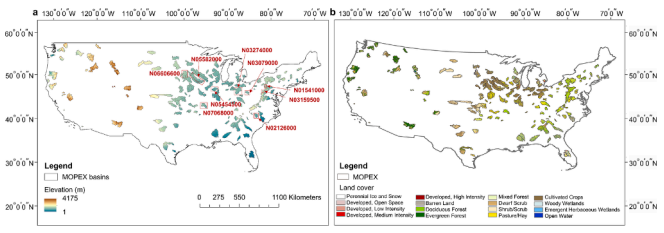

Fig. 2. a, DEM map with the boundaries of 201 MOPEX basins, highlighting selected ungauged basins with red rectangles. b, Land cover map with the boundaries of 201 MOPEX basins. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

framework for predicting FDCs at ungauged basins, four main types of data were utilized: Observed streamflow data, meteorological data, geographical information data, and remote sensing data. Streamflow data from 1982 to 2000 for 201 basins were from the MOPEX database, with the streamflow for the 1982 year selected as a warm-up period, almost 72 % of the streamflow data samples (1983–1995) were used for calibration, and a 5-year period spanning from 1996 to 2000 was utilized to verify the developed framework. Daily meteorological data from 1982 to 2000 including precipitation, mean, maximum, and minimum daily temperature, and potential evapotranspiration (ET_0) was also obtained from the MOPEX database (Duan et al., 2006; Tounsi et al., 2023). The MOPEX database provided data on 16 types of land use within these basins (Fig. 2b), according to the International Geosphere-Biosphere Programme (IGBP) classification, was utilized to calculate the effective depth (Z_E), a critical variable influenced by land cover and soil type that determines a basin's maximum sustainable evaporative flux from its water reserves (Daly et al., 2019). These data were used as clustering indicators in a K-means clustering approach to measure hydrological similarity across different basins. A 30 m resolution Digital Elevation Model (DEM) obtained from Advanced Spaceborne Thermal Emission (https://lpdaac.usgs.gov/products/astgtmv003/) and Reflection Radiometer Global Digital Elevation Model Version 3 (ASTGTMV003) was employed to obtain elevation information for the study area (Xin et al., 2021). Normalized Difference Vegetation Index (NDVI) data from 1982 to 2000, with a spatial resolution of 1/8° and a temporal resolution of hourly, were also employed in a regionalization method based on hydrological similarity for basin clustering.

3. Methodology

The developed framework (Fig. 3) is summarized as follows: (1) Developing a regionalization method based on hydrological similarity to transfer hydrological model parameters in ungauged basins. (2) Constructing the mixture copula model for predicting FDC, which involves

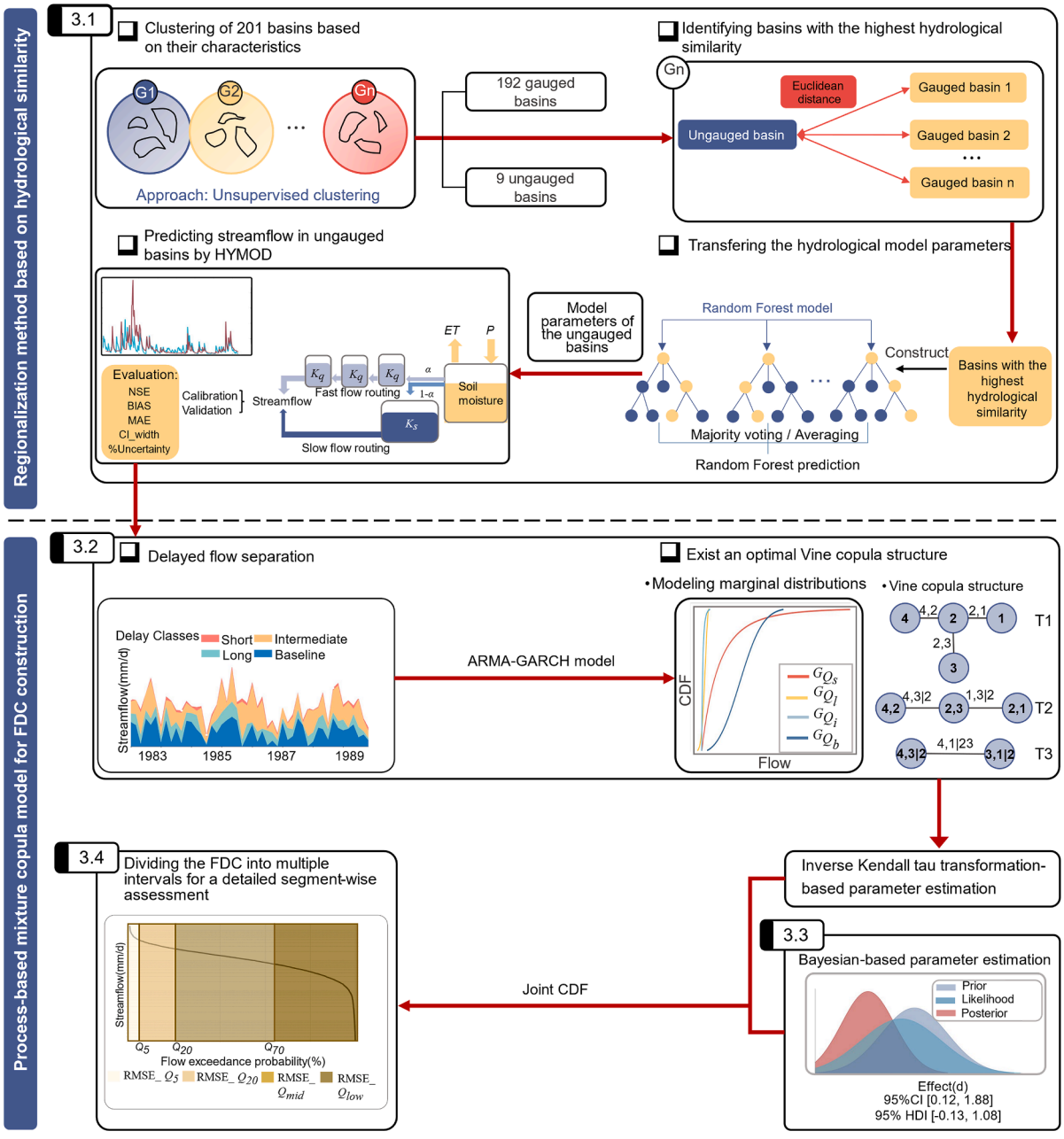

Fig. 3. Framework for integrating regionalization and process-based approaches for flow duration curve prediction in ungauged basins.

Table 1
Summary of various climate, topographic, land-use, and geopedological descriptors for the 201 MOPEX basins.

Notation	Descriptor	Computation	Unit	Data Source
P	Mean daily precipitation	Total depth of precipitations over the recorded period (19 years) divided by the number of days, aggregated spatially to the catchment scale	mm/d	MOPEX database
T_{max}	Maximum daily temperature	Total of maximum daily temperatures over the recorded period (19 years) divided by the number of days, aggregated spatially to the catchment scale	°C	MOPEX database
T_{min}	Minimum daily temperature	Total of minimum daily temperature over the recorded period (19 years) divided by the number of days, aggregated spatially to the catchment scale	°C	MOPEX database
ET_0	Mean daily potential evapotranspiration	Total of mean daily potential evapotranspiration over the recorded period (19 years) divided by the number of days, aggregated spatially to the catchment scale	mm/d	MOPEX database
Q	Mean streamflow	Total depth of streamflow over the recorded period (19 years) divided by the number of days, aggregated spatially to the catchment scale	mm/d	MOPEX database
Area	Catchment area	—	km^2	MOPEX database
Slope	Average basin slope	—	%	United States Geological Survey
NDVI	Half-month normalized vegetation index	Total of half-month normalized vegetation index over the recorded period (19 years) divided by the number of half-months, aggregated spatially to the catchment scale	—	MOPEX database
Land use	Evergreen Needleleaf Forest Evergreen Broadleaf Forest Deciduous Needleleaf Forest Deciduous Broadleaf Forest Mixed Forest Closed Shrublands Open Shrublands Woody Savannah Savannahs Grasslands Permanent Wetlands Croplands Urban and Built-Up Cropland / Natural Vegetation Mosaic Snow and Ice Barren or Sparsely Vegetated Water Bodies	—	—	(Daly et al., 2019)
$Z_{E,50\%}$ $Z_{E,60\%}$ $Z_{E,80\%}$	The characteristic depth of the soil layer	The characteristic depth of the soil layer supporting Z_E is calculated from the land use, defined according to the IGBP classification with 17 different land uses. Each land use is assumed to be characterized by a root distribution	m	(Daly et al., 2019)

separating the streamflow into diverse delayed flow contributions and capturing their dependence by a Vine copula structure. Through the constructed model, the characteristics of FDC are linked to underlying physical processes to improve the accuracy of FDC prediction. (3) Bayesian-based copula parameter estimation is compared with the inverse Kendall tau transformation-based parameter estimation to further improve the accuracy of the predicted FDC.

3.1. Regionalization method based on hydrological similarity

Streamflow prediction in ungauged basins is a premise for constructing the FDC, and accurate estimation of hydrological model parameters is crucial for streamflow prediction. Therefore, a regionalization method based on hydrological similarity has been designed to facilitate parameter transfer. Through a wide consideration of descriptors including climate, topography, land use, and soil geology, basin characteristics can be matched, and more appropriate hydrological model parameters and hydrological behavior predictions for ungauged basins are provided. Compared to traditional regionalization methods, the hydrological similarity based-regionalization method integrates the ensemble learning mechanism and bootstrap sampling principles. Through incorporating multiple decision trees, the risk of overfitting is reduced and the predictive accuracy is enhanced. In addition, the diversity of training data is increased through bootstrap sampling. The hydrological model could be of any type, with the Hydrological Model (HYMOD) serving as an illustrative example of the regionalization method. HYMOD is a daily-step hydrological model based on a nonlinear water storage capacity distribution function. The runoff generation process is described by a simple precipitation-excess model based on the probability-distributed principle (Quan et al., 2015). The regionalization method includes three steps: Basin clustering, selection of the basins with the highest hydrological similarity, and hydrological model parameters transfer. These steps are detailed as follows.

Step 1: Basin clustering is employed to refine the search for the basins with the highest hydrological similarity. After mapping out the hydrographical boundaries of the 201 MOPEX basins, each basin is characterized by a unique set of descriptors, as shown in Table 1. These descriptors are employed to group basins into clusters using the K-means process (Gaviria and Carvajal-Serna, 2022). The K-means clustering process begins with determining the optimal number of clusters (K) from a potential range of 2 to 10, adapting diverse clustering scenarios (Kodinariya and Makwana, 2013). Following this, the K-means clustering algorithm is applied to each number within this range, and the Correlation Distance (Eq. (2) is calculated for each cluster count. The K value is typically where the Correlation Distance is at its minimum (Gaviria and Carvajal-Serna, 2022), then, K centroids are randomly placed among the 201 basins, and the Euclidean Distance (Eq. (3) are calculated to assess the hydrological similarity between each basin and its cluster centroid. After the initial assignment, centroids are updated by recalculating the average values of descriptors for each cluster, thereby refining the centroids for the subsequent iterations. The iterative process continues until the algorithm converges, ensuring each basin is appropriately classified. The basin that lies closest to each cluster's centroid is then considered an ungauged basin within that group (Gaviria and Carvajal-Serna, 2022). N is the total number of descriptors, x_{k_i} represents the i th descriptor of any basin within the k th cluster, C_{k_i} represents the i th descriptor of the centroid in the k th cluster. The $\mathrm{corr}(x_{k_i}, C_{k_i})$ calculates the Pearson correlation coefficient between x_k and C_k (Eq. (1).

$$\mathrm{corr}(x_{k_i}, C_{k_i}) = \frac{\Sigma(x_k - \overline{x_k})(C_k - \overline{C_k})}{\sqrt{\Sigma(x_k - \overline{x_k})^2 \Sigma(C_k - \overline{C_k})^2}} \quad (1)$$

$$\text{Correlation Distance} = \frac{\Sigma_{i=1}^{N}(1 - \mathrm{corr}(x_{k_i}, C_{k_i})}{N} \quad (2)$$

$$\text{Euclidean Distance} = \sqrt{\Sigma_{i=1}^{N}(x_{k_i} - C_{k_i})^2} \quad (3)$$

Step 2: The choice of the basins with the highest hydrological similarity is determined through the application of distance metrics, the Euclidean distance is calculated within the cluster to which the ungauged basin belongs, serving as the measure of distance between the ungauged basin and other gauged basins. The basins with the lowest Euclidean distance are chosen as the basins with the highest hydrological similarity.

Step 3: The transfer of hydrological model parameters between ungauged basins and those with the highest hydrological similarity incorporates the mechanism of the Random Forest model. In the parameter transfer process, the strength of multiple decision trees is leveraged to enhance predictive accuracy and mitigate overfitting through result averaging (Zounemat-Kermani et al., 2021). The procedure for the transfer of hydrological model parameters is as follows.

Step 3.1: Climate, topography, land use, and geopedological descriptors (Table 1), along with hydrological model parameters from the basins with the highest hydrological similarity serve as the training inputs, and hydrological model parameters are the target outputs. The significance of each descriptor to the target variable is assessed using Pearson correlation coefficients, and descriptors with an absolute correlation value greater than 5 with the target variable are retained. Similar descriptors within the same category are averaged to combine features, simplifying the feature set to enhance model accuracy and efficiency. For example, for descriptors of various land use types, the average is calculated to represent the overall characteristics of that category.

Step 3.2: A Random Forest model with 100 decision trees is created to train on the selected descriptors. Bootstrap sampling is used to create diverse training sets for each decision tree, increasing data diversity and model generalization capabilities. The construction of multiple decision trees proceeds in parallel, with descriptor selection for splits being randomized to add further diversity. The predictions from all decision trees are then aggregated to produce the final output.

Step 3.3: Descriptors from ungauged basins are preprocessed to ensure compatibility with the trained model's descriptors. The processed descriptors are then input into the Random Forest model to predict the hydrological model parameters, and more detailed descriptions of the transfer of hydrological model parameters are presented in *Supporting Information* (section S1.1).

Step 4: These simulated and predicted streamflow with HYMOD driven by the hydrological parameters in the ungauged basins are crucial for the process-based mixture copula model for FDC construction. Therefore, several metrics are used to evaluate the prediction precision of streamflow for calibration and validation periods: Nash-Sutcliffe Efficiency (NSE), Mean Absolute Error (MAE), Bias, Confidence Interval Width (CI_width), and Percentage of Uncertainty (%Uncertainty) (Table S1) (Atieh et al., 2017). NSE serves as a robust indicator of hydrological model predictive accuracy and efficiency, which distinguishes between models of varying quality. MAE quantifies the average magnitude of errors in predictions, providing a straightforward assessment of model error magnitude. Bias helps identify systematic over or underestimations through the results of streamflow prediction (Chicco et al., 2021). CI_width and %Uncertainty is employed for uncertainty analysis. These evaluation metrics collectively afford a thorough insight into the accuracy and reliability of streamflow predictions in ungauged basins. The differences in metrics between the calibration and validation periods assess the temporal transferability of estimated hydrological model parameters. Effective temporal transferability is crucial for validating regionalization methods (Johannessen et al., 2019).

3.2. Process-based mixture copula model for FDC construction

3.2.1. Delayed flow separation

Since the focus of this paper is to illustrate a process-based mixture copula model for constructing FDCs by combining different delayed flows, as a first step, the four delayed components of streamflow are estimated from the predicted streamflow using a delayed flow separation method. The dynamics of multiple delayed contributions to streamflow are considered through the method. A more detailed assessment of water resource sustainability in basins during dry periods is allowed by this refinement (Stoelzle et al., 2020). The process of delayed flow separation is outlined as follows.

Step 1: To identify the minimum flow points within each time block, thereby recognizing characteristics representing delayed flows, the flow time series is divided into non-overlapping consecutive blocks of N days. The minimum flow within each time block is identified, and the minimum value of each block is compared to the minimum of the two adjacent blocks. If a factor $f = 0.9$ times the minimum value is less than or equal to the two adjacent minima, a turning point is defined. The turning points are linked with straight lines to become the delayed flow hydrograph (Wright et al., 1990; Yeh and Chen, 2022). Between turning points, the delayed flow values are derived through linear interpolation. If the interpolated delayed flow exceeds the original streamflow value, the interpolated delayed flow is replaced through the original streamflow (Stoelzle et al., 2020). In hydrology, delayed flow generally corresponds to baseflow, with instantaneous flow primarily contributed by surface runoff, equaling the total flow minus the baseflow (Wang et al., 2022). Constructing the delayed flow hydrograph enables the effective separation of baseflow and surface runoff, moreover, facilitating the calculation of the delayed flow's proportion in the streamflow, described as the DFI.

Step 2: The delayed flow index for a given N (DFI_N) is then calculated as the ratio of the sum of delayed flow to the sum of streamflow from 1 to 90 days. The calculation of the DFI relies on constructing the delayed flow hydrograph, intended to quantify the contributions of various delayed flows to the total flow within a basin (Anderson et al., 2023), and the quantification is crucial for understanding the contribution of groundwater recharge and storage processes to streamflow. DFI_0 represents the case of no separation, with the delayed flow series being identical to the streamflow series ($DFI_0 = 1$). With an increase in N, due to the turning points being set wider apart, more streamflow peaks (contributions with shorter delays) are excluded from the separation, resulting in a decrease in the DFI (Stoelzle et al., 2020).

Step 3: Based on the DFI for different N values, CDCs are plotted to illustrate the relationship between the block size N and DFI_N. CDCs converge to a basin-specific asymptotic value for a large N (Reimann et al., 2018). Accordingly, it is proposed by Stoelzle et al. (2020) that the proportion of delayed flow stays nearly constant, beyond a certain Nmax, even if N is further increased ($N >$ Nmax). This threshold, "Nmax", effectively represents the maximum delay for all contributing sources within the basin. For cases where N exceeds 60 days, the CDCs tend to flatten, indicating that 60 days are sufficient to capture all annual minimum flows across the basins, and increasing the N value no longer significantly changes the results of the delayed flow analysis, thus the Nmax is set to 60 days. A decrease in the slope of the CDC indicates a transition from faster to slower contributing sources (stores) in the basin (Stoelzle et al., 2020). Such specific values of N are defined as breakpoints splitting the CDCs into piecewise linear segments with different slopes (Miller et al., 2015). The CDC is divided into several segments with Nmax and two breakpoints, each segment representing different types of delayed flow. In addition, based on the DFI at specific points, the contribution ratio of each delayed flow category to the streamflow is quantified.

Step 4: The delayed contributions to streamflow are classified into the following four delayed classes and quantified as the ratio of each component to the total annual streamflow (ranging between 0 and 1) for each class. Therefore, the four classes are calculated as: D_s(short-delay class) = $DFI_0 - DFI_{BP1}$; D_i(intermediate-delay class) = $DFI_{BP1} - DFI_{BP2}$; D_l(long-delay class) = $DFI_{BP2} - DFI_{60}$; D_b(baseline-delay class) = DFI_{60}. Streamflow contributions in each delayed class are then calculated based on the basin-specific average streamflow (Stoelzle et al., 2020). The four types of delayed flow contributions are as follows. Short-delay flow (Q_s): Flow between no delay (N = 0, equivalent to the original flow series) and the first breakpoint BP1, which represents faster-responding water sources, such as near-surface runoff or transient groundwater contributions. Intermediate-delay flow (Q_i): Flow between two breakpoints, reflecting water sources in the basin with intermediate delays, such as certain groundwater flows or snowmelt contributions, whose response times lie between those of short-delay and long-delay flows. Long-delay flow (Q_l): Flow from the second breakpoint BP2 to N = 60 days. This type of flow represents longer time-scale water source contributions, which have a longer response time to precipitation events. Baseline-delay flow (Q_b): Equal to the DFI_{60} value (N = 60 days), this flow represents the slowest water source contributions, typically associated with deep groundwater, which maintains the stability of flow during dry periods. Through delayed flow separation, the formation mechanism of the FDC shape and the variations in different segments of the FDC are better understood and explained.

3.2.2. Capturing the dependence of delayed flows

The FDC is understood to be the complementary Cumulative Distribution Function (CDF) of the observed streamflow, and the CDF of the observed streamflow can be approximated by the joint CDF of four delayed flows (Eq. (4). The accuracy of the predicted FDC is enhanced through an effective dependency structure among these four delayed flows. The joint CDF of four delayed flows (derived from the predicted streamflow) is quantified through a Vine copula structure (Eq. (5).

$$G(q) = G(q_s, q_i, q_l, q_b) \tag{4}$$

$$G(q_s, q_i, q_l, q_b) = C(G_{Q_s}(q_s), G_{Q_i}(q_i), G_{Q_l}(q_l), G_{Q_b}(q_b)) \tag{5}$$

where Q_s, Q_i, Q_l, and Q_b are random variables representing the short-, intermediate-, long- and baseline-delay flow, respectively. $G(q)$ is the CDF of the observed streamflow, and $G(q_s, q_i, q_l, q_b)$ is the joint distribution function of four delayed flows. $G_{Q_s}(q_s)$ is the CDF of short-delay flow. $G_{Q_i}(q_i)$ is the CDF of intermediate-delay flow. $G_{Q_l}(q_l)$ is the CDF of long-delay flow. $G_{Q_b}(q_b)$ is the CDF of baseline-delay flow and C is a copula function that quantifies the joint CDF of $G_{Q_s}(q_s)$, $G_{Q_i}(q_i)$, $G_{Q_l}(q_l)$ and $G_{Q_b}(q_b)$. More information for the Vine copula is provided in the *Supporting Information* (section S1.5). Modeling the joint distribution through the Vine copula structure involves three main steps: Estimating the marginal distributions of the four delayed flows (assumed to be independent and identically distributed, iid); selecting the optimal Vine copula structure; calculating the joint distribution of the four delayed flows through the Vine copula structure.

Step 1: Transforming delayed flow data into a time series format enhances the capability for trend and seasonal analyses. To ensure that the time series satisfies statistical properties and underlying assumptions, various tests are conducted to analyze different characteristics of the time series (section S1.4). The copula estimation based on pseudo copula data requires that the data be iid. However, this is not the case with multivariate time series data, which exhibit heteroscedasticity and autocorrelation, making them unsuitable for standard copula modeling without prior transformations (Czado and Nagler, 2022). Therefore, before constructing a copula model, it is essential to choose a marginal distribution for the time series model. To eliminate conditional heteroscedasticity and autocorrelation, the ARMA(*p,q*)-GARCH(1,1)-τ model is selected to construct the marginal distributions, and then, the optimal Vine copula structure is chosen to establish the joint CDF for the four delayed flows (Mejdoub and Ghorbel, 2018). The ARMA(*p,q*)-GARCH(1,1)-τ model assumes that the error terms follow a Student's *t*-distribution. The ARMA(*p,q*) component captures the short-term dependencies within the time series, while the conditional variance (volatility) of the time series is constructed by the GARCH(1,1). The assumption of a skewed τ distribution is utilized to handle the thickness of the tails in the data (the occurrence of extreme values) (Ling and McAleer, 2003), expressed as:

$$\begin{cases} r_{i,t} = \varphi_0 + \varphi_1 r_{i,1} \times r_{i,t\text{-}1} + \ldots + \varphi_p r_{i,p} \times r_{i,t\text{-}p} + \theta_1 a_{i,1} \times \varepsilon_{i,t\text{-}1} + \ldots + \theta_q a_{i,q} \times \varepsilon_{i,t\text{-}q} + \varepsilon_{i,t} \\ \varepsilon_{i,t} = \sigma_{i,t} \times \mu_{i,t} \\ \mu_{i,t} \text{ skewed t}(\text{skew}_i, \text{shape}_i) \\ {\sigma_{i,t}}^2 = \omega_i + \alpha_{i,t} \times {\varepsilon_{i,t\text{-}1}}^2 + \beta_{i,1} \times {\sigma_{i,t\text{-}1}}^2 \end{cases} \tag{6}$$

where the return series $r_{i,t}$ is modeled with an ARMA(*p, q*) model with parameters $\varphi_1, \ldots, \varphi_p$ and $\theta_1, \ldots, \theta_q$; $\varepsilon_{i,t}$ is the residual of the return series; $\sigma_{i,t}$ is the conditional variance (or standard deviation) at time *t*; $\mu_{i,t}$ is used to represent the location parameter of the skewed τ distribution. Skewness and kurtosis are captured through the skewed τ distribution with parameters skew_i and shape_i; ω_i, $\alpha_{i,t}$ and $\beta_{i,1}$ are parameters of the GARCH(1,1) model.

Step 2: Residuals and conditional variances are extracted from the fitted ARMA(*p,q*)-GARCH(1,1)-τ model. Model residuals represent the error terms after accounting for the mean and volatility of the time series, while conditional variances reflect the volatility of the time series at different time points. By dividing each residual following its corresponding conditional variance, standardized residuals are obtained, thereby removing the effects of volatility, and providing a standardized data basis for subsequent estimation of marginal distributions. To determine whether the ARMA(*p,q*)-GARCH(1,1)-τ model has successfully eliminated the effects of autoregressive conditional heteroscedasticity and autocorrelation, the standardized residuals and their squared values are subjected to the Ljung-Box (LB) test and the ARCH-LM test. The detailed procedure and results of the tests are documented in Appendix B. The results indicate that after employing the ARMA(*p,q*)-GARCH(1,1)-τ model, the percentiles of the FDCs (the different parts of the streamflow) are iid. Then, the Empirical Cumulative Distribution Function (ECDF) is applied to the standardized residuals, mapping the residuals to values in the [0,1] interval as estimates of marginal distributions. Through ECDF, each standardized residual is transformed into its relative position within its distribution, and these relative position values can be directly used for constructing copula models.

Step 3: Before applying the copulas, the dependence between the variables must be investigated through Kendall's tau (Dehling et al., 2017), the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are applied to evaluate multiple copula families for each pair of variables. Both AIC and BIC are statistical measures commonly used to evaluate and compare the goodness of fit of statistical

models, while avoiding overfitting (Lin et al., 2017). Vine copula structures with lower information criteria are considered superior, as they provide a good fit to the data while maintaining structure simplicity (Czado and Nagler, 2022). Then, Kendall's tau is utilized to estimate the copula parameters, for any given pair of variables X and Y, the calculation formula for Kendall's Tau is as presented in Eq. (7). Besides, employing the relationship between Kendall's Tau and copula parameters within copula theory (Eq. (8), Kendall's Tau value is converted into parameters for a specific copula model through an inverse transformation. Here, n represents the number of observation pairs; $C(u, v)$ is the bivariate copula function; u and v is the CDF of the marginal distributions.

$$\tau = \frac{2}{n(n\text{-}1)} \sum_{i<j} \text{sign}(x_i\text{-}x_j) \times \text{sign}(y_i\text{-}y_j) \tag{7}$$

$$\tau = 4 \iint C(u, v) dC(u, v)\text{-}1 \tag{8}$$

Step 4: Vine copula utilizes a series of tree structures (R-Vines) to model complex dependencies in multidimensional data. Each tree represents a layer of dependency, with the first tree defining the basic dependencies between variables, and subsequent trees modeling conditional dependencies. Each edge corresponds to a bivariate copula, which is applied to describe the dependency between two variables (or combinations of variables) (Coblenz, 2021). Based on the selected optimal Vine copula structure, conditional distributions are constructed layer by layer. By integrating the marginal distributions with bivariate copulas, the joint CDF of the four delayed flows is calculated. In the first layer (Tree1), the joint CDFs for pairs of variables are calculated through the specified copula function. For example, the relationships between (u_1, u_2), (u_2, u_3) and (u_3, u_4) are expressed as $C_{1,2}(u_1, u_2)$, $C_{2,3}(u_2, u_3)$ and $C_{3,4}(u_3, u_4)$. Then, conditional CDFs are computed for the connected pairs from the first tree, such as: $C_{13|2}(u_1, u_3|u_2)$ and $C_{24|3}(u_2, u_4|u_3)$, where the calculations for these conditional probabilities are detailed in Eq. (9). Finally, under the conditions set by Tree2, the highest-order conditional CDF, $C_{14|23}(u_1, u_4|u_2, u_3)$, is calculated. The formula for the joint CDF of u_1, u_2, u_3 and u_4 is shown in Eq. (10), and the exceedance probability along the horizontal axis of the FDC is determined by the equation: Exceedance probability = 1 - G (q_s, q_i, q_l, q_b). The vertical axis of the FDC represents predicted streamflow values, arranged from highest to lowest. Each streamflow value is connected to its corresponding probability point, forming the FDC, more descriptions of the process-based mixture copula model are presented in *Supporting Information* (section S1.3)).

$$C_{13|2}(u_1, u_3|u_2) = \frac{C_{12,23}(u_1, u_2, u_3)}{C_2(u_2)} \tag{9}$$

$$\begin{aligned} C(u_1, u_2, u_3, u_4) =& C_{1,2}(u_1, u_2) \times C_{2,3}(u_2, u_3) \times C_{3,4}(u_3, u_4) \\ & \times C_{13|2}(u_1, u_3|u_2) \times C_{24|3}(u_2, u_4|u_3) \times C_{14|23}(u_1, u_4|u_2, u_3) \end{aligned} \tag{10}$$

3.3. Bayesian-based parameter estimation for copula model

An inverse Kendall tau transformation, as a non-parametric statistical method, removes the necessary for strict assumptions regarding the distribution of data, thus broadening the estimated copula parameters' applicability, however, the accuracy of the estimation results may be affected when the data structure is complex (Dehling et al., 2017). Bayesian-based parameter estimation, integrates prior knowledge with observed data, thus enabling the provision of relatively accurate estimates even when sample sizes are small. In addition, the capability to handle models with complex data structures is facilitated through MCMC numerical methods. As a result, copula parameters are estimated through Bayesian-based parameter estimation and compared with the inverse Kendall tau transformation-based parameter estimation to construct more precise FDCs (Hill and Spall, 2019; Krapu and Borsuk, 2022). Bayesian-based parameter estimation consists of the following steps.

Step 1: For each copula function parameter in Bayesian analysis, a prior distribution is established to represent initial understanding. When specific information about a parameter's prior distribution is lacking, a flat uniform prior that states there is an equal probability of any parameter value within the parameter's physiological ranges is adopted (Eq. (11) (Gelman et al., 2017). α is the parameter of the copula function, the upper and lower bounds of all kinds of copula parameters are listed in Table S2. $P(\alpha)$ represents the prior distribution of copula parameters (uniform in our case).

$$P(\alpha) = \frac{1}{\alpha_{max}\text{-}\alpha_{min}} \tag{11}$$

Step 2: In general, if a model could produce good estimates for one or several process behaviors, the residual between the actual measurements and their corresponding estimates is expected to follow a Gaussian distribution, or at least be much closer to a Gaussian distribution compared to the original measurements (Tong et al., 2019). Based on this argument, if assuming error residuals (The difference between the copula predicted probability values, $d_i(\alpha)$, and the joint probability of input variables,$\overline{d_i}$) are uncorrelated, Gaussian distributed with a zero mean and a constant variance (homoscedastic), the likelihood function could be defined as:

$$P(\widetilde{D}|\alpha) \cong L\left(\alpha|\widetilde{D}\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\overline{\sigma}^2}} \exp\left\{ -\frac{1}{2}\overline{\sigma}^{-2}[\overline{d_i}\text{-}d_i(\alpha)]^2 \right\} \tag{12}$$

Where $\overline{\sigma}$ is an estimate of the standard deviation of measurement error, which can be dynamically estimated during MCMC simulations. $P(\widetilde{D}|\alpha)$ denotes the likelihood function, and $\widetilde{D}$ denotes the empirical joint probability vector. Appropriate assumptions about residuals enhance the robustness of statistical models used for parameter estimation, increasing their tolerance for outliers. (Nolan et al., 2021). Moreover, through comparing the actual residuals with the assumed distribution, the applicability of these statistical models is validated. A good match between the residuals and the assumed distribution indicates that the models are appropriately set. If discrepancies are found, however, it may be necessary to adjust the models or revise assumptions about the residuals (Linden et al., 2022). Further clarification is warranted here; no assumptions are imposed on the posterior distribution of parameters. The assumptions of "homoscedasticity, no correlation, and Gaussian distribution with mean zero" only apply to the distribution of error residuals which is used to construct the likelihood function that summarizes the distance between observations (empirical bivariate probability values) and model simulations (copula predicted bivariate probability values) into a single scalar.

Step 3: Bayesian analysis is applied to calculate the posterior probabilities of the parameters, combining the prior distribution and the likelihood function (Eq. (13). Due to the Bayesian equation being difficult to solve analytically, a hybrid-evolution MCMC approach (Fig. S3) is developed to sample from the posterior distribution to delineate the posterior parameter region (Sadegh et al., 2017). Before initiating MCMC simulations, intelligent prior sampling is utilized to identify the most likely samples to serve as starting points for the Markov chains (Wolf et al., 2021). Specifically, *LN* (>*N*) samples are extracted from the prior distribution through Latin Hypercube Sampling (LHS). These samples are then randomly distributed among *N* complexes. Within each complex, the sample demonstrating the optimal value is chosen as the starting point for a Markov chain. Here, *LN* represents the number of

samples drawn using LHS, and N denotes the number of Markov chains. The intelligent prior sampling strategically initiates chains from points with the highest likelihood of success, thereby enhancing the overall acceptance rate of the simulations. Then, adaptive metropolis (Haario et al., 2001), differential evolution (Pant et al., 2020), and snooker update (Ter Braak and Vrugt, 2008) algorithms are employed to search the feasible space, thereby increasing sample diversity and improving sampling efficiency (A detailed introduction to the hybrid-evolution MCMC method is in *Supporting Information* (section S1.6)). $P(\widetilde{D})$ is the evidence, and $P\left(\alpha|\widetilde{D}\right)$ denotes the posterior distribution of copula parameters.

$$P\left(\alpha|\widetilde{D}\right) = \frac{P\left(\widetilde{D}|\alpha\right) \times P(\alpha)}{P(\widetilde{D})} \propto P\left(\widetilde{D}|\alpha\right) \times P(\alpha) \tag{13}$$

Step 4: Quantile-Quantile plots (QQ plots) are utilized for the posterior evaluation of residual assumptions. Specifically, the quantiles of the residuals are compared to the quantiles of a theoretical distribution. If the data points in the QQ plot align approximately along the 45-degree line, the distribution of residuals is consistent with the assumed theoretical distribution, suggesting that the assumptions are reasonable. Significant deviations from this line suggest that the distribution of residuals does not match the assumed distribution, adjustments to the statistical model or its residual assumptions are necessary. Several goodness of fit measures are employed to evaluate the performance of copula parameters, including the likelihood value (Eq. (12), RMSE, and NSE. A parameter set that provides the maximum likelihood minimizes the residuals between model simulations and observations. It therefore provides, in this sense, the best fit for the observed data. NSE and RMSE are widely recognized as measures of goodness of fit, focusing on the minimization of residuals. The parameter sets that achieve the highest goodness of fit are selected, with their 95 % uncertainty range determined. These are then subjected to a detailed uncertainty analysis (as detailed in Appendix C) (Li et al., 2013). Estimated copula parameters are applied in the computation of the CDF within a Vine copula structure (this step is detailed in section 3.3.2), thereby constructing the FDC for ungauged basins.

3.4. Copula model evaluation with streamflow segmentation

The performance evaluation of the mixture copula model includes (1) the analysis of its effectiveness across various streamflow segments and (2) the assessment of the applicability of the mixture copula model in different periods, employing four performance metrics, including a four-segment flow duration curve with the RMSE. Their detailed descriptions are listed in Table S1 (Pfannerstill et al., 2014; Yilmaz et al., 2008). RMSE is noted for its ability to sensitively detect both extremely positive and negative errors, making it an essential tool for identifying unsatisfactory model performance (Singh et al., 2012). Both the four-segment flow duration curve and RMSE are applied in analyzing the operational performance of the mixture copula model across various segments (extremely low, low, medium, and high streamflow) of the FDC, RMSE_low is utilized to assess the FDC segment above Q_{70}, reflecting river flow levels under nonextreme but drought conditions. Accurate prediction for this low streamflow is crucial for water quality management, ecosystem maintenance, and the formulation of groundwater recharge strategies. RMSE_mid for the FDC segment between Q_{20} and Q_{70}, characterizing the river's normal flow state, which is crucial for the planning and management of water resources. RMSE_5 and RMSE_20 serve as metrics designated to evaluate the FDC segments below Q_5 and between Q_5 and Q_{20}, respectively, associated with extreme hydrological events. In addition, the differences in these metrics between the calibration period and the validation period are adopted to assess the applicability of the mixture copula model (Broderick et al., 2016; Shen et al., 2022).

4. Results

4.1. Hydrological similarity-based regionalization

The results of the regionalization method in hydrological model parameters transfer are presented in *Supporting Information* (section S2.1), and the performance of the hydrological similarity-based regionalization method for streamflow prediction is detailed in Fig. 4 and Table 2. Table 2 demonstrates a good fit between simulated and observed streamflow, indicated by NSE values around 0.8. In addition, the variability in NSE values between calibration and validation periods is approximately 3 %, suggesting that the performance of the regionalization method is satisfactory. Due to generally diminished sensitivity to short-term meteorological changes, low streamflow exhibits stability and coherence (Hannaford et al., 2013), applying the regionalization approach, allows for the effective extraction and transfer of parameters from basins with hydrological similarity, the closer clustering of data points around the diagonal in low streamflow proves the enhanced reliability of predictions for normal flow events, as shown in Fig. 4. However, the scatters are quite distant from the 1:1 line, indicating limitations in the accuracy of the hydrological modeling forecast. These discrepancies may be due to several factors, including data limitations such as poor quality or insufficient input data, parameter uncertainties, as well as oversimplified model structures that fail to represent complex hydrological processes (Engeland et al., 2016). In addition, the calibration of hydrological models typically assumes that parameters remain constant throughout the calibration period (Seibert et al., 2019; Zhou et al., 2021). In the N05582000 basin in central Illinois' Salt Creek, the Bias value is 98 % higher than the average, and the MAE exceeds the average across all basins by 50 %. Additionally, significant deviations are observed in the predictions of high streamflow. The confluence of multiple tributaries like Addison Creek and Spring Brook, complexity in streamflow patterns may increase prediction uncertainty, on the other hand, the DuPage County Stormwater Management Website (https://ec.dupagecounty.gov/SaltCreek/SaltCreek.aspx) acknowledges the possibility of inaccuracies in gage data, further complicating these predictive efforts. Additionally, the small CI_width and Uncertainty% in Table 2, with minimal variation across different assessment periods, suggest consistent model reliability and precision. In sum, the hydrological similarity-based regionalization method demonstrates effectiveness and robustness in estimating hydrological model parameters for these ungauged basins, particularly in normal and low streamflow. Furthermore, the accurate prediction of streamflow in ungauged basins enables the constructed FDC to reflect the distribution and variability of streamflow comprehensively. Despite initial discrepancies existing, the subsequent application of a process-based mixture copula model attributes the shape of FDCs to physical processes, thereby capturing the complex interdependencies among various streamflow phases, and enhancing the accuracy and reliability of FDCs.

4.2. Construction FDC through a process-based model

4.2.1. Construction of FDC coupling the dependency of delayed flows

Understanding the components of streamflow is essential for elucidating the mechanisms that shape FDCs and their variations across different segments. Delayed flow separation has been utilized to categorize the predicted streamflow into four components, and an analysis of the variability in delayed contributions to streamflow across ungauged basins during both calibration and validation periods is detailed in *Supporting Information* (section S2.2). After identifying the four delayed flows, the Vine copula structure was employed to model their dependence, and detailed results on the construction of the Vine copula are in *Supporting Information* (section S2.3). As shown in Fig. 5, the development of the mixture copula model for constructing FDCs across various basins highlights the significant influence of physical factors on the shape of the FDC. For instance, the N03274000 basin, situated in

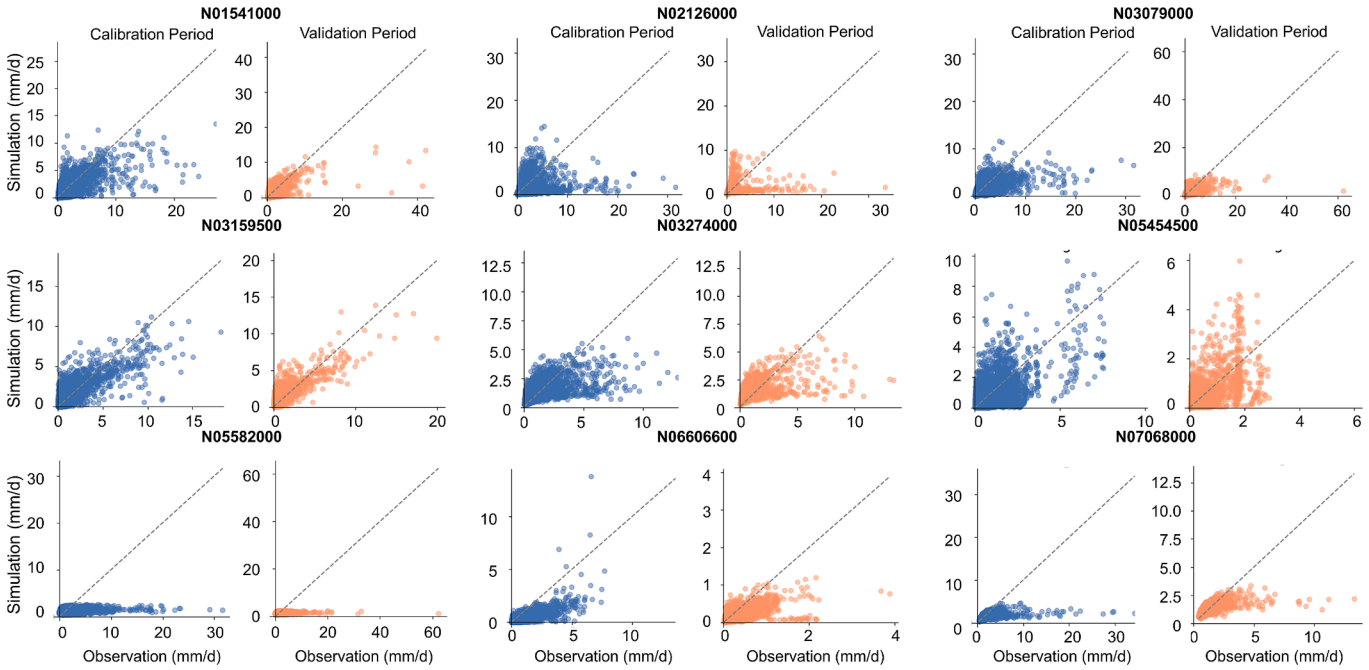

Fig. 4. The comparison between simulations and observations during calibration and validation periods in nine ungauged basins, the black dotted line represents the 1:1 line.

Table 2
The summary of the statistics of hydrological model evaluation indices for different basins during calibration and validation periods, including the NSE, BIAS, MAE, CI_width, and %Uncertainty.

Station	Calibration period (1982–1995)					Validation period (1996–2000)				
	NSE	BIAS	MAE	CI_width	%Uncertainty	NSE	BIAS	MAE	CI_width	%Uncertainty
N01541000	0.653	0.379	0.877	0.086	45.690	0.673	0.443	0.896	0.150	44.975
N05454500	0.789	0.869	0.574	0.054	43.320	0.862	0.156	0.436	0.067	44.615
N06606600	0.664	0.101	0.381	0.025	43.342	0.739	0.763	0.184	0.017	47.018
N07068000	0.800	0.093	0.556	0.028	20.640	0.653	0.117	0.413	0.042	26.924
N03079000	0.644	0.414	0.859	0.067	40.580	0.685	0.528	0.942	0.040	24.820
N05582000	0.989	1.708	0.939	0.019	26.057	0.965	1.694	0.997	0.040	24.820
N02126000	0.987	0.467	0.981	0.069	26.335	0.965	1.206	0.764	0.104	30.718
N03274000	0.783	1.304	0.67	0.040	30.631	0.699	1.496	0.76	0.074	30.718
N03159500	0.670	1.135	0.6	0.075	41.788	0.618	1.27	0.625	0.141	40.128

western Ohio's hilly terrain, features steep slopes that expedite precipitation accumulation and upstream runoff, resulting in a steeper FDC. Seasonal growth variations influence evapotranspiration rates and surface runoff. In the N06606600 basin, surface runoff is reduced in the summer's abundant vegetation, exhibiting a gradual FDC at low streamflow, while reduced vegetation cover in autumn and winter may lead to increased runoff and a steeper FDC in mid to high streamflow (Chouaib et al., 2018; Ye et al., 2012).

Fig. 5 illustrates the results of constructing FDCs in ungauged basins. Preliminary analysis suggests that the FDCs when constructed using a mixture copula model, align more closely with observed FDCs compared to those derived directly from predicted streamflow. This demonstrates that the application of the mixture copula model corrects biases in streamflow prediction generated by the hydrological similarity-based regionalization method. Such effectiveness may be attributed to the model's capability to capture the complex multidimensional dependence of multiple delayed flows within a basin. In the majority of the study cases, FDCs constructed through the mixture copula model, which employs Bayesian-based estimation for copula parameters, demonstrate a closer alignment with the observed FDCs than those derived through inverse Kendall tau transformation-based copula parameter estimation. During the validation period, this model enhances the accuracy of the constructed FDCs more than during the calibration period, highlighting its generalizability and robustness across different periods within these basins. However, Fig. 5a and 5b indicate that in the N03274000 and N05582000 basins, significant discrepancies remain between the constructed and observed FDCs, with the mixture copula model leading to minimal accuracy improvements. The minimal enhancement in precision mainly arises from structural limitations within the hydrological model, which impact the streamflow predictions in these two basins and then, the accuracy of the constructed FDCs. In sum, throughout both the calibration and validation periods within these basins, the mixture copula model has corrected biases in the predicted streamflow, thereby improving the accuracy of the constructed FDC, especially in the middle streamflow (20–70th exceedance probability of FDC).

4.2.2. *Model evaluation considering various streamflow segments*

To evaluate the effectiveness of the mixture copula model in generating FDCs across various streamflow segments, a detailed analysis of Table A1 is undertaken. In most study cases, the FDC directly calculated from the predicted streamflow has RMSE_mid and RMSE_low values around 1 mm/d, and when constructed through the mixture copula model, the RMSE_low and RMSE_mid are both around 0.5 mm/d, detailly, RMSE_mid decreases by approximately 60 %, and RMSE_low decreases by about 90 %. In addition, Bayesian-based copula parameter estimation generally results in lower RMSE_low and RMSE_mid

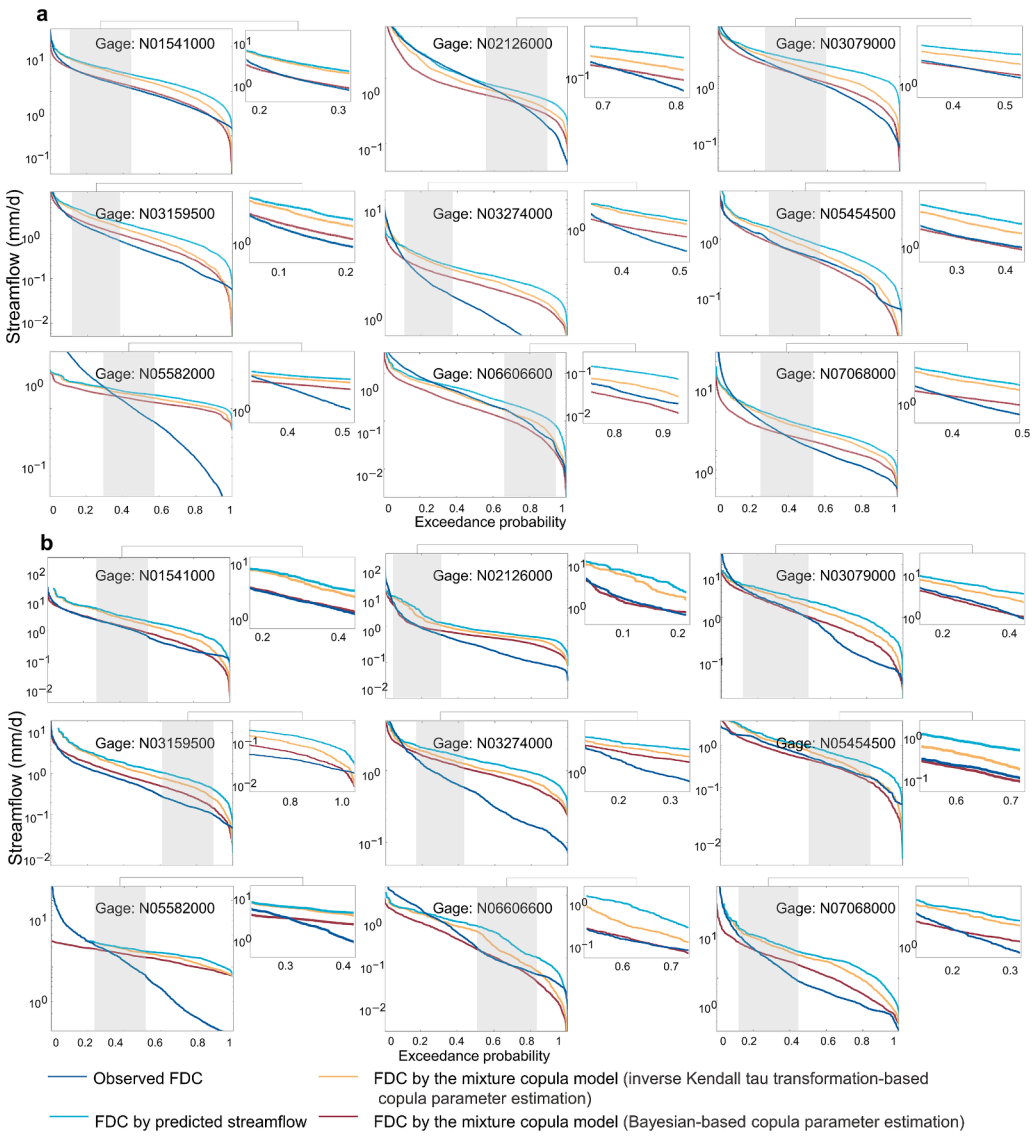

Fig. 5. a, Comparative analysis of FDCs for ungauged basins during the calibration period. It contrasts FDCs derived directly from predicted flows with those constructed using the mixture copula model, employing both inverse Kendall tau transformation-based and Bayesian-based methods for copula parameter estimation, against the observed FDCs. b, Comparison during the validation period.

compared to inverse Kendall tau transformation-based parameter estimation, indicating that Bayesian-based copula parameter estimation achieves superior prediction accuracy in these streamflow segments, highlighting its efficacy in capturing the dynamics of various streamflow patterns. In very high (below 5th exceedance probability of FDC) and high streamflow (5-20th exceedance probability of FDC), the regionalization method based on hydrological similarity needs to accurately simulate and predict rapid streamflow changes within short time scales. It is a challenge for the regionalization method, resulting in larger errors in streamflow prediction and, therefore in FDC construction, the RMSE_5 and RMSE_20, which correspond to RMSE_mid and RMSE_low, show significant increases. However, as shown in Table A1, after applying the mixture copula model, RMSE_5 and RMSE_20 decreased by approximately 70 %, showing that the mixture copula model is able to correct these errors. In most study cases, the mixture copula model exhibits superior accuracy and robustness, capable of maintaining or enhancing its predictive performance across calibration and validation periods. In certain exceptional instances, such as due to the absence of flow data in the high streamflow of the N05582000 basin, RMSE values are missing for both the FDC directly calculated from the predicted streamflow and the FDC constructed through the mixed copula model. In the N03274000 basin, an increase in RMSE values during the validation period, as compared to the calibration period, has been observed. This increase could be attributed to the basin incorporating new hydrological conditions not observed during the calibration period, or to changes within the basin that influence flow patterns (such as changes in land use and construction of water management facilities). In sum, the mixture copula model has exhibited good robustness and stability for constructing FDC during various periods, significantly in low and middle streamflow, with Bayesian-based copula parameter estimation proving effective.

5. Discussion

5.1. Application of the FDC model in the Hanjiang River basin for comparative analysis

Spring snowmelt and seasonal precipitation are identified as the main factors influencing seasonal flow variations in the selected MOPEX basins, with monsoon precipitation playing a minimal role (Ghotbi et al.,

2020b). The application of the mixture copula model to construct FDC in these basins, situated in the Mid-Eastern United States, may not demonstrate the model's adaptability or predictive accuracy across diverse climatic areas. In MOPEX basins, disparities in FDCs constructed from inverse Kendall tau transformation-based and Bayesian-based parameter estimations for copula functions are insignificant, failing to distinguish the effectiveness of these methods. Therefore, the necessity to deploy the model in various basins for research is underscored. The Hanjiang River is the largest tributary of the Yangtze River in China, located in the monsoon region of the East Asia subtropical zone, the primary drivers of streamflow seasonal variation and extreme streamflow events differ across the MOPEX basins. By further application of the mixture copula model to construct the FDC for the Hanjiang River Basin and contrasting these with results from the MOPEX basins, it is possible to evaluate the model's applicability in monsoon climate regions. Therefore, the proposed mixture copula model was developed for three basins with different spatial scales in the Hanjiang River Basin, including the Hanzhong River Basin, Mumahe River Basin, and Xunhe River Basin as shown in Fig. 6a. The application results including the results of the delayed flow separation and a summary of the estimated parameters for the Vine copula structure at each node, are detailed in *Supporting Information* (sections S2.2 and S2.3). Streamflow from 1980

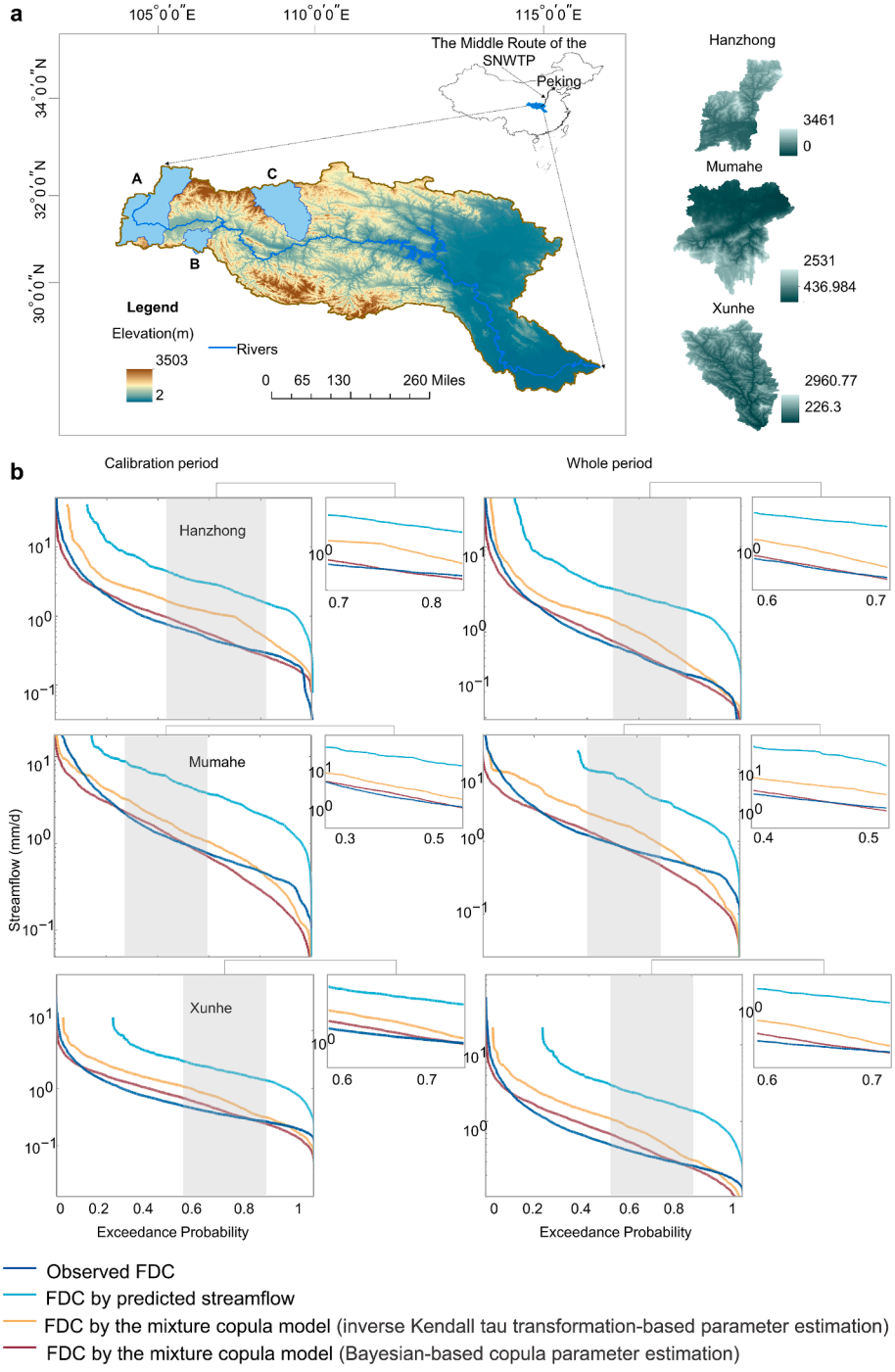

Fig. 6. a, Information on the Hanjiang River Basin. b, Comparative analysis of FDCs for ungauged basins during the calibration and whole periods. It contrasts FDCs derived directly from predicted flows with those constructed using the mixture copula model, employing both inverse Kendall tau transformation-based and Bayesian-based methods for copula parameter estimation, against the observed FDCs.

to 1990 was employed. Nearly 73 % of the data samples (1980–1987) were used for calibration, and the remainder (1988–1990) was utilized to verify the developed model. Model application analysis was conducted for both the calibration and the entire periods separately, due to the short time series during the validation period.

The Hanjiang River Basin experiences distinct wet and dry periods, leading to significant seasonal runoff variations, reflected on the FDCs as differing slopes. For example, from Fig. 6b, during the wet period, the precipitation brought by the monsoon is often intense and of short duration, represented on the FDC as a rapid increase in flow within a brief exceedance probability interval. In the dry period, a reduction in precipitation leads to decreased streamflow, which in turn causes an expansion of the FDC's low streamflow. The comparison of FDCs for three scenarios (FDC calculated directly from predicted streamflow, FDCs constructed through the mixture copula model with inverse Kendall tau transformation-based and Bayesian-based copula parameter estimation) and observed FDC in nine basins is shown in Table A2. Similar to the MOPEX basins, the FDCs derived from the mixture copula model during both calibration and the entire period exhibit a closer approximation to the observed FDCs than those directly calculated from predicted streamflow. As demonstrated in Tables A1 and A2, within the range of low to medium flows, the FDC constructed through the mixture copula model in the Hanjiang River Basin shows a greater improvement in accuracy compared to MOPEX basins. Specifically, the value of RMSE_low in the Hanjiang River Basin decreased by an average of 90 %, and in MOPEX basins, the average reduction in RMSE_low is 65 %. However, at high streamflow levels, RMSE_5 indicates that the mixture copula model does not demonstrate strengths in simulating extremely high streamflow rates, similar to the performance observed in MOPEX basins. Notably, when the validation period is included, the mixture copula model demonstrates significant advantages. In sum, the FDC constructed through the mixture copula model, utilizing Bayesian-based copula parameter estimation, aligns more closely with observed FDCs, particularly within low to middle streamflow, thereby ensuring accurate reflection of actual FDC shapes.

5.2. Comparative analysis of regional methods for FDC prediction in ungauged basins

A novel framework integrating regionalization and process-based methods is proposed to enhance the prediction of FDCs in ungauged basins through deepening the understanding of the physical characteristics governing FDCs. The framework's potential has been explored in MOPEX basins, in general, the framework showed high accuracy, particularly for low to medium streamflow, and demonstrated superior simulation precision during the validation period. However, due to hydrological variations across basins, there are some discrepancies in predictive performance. Traditional regionalization of FDCs is classified into two main categories: (1) those derived from probability distribution methods, which are statistical approaches, and (2) those that dismiss the relationship between FDCs and probability theory (Castellarin et al., 2004), which further divides into parametric approaches that use analytical equations to derive FDCs, and graphical approaches that employ standardized graphical representations of FDCs with regional applicability (Castellarin et al., 2004). Fouad and Loáiciga (2020) examined various variable selection methods to develop percentile flow regression models across a large and representative sample of 918 basins in the United States. They find that the best predictive performance for percentile flows did not align with any single method, suggesting that combining variable selection techniques, whether through ranking variables or developing integrated models, enhances the regression models for percentile flows. However, predictive performance was less accurate for high and low percentile flows. In this study, a parameter transfer method based on hydrological similarity is developed for predicting streamflow in ungauged basins. In addition, process controls of FDCs are explored, including separating predicted streamflow into multiple delay flows to model their interdependencies separately and then integrating them through a Vine copula structure to construct FDCs. In these regards, this approach enhanced the understanding and prediction of basin responses to hydrological events and the role of various hydrological processes in the total streamflow, also biases in FDCs derived directly from predicted streamflow are effectively corrected, maintaining high accuracy in high and low percentile flows.

Moreover, current approaches that enhance the understanding of hydrological processes either utilize stochastic characterization of streamflow time series or process-based models to analyze processes and reconstruct FDCs (Leong and Yokoo, 2021). Botter et al. (2007) developed an analytical probabilistic model based on baseflow probability modeling to characterize streamflow responses. The model focuses on the slow percolation process, which is influenced by the frequency distribution of the percolation component (modeled with a Gamma distribution). The study illustrates that simple but significant parameters related to soil moisture dynamics, decay time constants, and steady-state precipitation events are instrumental in reproducing the baseflow component of FDCs, controlled by the ratio of baseflow frequency to decay time constants. The understanding of the interactions between climate and catchment relationships is significantly advanced through the probabilistic approach. However, this approach, like other similar process-based modeling approaches (Blum et al., 2017), incorporates a simplified assumption of spatial homogeneity within basins (Hermans et al., 2023), which may compromise the model's universality compared to empirical approaches. For instance, Poisson rainfall inputs assume constant parameters applicable to the assigned season, neglecting the carryover of seasonal soil moisture. This restricts the framework's applicability outside the wet season, resulting in limited capability to address seasonal variations, thereby impacting the derivation of annual FDC in highly seasonal basins (Yaeger et al., 2012). With respect to this, the developed framework is applied to MOPEX basins with significant seasonal variation. The framework incorporates a comprehensive set of hydrological indicators that are closely related to seasonal changes in streamflow, enhancing predictions based on hydrological similarity. It demonstrates strong predictive performance, particularly during the validation period, where there was a significant improvement in the accuracy of the predicted FDCs.

5.3. Limitations and future research

There are still some limitations in the current work that need to be further studied and explored in future work. (1) The regionalization method based on hydrological similarity, leaves out groundwater level fluctuations due to data constraints, potentially affecting clustering results. For this reason, integrating descriptors of groundwater level variations is essential for future research. (2) The examination of the mixture copula model concentrated on deterministic prediction, utilizing four delayed flows for enhanced FDC prediction in ungauged basins, beneficial for water resource management. However, model evaluation needs to account for uncertainties in inputs, parameters, and structure. Research into quantifying these uncertainties is a future direction.

6. Conclusions

FDCs are crucial for depicting streamflow behavior within basins. The accuracy and applicability of FDCs, however, are affected by the lack of data in ungauged and poorly gauged basins. Guided by this consideration, this study introduces a novel framework aimed at enhancing predictions of FDCs in ungauged basins by deepening the understanding of the physical characteristics governing FDCs. The framework has been successfully applied to the MOPEX basins, also, to evaluate the applicability of the framework's process-based approach in monsoon climate regions, the approach is further discussed in the Hanjiang River Basin.

- A regionalization method based on hydrological similarity is designed to estimate the hydrological model parameters in ungauged basins, thereby enabling streamflow prediction. The ensemble method is better equipped to handle the complex interactions and nonlinear relationships among basin characteristics. The NSE values between predicted and observed streamflow in the calibration period and validation periods are mostly approximately 0.76, indicating the hydrological similarity-based regionalization method is both effective and robust in estimating parameters for the hydrological model. The constructed FDC thus accurately represents the distribution and variability of streamflow within ungauged basins.
- To improve the accuracy of FDC predictions in ungauged basins and deepen comprehension of the process controls influencing the shape of FDCs, a mixture copula model has been constructed. This model explores process controls of FDCs, including dividing the streamflow into multiple delayed flows to model their interdependence separately and later combining them to model the FDC through the Vine copula structure. Applied to the MOPEX basins, it was revealed that quantifying contributions from short-, intermediate-, long-, and baseline-delay flows through delayed flow separation improves understanding and prediction of basin responses to hydrological events and the roles of various hydrological processes in total flow. Also, biases in the FDCs directly derived from predicted streamflow are effectively corrected by the mixture copula model, thus demonstrating the model's enhanced accuracy of FDCs. Moreover, it has been noted that the performance of the mixture copula model exhibits significant improvement during the validation period, and similar results have been observed in the Hanjiang River Basin.
- To further improve the accuracy of the predicted FDC, Bayesian-based parameter estimation for the construction of Vine-copula structures is developed. Application results from both the MOPEX and Hanjiang River Basins show that the mixture copula model with Bayesian-based copula parameter estimation has exhibited good robustness and stability for constructing FDC, particularly in low and middle streamflow, as the ranges of RMSE_low and RMSE_mid are respectively 0.015–0.78 mm/d and 0.047–1.313 mm/d.

In conclusion, this study has developed an effective framework for predicting FDCs in ungauged basins. The results indicate that the framework offers a reliable framework for predicting FDCs, attributing the shape of the FDCs to physical processes. Future work will concentrate on integrating groundwater impacts for precise future hydrological predictions and expanding the framework's application to a wider range of basins to assess its generalizability and adaptability.

CRediT authorship contribution statement

Tian Lan: Writing – review & editing. **Jiajia Zhang:** Writing – original draft. **Huanhuan Li:** Investigation. **Hongbo Zhang:** Conceptualization. **Xinghui Gong:** Visualization. **Jing Sun:** Methodology. **Yongqin David Chen:** Resources. **Chong-Yu Xu:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study is financially supported by the Science and Technology Plan Project of Shanxi Province Department of water resources (2023slkj-3).

Appendix 1

Appendix A

Table A1
The summary of RMSE values for three scenarios: FDC calculated directly from predicted streamflow, FDC constructed by the mixture copula model (inverse Kendall tau transformation-based and Bayesian-based copula parameter estimation methods) in the calibration and validation periods. RMSE values are categorized based on different streamflow segments (exceedance probability below 5 %, 5–20 %, 20 %-70 %, and greater than 70 %).

		FDC by predicted streamflow		FDC by the mixture copula model (inverse Kendall tau transformation-based copula parameter estimation)		FDC by the mixture copula model (Bayesian-based copula parameter estimation)	
		Calibration	Validation	Calibration	Validation	Calibration	Validation
Station	RMSE (mm/d)						
N01541000	RMSE_5	13.36	8.99	10.89	7.20	5.61	2.84
	RMSE_20	1.57	1.13	0.54	0.13	0.66	0.32
	RMSE_mid	0.26	0.39	0.22	0.38	0.15	0.22
	RMSE_low	1.05	0.92	0.37	0.35	0.16	0.11
N02126000	RMSE_5	4.66	1.48	2.40	1.09	1.96	0.86
	RMSE_20	1.25	0.56	0.70	0.40	0.72	0.40
	RMSE_mid	0.30	0.06	0.10	0.05	0.18	0.09
	RMSE_low	0.73	0.27	0.03	0.02	0.02	0.01
N03079000	RMSE_5	14.58	–	12.45	–	8.20	8.56
	RMSE_20	3.27	3.32	2.60	3.15	2.04	1.66
	RMSE_mid	0.17	0.47	0.41	0.81	0.41	0.57
	RMSE_low	1.24	1.58	0.84	1.00	0.70	0.78
N03159500	RMSE_5	5.79	2.35	1.04	1.69	0.61	0.10
	RMSE_20	1.94	1.30	0.59	0.87	0.16	0.33
	RMSE_mid	0.82	0.59	0.21	0.11	0.09	0.16
	RMSE_low	1.05	1.13	0.04	0.05	0.04	0.03

(*continued on next page*)

Table A1 (*continued*)

		FDC by predicted streamflow		FDC by the mixture copula model (inverse Kendall tau transformation-based copula parameter estimation)			FDC by the mixture copula model (Bayesian-based copula parameter estimation)
N03274000	RMSE_5	6.16	5.73	4.14	2.91	2.34	2.34
	RMSE_20	1.53	1.02	0.27	0.31	0.35	0.37
	RMSE_mid	0.71	0.74	0.67	0.62	0.47	0.51
	RMSE_low	1.50	1.50	0.55	0.52	0.38	0.37
N05454500	RMSE_5	9.46	14.92	2.63	4.78	1.87	1.42
	RMSE_20	1.93	1.30	0.77	2.36	0.26	0.55
	RMSE_mid	1.46	1.29	0.57	0.86	0.25	0.27
	RMSE_low	1.84	2.49	0.18	0.28	0.08	0.09
N05582000	RMSE_5	13.30	13.42	7.66	7.25	5.16	5.89
	RMSE_20	3.16	3.61	0.58	0.57	0.97	0.70
	RMSE_mid	0.49	0.91	0.30	0.41	0.14	0.19
	RMSE_low	2.43	3.07	0.38	0.42	0.11	0.18
N06606600	RMSE_5	13.34	17.02	6.03	9.08	4.45	5.78
	RMSE_20	3.43	1.24	1.11	1.21	1.68	0.18
	RMSE_mid	0.85	1.64	0.37	0.38	0.57	0.23
	RMSE_low	2.10	1.55	0.18	0.31	0.06	0.19
N07068000	RMSE_5	14.89	28.61	7.36	17.26	3.90	5.13
	RMSE_20	2.74	1.81	0.79	1.76	0.31	0.14
	RMSE_mid	0.94	0.87	0.77	0.96	0.12	0.12
	RMSE_low	2.68	2.96	0.34	0.29	0.05	0.08

Table A2
The summary of RMSE values for three scenarios in the Hanjiang River Basin.

		FDC by predicted streamflow		FDC by the mixture copula model (inverse Kendall tau transformation-based copula parameter estimation)			FDC by the mixture copula model (Bayesian-based copula parameter estimation)
		Calibration	Whole	Calibration	Whole	Calibration	Whole
Station	RMSE(mm/d)						
	RMSE_5	–	–	–	–	–	–
Hanzhong	RMSE_20	11.30	–	9.98	–	6.80	4.39
	RMSE_mid	1.15	1.05	1.24	1.86	1.06	1.45
	RMSE_low	2.43	2.41	4.37	3.92	5.82	5.04
Mumahe	RMSE_5	44.14	12.86	27.28	34.33	36.39	19.89
	RMSE_20	3.88	2.00	2.03	1.03	4.79	0.62
	RMSE_mid	0.93	1.31	0.66	0.62	0.80	0.48
	RMSE_low	0.37	0.13	0.057	0.06	0.09	0.15
Xunhe	RMSE_5	21.57	5.71	18.42	18.35	13.29	11.82
	RMSE_20	4.43	0.29	2.56	2.36	1.04	0.99
	RMSE_mid	0.66	0.23	0.13	0.15	0.17	0.13
	RMSE_low	0.21	0.04	0.17	0.19	0.032	0.03

Appendix B

The Ljung-Box test is a statistical test used to determine whether there is significant autocorrelation at lag k in a time series (Uyanto, 2020). The LB test applied to the residuals of an ARMA(p,q) model checks for any remaining autocorrelations in the residuals that the model did not account for. The statistic Q(m) is calculated as follows:

$$Q(m) = T(T+2)\sum_{i=1}^{m}\frac{\widehat{p_i}(a_t^2)}{T\text{-}i} \tag{14}$$

where $\widehat{p_i}$ represents the autocorrelations of the residuals at lag i. The null hypothesis (H0) states that all autocorrelations up to lag m are zero. H0:$\beta_1 = \beta_2 = \ldots = \beta_m = 0$, β_i represents the coefficient of $a_{t\text{-}i}{}^2$:

$$a_t{}^2 = \beta_0 + \beta_1 a_{t\text{-}1}{}^2 + \ldots + \beta_m a_{t\text{-}m}{}^2 + e_t \tag{15}$$

where $t = m+1\ldots$, T, Q(m) is derived from residuals rather than direct observations. Hence, the degrees of freedom for this statistic are adjusted to m-p-q. H1: There exists at least one$\beta_k \neq 0$, for $k \leq m$. In this study, the lag order m is set to 20 to test for autocorrelation up to the 20th lag. Under the null hypothesis, the test statistic Q(m) follows a chi-squared distribution with m degrees of freedom. At a significance level α, the rejection region is defined as $Q > \chi_{1\text{-}\alpha,m}{}^2$ (where the p-value is less than 0.05). Accepting the null hypothesis implies that the series is considered white noise, whereas rejection suggests the presence of autocorrelation. ARCH-LM refers to Engle's LM test for ARCH effects in residuals, which is used to detect the

presence of autoregressive conditional heteroskedastic effects in time series data (Sjölander, 2011). For selected MOPEX basins, the results for the calibration and validation periods in Tables A3 and A4 show the Ljung-Box (LB and LB2) test *p*-values (>0.9) are not significant, proving no serial correlation in the squared residuals. Similarly, non-significant *p*-values in the ARCH-LM test suggest the absence of GARCH effects in model residuals, confirming the appropriateness of copula modeling under assumptions of homoscedasticity and no autocorrelation.

Table A3
Assessment results for four delayed flows in selected MOPEX basins in the calibration period.

Station	Test	Short-dealy	Intermediate-dealy	Long-dealy	Baseline-dealy
N01541000	LB	7.65[0.86]	2.27[1.00]	3.04[0.78]	0.01[1.00]
	LB2	1.85[0.87]	0.08[1.00]	0.55[1.00]	0.07[1.00]
	ARCH-LM	1.36[0.69]	0.06[0.99]	0.04[1.00]	0.05[0.99]
N02126000	LB	2.34[0.76]	2.51[0.81]	7.51[0.98]	2.06[1.00]
	LB2	0.31[0.99]	2.45[0.84	0.08[1.00]	0.01[1.00]
	ARCH-LM	0.24[0.99]	2.12[0.79]	0.05[0.99]	0.01[1.00]
N03079000	LB	5.65[0.81]	6.60 [0.83]	4.41[0.86]	0.32[1.00]
	LB2	1.90[0.92]	0.24[0.99]	1.29[0.93]	0.14[1.00]
	ARCH-LM	1.51[0.82]	0.17[0.99]	2.32[0.94]	0.10[1.00]
N03159500	LB	7.51[0.85]	1.39[0.82]	4.14[0.81]	1.91[1.00]
	LB2	3.42[0.87]	0.52[0.99]	2.12[0.88]	0.03[1.00]
	ARCH-LM	2.55[0.80]	0.36[0.99]	2.31[0.85]	2.32[1.00]
N03274000	LB	3.60[0.91]	3.94[0.88]	3.91[0.84]	3.16[0.86]
	LB2	2.14[0.93]	1.99[0.94]	1.22[1.00]	0.08[1.00]
	ARCH-LM	1.98[0.83]	2.32[0.82]	0.01[1.00]	0.01[1.00]
N05454500	LB	2.55[0.91]	3.39[0.84]	3.48[0.83]	0.03[1.00]
	LB2	0.18[0.99]	0.91[0.98]	2.05[0.91]	0.01[1.00]
	ARCH-LM	0.13[0.99]	0.23[0.99]	2.31[0.92]	0.01[1.00]
N05582000	LB	3.67[1.00]	3.29[0.71]	2.49[0.81]	2.92[0.91]
	LB2	0.02[1.00]	3.21[0.72]	0.04[1.00]	0.05[1.00]
	ARCH-LM	0.02[1.00]	2.11[0.81]	0.02[1.00]	0.01[1.00]
N06606600	LB	2.18[1.00]	2.85[0.80]	0.01[1.00]	2.93[0.82]
	LB2	0.01[1.00]	1.17[0.98]	0.01[1.00]	1.99[0.91]
	ARCH-LM	0.01[1.00]	0.81[0.94]	0.01[1.00]	2.13[0.93]
N07068000	LB	3.51[0.81]	1.25[0.95]	1.36[0.91]	0.34[1.00]
	LB2	0.25[0.99]	1.12[0.96]	1.17[0.95]	0.01[1.00]
	ARCH-LM	0.20[0.99]	1.56[0.93]	1.54[0.94]	0.01[1.00]

Table A4
Assessment results for four delayed flows in selected MOPEX basins in the validation period.

Station	Test	Short-dealy	Intermediate-dealy	Long-dealy	Baseline-dealy
N01541000	LB	2.19[0.85]	1.52[0.94]	1.63[0.92]	0.01[1.00]
	LB2	1.05[0.98]	1.74[0.93]	1.35[0.94]	0.01[1.00]
	ARCH-LM	0.76[0.95]	1.15[0.89]	1.44[0.87]	0.01[1.00]
N02126000	LB	6.36[0.95]	4.12[0.99]	1.04[0.84]	0.39[0.97]
	LB2	0.12[1.00]	1.16[0.98]	1.49[0.96]	0.01[1.00]
	ARCH-LM	0.09[0.99]	0.97[0.92]	0.27[0.99]	0.01[1.00]
N03079000	LB	1.41[0.84]	1.95[0.87]	0.01[1.00]	3.64[0.81]
	LB2	1.60[0.95]	0.93[0.99]	0.01[1.00]	1.58[0.95]
	ARCH-LM	0.56[0.97]	0.41[0.99]	0.01[1.00]	0.01[1.00]
N03159500	LB	2.38[0.87]	4.94[0.90]	4.38[1.00]	0.01[1.00]
	LB2	2.05[0.91]	0.39[0.99]	0.02[1.00]	0.01[1.00]
	ARCH-LM	3.06[0.78]	0.23[0.99]	0.01[1.00]	0.01[1.00]
N03274000	LB	1.44[0.89]	3.22[0.99]	0.34[1.00]	0.01[1.00]
	LB2	2.02[0.90]	0.85[0.99]	0.01[1.00]	0.01[1.00]
	ARCH-LM	1.51[0.82]	0.56[0.97]	0.01[1.00]	0.01[1.00]
N05454500	LB	1.27[0.87]	4.54[0.99]	1.42[0.82]	0.01[1.00]
	LB2	1.13[0.98]	1.22[0.98]	1.60[0.93]	0.01[1.00]
	ARCH-LM	1.00[0.91]	0.85[0.94]	1.35[0.87]	0.01[1.00]
N05582000	LB	2.29[0.90]	1.82[0.93]	0.01[1.00]	0.02[1.00]
	LB2	1.83[0.92]	0.07[1.00]	0.01[1.00]	0.01[1.00]
	ARCH-LM	1.16[0.88]	0.05[0.99]	0.01[1.00]	0.01[1.00]
N06606600	LB	1.86[0.96]	0.60[1.00]	2.87[0.82]	2.19[1.00]
	LB2	0.82[0.99]	0.01[1.00]	1.08[0.91]	0.01[1.00]
	ARCH-LM	0.96[0.92]	0.01[1.00]	1.92[0.83]	0.01[1.00]
N07068000	LB	1.65[0.88]	5.08[0.99]	2.37[0.85]	1.76[0.86]
	LB2	1.06[0.92]	0.86[0.99]	0.02[1.00]	0.01[1.00]
	ARCH-LM	1.11[0.93]	0.62[0.97]	0.02[1.00]	0.01[1.00]

Notes: “LB” and “LB2″ respectively denote the Ljung-Box statistics associated with serial correlation in the residual and squared residual models. The ”ARCH-LM“ refers to Engle’s LM test for detecting ARCH effects within the residuals. P-values [in square brackets] below 0.05 indicate that the null hypothesis of adequate model specification is correctly rejected.

Appendix C:. Analyzing uncertainty in Bayesian parameter estimation results

In Appendix C, an uncertainty analysis of the Bayesian parameter estimation results for the Joe copula function used to calculate the joint CDF between short-delay flow and intermediate-delay flow in the N0154100 basin is conducted, and uncertainty analyses for other basins are presented in the *Supporting Information* (section S2.4). QQ plots (Fig. A1) show that data points generally align along the 45-degree line, indicating that the residual distribution is consistent with the assumed theoretical distribution, thereby validating the assumptions. The cumulative distribution fitted by the Joe copula function and the empirical cumulative distribution are indicated in Fig. A2a and b. As observed, the scatter points are evenly distributed along the 45-degree diagonal, indicating that the Joe copula function performs well in fitting the data. To quantitatively assess the goodness-of-fit between the theoretical joint frequency of the copula function and the empirical joint frequency, the RMSE, AIC, and BIC were employed to further evaluate the fit of the Joe copula. With RMSE = 0.40, AIC = $-$6649.75, and BIC = -6644.73 based on the criterion that smaller RMSE, AIC, and BIC values indicate a better fit, these results confirm the adequacy of the Joe copula in capturing the joint probability distribution. Moreover, the posterior parameter distributions of the Joe copula, as shown in Fig. A2c, are well-constrained. The RMSE of 0.40 for the best parameter value obtained through MCMC simulations demonstrates the Bayesian-based parameter estimation method's ability to search target values reliably, providing a robust solution for dependency analysis issues (Sadegh et al., 2017).

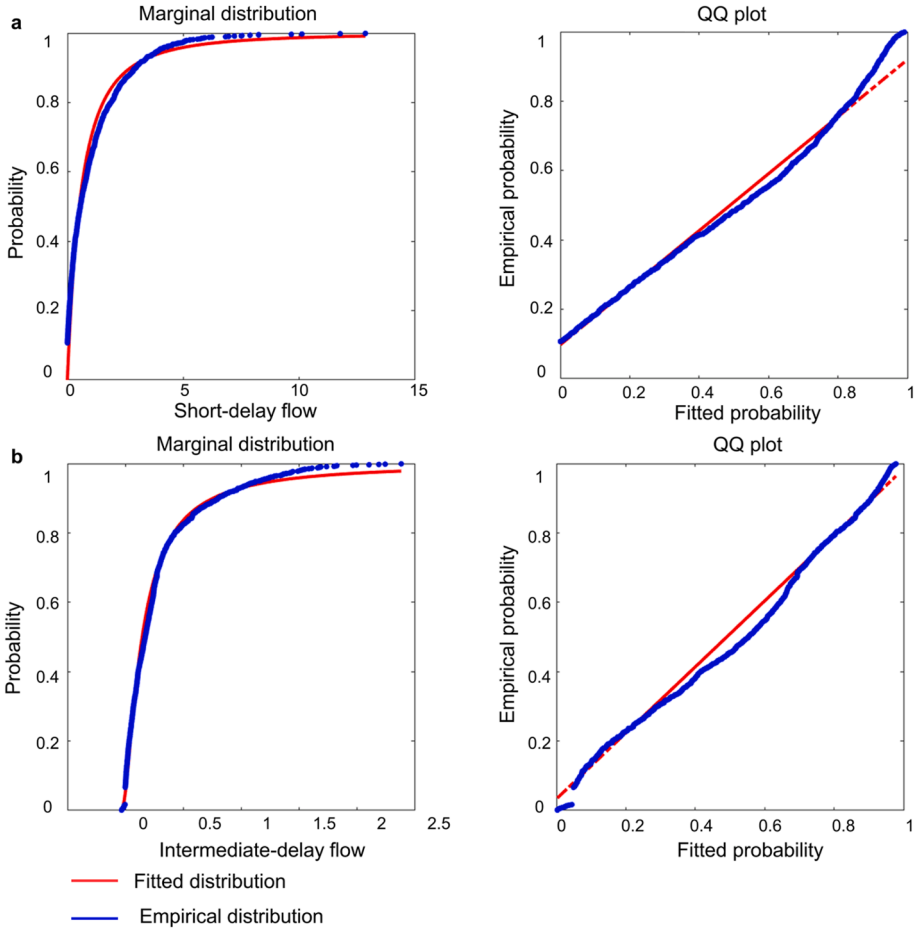

Fig. A1. Marginal distributions of short-delay flow (a) and intermediate-delay flow (b). A selected distribution to fit each data set (red line). Input data are coded with blue dots. QQ plots visually show the goodness-of-fit of distributions to input data.

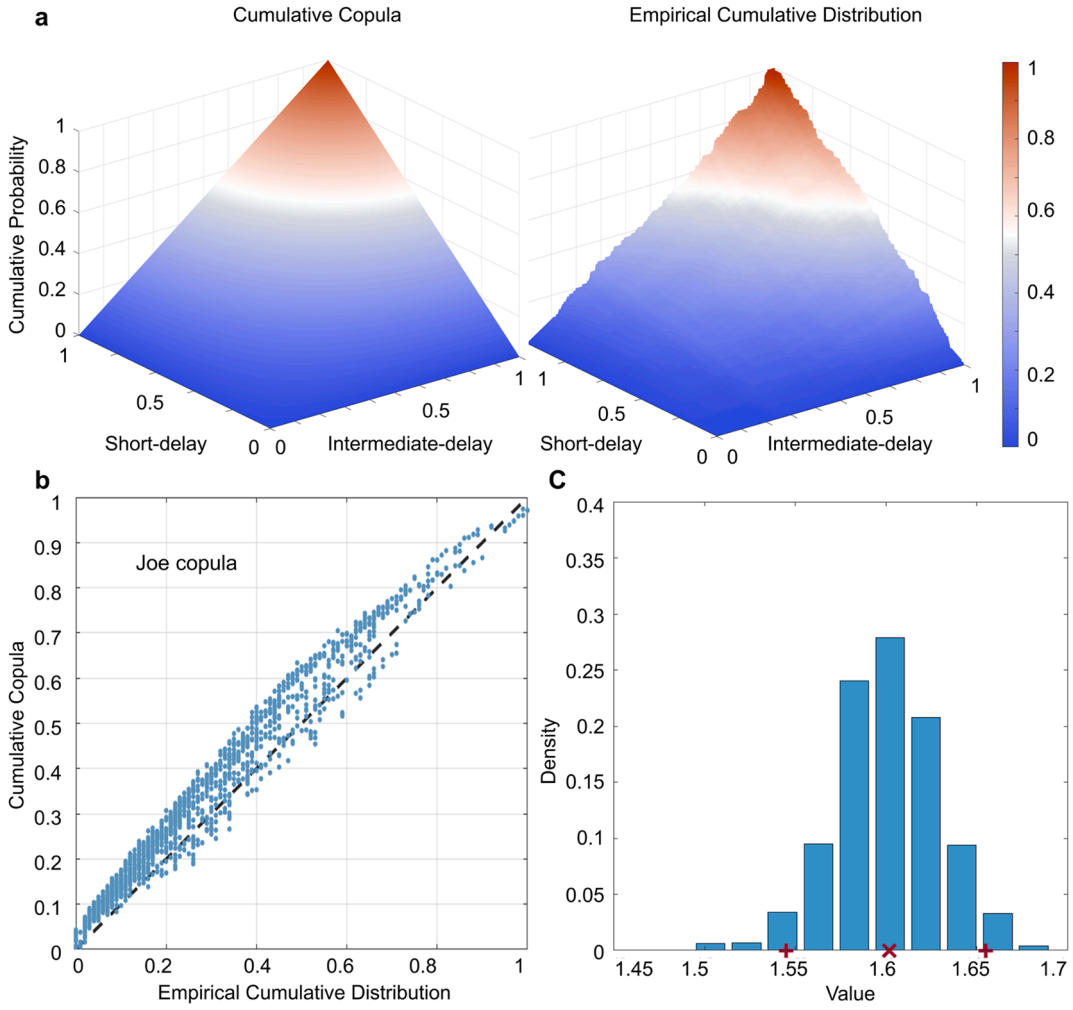

Fig. A2. **a,** Joint cumulative probability distribution based on the Joe copula and the ECDF for short-delay and intermediate-delay flows; **b,** The scatter plot for comparing the theoretical joint probability distribution with the empirical joint probability. **c,** Posterior distribution of the Joe copula derived by the MCMC simulation within a Bayesian framework. The blue bins are the MCMC-derived parameters and the red cross shows the maximum likelihood parameter of the MCMC.

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2024.132364.

Data availability

Data will be made available on request.

References

Aboelnour, M.A., Engel, B.A., Frisbee, M.D., Gitau, M.W., Flanagan, D.C., 2021. Impacts of watershed physical properties and land use on baseflow at regional scales. J. Hydrol.: Reg. Stud. 35, 100810. https://doi.org/10.1016/j.ejrh.2021.100810.

Anderson, B.J., Brunner, M.I., Slater, L.J., Dadson, S.J., 2023. Elasticity curves describe streamflow sensitivity to precipitation across the entire flow distribution. Hydrol. Earth Syst. Sci. Discuss. 2023, 1–27. https://doi.org/10.5194/hess-28-1567-2024.

Archfield, S.A., et al., 2015. Accelerating advances in continental domain hydrologic modeling. Water Resour. Res. 51 (12), 10078–10091. https://doi.org/10.1002/2015WR017498.

Arsenault, R., Breton-Dufour, M., Poulin, A., Dallaire, G., Romero-Lopez, R., 2019. Streamflow prediction in ungauged basins: analysis of regionalization methods in a hydrologically heterogeneous region of Mexico. Hydrol. Sci. J. 64 (11), 1297–1311. https://doi.org/10.1080/02626667.2019.1639716.

Asadollah, S.B.H.S., Sharafati, A., Motta, D., Yaseen, Z.M., 2021. River water quality index prediction and uncertainty analysis: a comparative study of machine learning models. J. Environ. Chem. Eng. 9 (1), 104599. https://doi.org/10.1016/j.jece.2020.104599.

Atieh, M., Taylor, G., Sattar, A.M., Gharabaghi, B., 2017. Prediction of flow duration curves for ungauged basins. J. Hydrol. 545, 383–394. https://doi.org/10.1016/j.jhydrol.2015.08.068.

Bangyal, W.H., Hameed, A., Alosaimi, W., Alyami, H., 2021. A new initialization approach in particle swarm optimization for global optimization problems. Comput. Intell. Neurosci. 2021 (1), 6628889. https://doi.org/10.1155/2021/6628889.

Booker, D.J., Snelder, T.H., 2012. Comparing methods for estimating flow duration curves at ungauged sites. J. Hydrol. 434, 78–94. https://doi.org/10.1016/j.jhydrol.2012.02.031.

Bozchaloei, S.K., Vafakhah, M., 2015. Regional analysis of flow duration curves using adaptive neuro-fuzzy inference system. J. Hydrol. Eng. 20 (12), 6. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001243.

Broderick, C., Matthews, T., Wilby, R.L., Bastola, S., Murphy, C., 2016. Transferability of hydrological models and ensemble averaging methods between contrasting climatic periods. Water Resour. Res. 52 (10), 8343–8373. https://doi.org/10.1002/2016WR018850.

Castellarin, A., 2014. Regional prediction of flow-duration curves using a three-dimensional kriging. J. Hydrol. 513, 179–191. https://doi.org/10.1016/j.jhydrol.2014.03.050.

Castellarin, A., Galeati, G., Brandimarte, L., Montanari, A., Brath, A., 2004. Regional flow-duration curves: reliability for ungauged basins. Adv. Water Resour. 27 (10), 953–965. https://doi.org/10.1016/j.advwatres.2004.08.005.

Cheng, L., et al., 2012. Exploring the physical controls of regional patterns of flow duration curves–Part 1: Insights from statistical analyses. Hydrol. Earth Syst. Sci. 16 (11), 4435–4446. https://doi.org/10.5194/hess-16-4435-2012.

Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput. Sci. 7. https://doi.org/10.7717/peerj-cs.623.e623.
Chouaib, W., Caldwell, P.V., Alila, Y., 2018. Regional variation of flow duration curves in the eastern United States: Process-based analyses of the interaction between climate and landscape properties. J. Hydrol. 559, 327–346. https://doi.org/10.1016/j.jhydrol.2018.01.037.
Cislaghi, A., Masseroni, D., Massari, C., Camici, S., Brocca, L., 2020. Combining a rainfall–runoff model and a regionalization approach for flood and water resource assessment in the western Po Valley. Italy. Hydrological Sciences Jounal 65 (3), 348–370. https://doi.org/10.1080/02626667.2019.1690656.
Coblenz, M., 2021. MATVines: A vine copula package for MATLAB. SoftwareX 14, 100700. https://doi.org/10.1016/j.softx.2021.100700.
Coopersmith, E., Yaeger, M., Ye, S., Cheng, L., Sivapalan, M., 2012. Exploring the physical controls of regional patterns of flow duration curves–Part 3: a catchment classification system based on regime curve indicators. Hydrol. Earth Syst. Sci. 16 (11), 4467–4482. https://doi.org/10.5194/hess-16-4435-2012.
Costa, V., Fernandes, W., 2017. Bayesian estimation of extreme flood quantiles using a rainfall-runoff model and a stochastic daily rainfall generator. J. Hydrol. 554, 137–154. https://doi.org/10.1016/j.jhydrol.2017.09.003.
Costa, V., Fernandes, W., 2021. Regional modeling of long-term and annual flow duration curves: reliability for information transfer with evolutionary polynomial regression. J. Hydrol. Eng. 26 (2), 12. https://doi.org/10.1061/(ASCE)HE.1943-5584.0002051.
Czado, C., Nagler, T., 2022. Vine copula based modeling. Annu. Rev. Stat. Appl. 9, 453–477. https://doi.org/10.1016/j.jfds.2022.11.003.
Daly, E., Calabrese, S., Yin, J., Porporato, A., 2019. Hydrological spaces of long-term catchment water balance. Water Resour. Res. 55 (12), 10747–10764. https://doi.org/10.1029/2019WR025952.
Dehling, H., Vogel, D., Wendler, M., Wied, D., 2017. Testing for changes in Kendall's tau. Economet. Theor. 33 (6), 1352–1386. https://doi.org/10.1017/S026646661600044X.
Dey, P., Mathai, J., Sivapalan, M., Mujumdar, P.P., 2024. On the regional-scale variability in flow duration curves in Peninsular India. Hydrol. Earth Syst. Sci. 28 (7), 1493–1514. https://doi.org/10.5194/hess-28-1493-2024.
Duan, Q., et al., 2006. Model Parameter Estimation Experiment (MOPEX): an overview of science strategy and major results from the second and third workshops. J. Hydrol. 320 (1–2), 3–17. https://doi.org/10.1016/j.jhydrol.2005.07.031.
Duncan, H.P., 2019. Baseflow separation - a practical approach. J. Hydrol. 575, 308–313. https://doi.org/10.1016/j.jhydrol.2019.05.040.
Engeland, K., Steinsland, I., Johansen, S.S., Petersen-Overleir, A., Kolberg, S., 2016. Effects of uncertainties in hydrological modelling. A case study of a mountainous catchment in Southern Norway. J. Hydrol. 536, 147–160. https://doi.org/10.1016/j.jhydrol.2016.02.036
Farmer, W.H., Vogel, R.M., 2013. Performance-weighted methods for estimating monthly streamflow at ungauged sites. J. Hydrol. 477, 240–250. https://doi.org/10.1016/j.jhydrol.2012.11.032.
Feng, D., Liu, J., Lawson, K., Shen, C., 2022. Differentiable, learnable, regionalized process-based models with multiphysical outputs can approach state-of-the-art hydrologic prediction accuracy. Water Resour. Res. 58 (10), e2022WR032404. https://doi.org/10.1029/2022WR032404.
Gaviria, C., Carvajal-Serna, F., 2022. Regionalization of flow duration curves in Colombia. Hydrol. Res. 53 (8), 1075–1089. https://doi.org/10.2166/nh.2022.022.
Gelman, A., Simpson, D., Betancourt, M., 2017. The prior can often only be understood in the context of the likelihood. Entropy 19 (10), 555. https://doi.org/10.3390/e19100555.
Ghotbi, S., Wang, D.B., Singh, A., Bloschl, G., Sivapalan, M., 2020a. A new framework for exploring process controls of flow duration curves. Water Resour. Res. 56 (1), 15. https://doi.org/10.1029/2020WR028041.
Ghotbi, S., Wang, D.B., Singh, A., Mayo, T., Sivapalan, M., 2020b. Climate and landscape controls of regional patterns of flow duration curves across the continental United States: statistical approach. Water Resour. Res. 56 (11), 22. https://doi.org/10.1029/2019WR026083.
Gräler, B., et al., 2013. Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. Hydrol. Earth Syst. Sci. 17 (4), 1281–1296. https://doi.org/10.5194/hess-17-1281-2013.
Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive Metropolis algorithm. Bernoulli 223–242. https://doi.org/10.2307/3318737.
Hannaford, J., Buys, G., Stahl, K., Tallaksen, L., 2013. The influence of decadal-scale variability on trends in long European streamflow records. Hydrol. Earth Syst. Sci. 17 (7), 2717–2733. https://doi.org/10.5194/hess-17-2717-2013.
Hao, Z.C., Singh, V.P., 2016. Review of dependence modeling in hydrology and water resources. Prog. Phys. Geogr. 40 (4), 549–578. https://doi.org/10.1177/0309133316632460.
Hill, S.D., Spall, J.C., 2019. Stationarity and convergence of the metropolis-hastings algorithm: Insights into theoretical aspects. IEEE Control Syst. Mag. 39 (1), 56–67. https://doi.org/10.1109/MCS.2018.2876959.
Huang, Q., et al., 2020. Using remote sensing data-based hydrological model calibrations for predicting runoff in ungauged or poorly gauged catchments. Water Resour. Res. 56 (8), e2020WR028205. https://doi.org/10.1029/2020WR028205.
Jafarzadegan, K., Merwade, V., Moradkhani, H., 2020. Combining clustering and classification for the regionalization of environmental model parameters: Application to floodplain mapping in data-scarce regions. Environ. Model. Softw. 125, 104613. https://doi.org/10.1016/j.envsoft.2019.104613.
Janssen, J., Ameli, A.A., 2021. A hydrologic functional approach for improving large-sample hydrology performance in poorly gauged regions. Water Resour. Res. 57 (9), e2021WR030263. https://doi.org/10.1029/2021WR030263.
Johannessen, B.G., Hamouz, V., Gragne, A.S., Muthanna, T.M., 2019. The transferability of SWMM model parameters between green roofs with similar build-up. J. Hydrol. 569, 816–828. https://doi.org/10.1016/j.jhydrol.2019.01.004.
Kaplan, D., 2021. On the quantification of model uncertainty: A Bayesian perspective. Psychometrika 86 (1), 215–238. https://doi.org/10.1007/s11336-021-09754-5.
Kodinariya, T.M., Makwana, P.R., 2013. Review on determining number of Cluster in K-Means Clustering. Int. J. 1 (6), 90–95. https://www.researchgate.net/publication/313554124.
Krapu, C., Borsuk, M., 2022. A differentiable hydrology approach for modeling with time-varying prameters. Water Resour. Res. 58 (9), e2021WR031377. https://doi.org/10.1029/2021WR031377.
Kratzert, F., et al., 2019. Toward improved predictions in ungauged basins: exploiting the power of machine learning. Water Resour. Res. 55 (12), 11344–11354. https://doi.org/10.1029/2019WR026065.
Kwon, H.H., Lall, U., 2016. A copula-based nonstationary frequency analysis for the 2012–2015 drought in California. Water Resour. Res. 52 (7), 5662–5675. https://doi.org/10.1002/2016WR018959.
Leong, C., Yokoo, Y., 2019. Estimating flow duration curves in perennial and ephemeral catchments using a disaggregated approach. Hydrol. Res. Lett. 13 (2), 14–19. https://doi.org/10.3178/hrl.13.14.
Leong, C., Yokoo, Y., 2021. A step toward global-scale applicability and transferability of flow duration curve studies: A flow duration curve review (2000-2020). J. Hydrol. 603, 16. https://doi.org/10.1016/j.jhydrol.2021.126984.
Li, L., Xu, C.Y., Engeland, K., 2013. Development and comparison in uncertainty assessment based Bayesian modularization method in hydrological modeling. J. Hydrol. 486, 384–394. https://doi.org/10.1080/10705511.2017.1363652.
Lin, L.-C., Huang, P.-H., Weng, L.-J., 2017. Selecting path models in SEM: a comparison of model selection criteria. Struct. Equ. Model. Multidiscip. J. 24 (6), 855–869. https://doi.org/10.5555/3495724.3495854.
Linden, N.J., Kramer, B., Rangamani, P., 2022. Bayesian parameter estimation for dynamical models in systems biology. PLoS Comput. Biol. 18 (10), e1010651.
Ling, S., McAleer, M., 2003. Asymptotic theory for a vector ARMA-GARCH model. Economet. Theor. 19 (2), 280–310. https://doi.org/10.1017/S0266466603192092.
Mazdiyasni, B.S., Matthew, R., AghaKouchak, A., 2018. Multi-hazard scenarios for analysis of compound extreme events. Geophys. Res. Lett. https://doi.org/10.1029/2018GL077317.
Mejdoub, H., Ghorbel, A., 2018. Conditional dependence between oil price and stock prices of renewable energy: a vine copula approach. Economic and Political Studies 6 (2), 176–193. https://doi.org/10.1080/20954816.2018.1463600.
Miller, M.P., Johnson, H.M., Susong, D.D., Wolock, D.M., 2015. A new approach for continuous estimation of baseflow using discrete water quality data: method description and comparison with baseflow estimates from two existing approaches. J. Hydrol. 522, 203–210. https://doi.org/10.1016/j.jhydrol.2014.12.039.
Müller, M.F., Thompson, S.E., 2016. Comparing statistical and process-based flow duration curve models in ungauged basins and changing rain regimes. Hydrol. Earth Syst. Sci. 20 (2), 669–683. https://doi.org/10.5194/hess-20-669-2016.
Nazeri Tahroudi, M., Ramezani, Y., De Michele, C., Mirabbasi, R., 2021. Flood routing via a copula-based approach. Hydrol. Res. 52 (6), 1294–1308. https://doi.org/10.2166/nh.2021.008.
Niemeyer, R.J., Bladon, K.D., Woodsmith, R.D., 2020. Long-term hydrologic recovery after wildfire and post-fire forest management in the interior Pacific Northwest. Hydrol. Process. 34 (5), 1182–1197. https://doi.org/10.1002/hyp.13665.
Nolan, S., Smerzi, A., Pezzè, L., 2021. A machine learning approach to Bayesian parameter estimation. Nature Partner Journals Quantum Information 7 (1), 169. https://doi.org/10.1038/s41534-021-00497-w.
Owolabi, S.T., Madi, K., Kalumba, A.M., Alemaw, B.F., 2020. Assessment of recession flow variability and the surficial lithology impact: a case study of Buffalo River catchment, Eastern Cape, South Africa. Environmental Earth Sciences, 79: 1-19, https://doi.org/ 10.1007/s12665-020-08925-4.
Pant, M., Zaheer, H., Garcia-Hernandez, L., Abraham, A., 2020. Differential Evolution: A review of more than two decades of research. Eng. Appl. Artif. Intel. 90, 103479. https://doi.org/10.1016/j.engappai.2020.103479.
Parent, E., Favre, A.-C., Bernier, J., Perreault, L., 2014. Copula models for frequency analysis what can be learned from a Bayesian perspective? Adv. Water Resour. 63, 91–103. https://doi.org/10.1016/j.advwatres.2013.10.013.
Persiano, S., et al., 2022. Streamflow data availability in Europe: a detailed dataset of interpolated flow-duration curves. Earth Syst. Sci. Data 14 (9), 4435–4443. https://doi.org/10.5194/essd-14-4435-2022.
Pfannerstill, M., Guse, B., Fohrer, N., 2014. Smart low flow signature metrics for an improved overall performance evaluation of hydrological models. J. Hydrol. 510, 447–458. https://doi.org/10.1016/j.jhydrol.2013.12.044.
Poovathingal, S.K., Gunawan, R., 2010. Global parameter estimation methods for stochastic biochemical systems. BMC Bioinf. 11, 1–12. https://doi.org/10.1186/1471-2105-11-414.
Pouliasis, G., Torres-Alves, G.A., Morales-Napoles, O., 2021. Stochastic modeling of hydroclimatic processes using Vine copulas. Water 13 (16), 19. https://doi.org/10.3390/w13162156.
Qamar, M.U., et al., 2016. Model swapping: a comparative performance signature for the prediction of flow duration curves in ungauged basins. J. Hydrol. 541, 1030–1041. https://doi.org/10.1016/j.jhydrol.2016.08.012.
Quan, Z., Teng, J., Sun, W., Cheng, T., Zhang, J., 2015. Evaluation of the HYMOD model for rainfall–runoff simulation using the GLUE method. Proc. Int. Assoc. Hydrol. Sci. 368, 180–185. https://doi.org/10.1007/s11269-015-0939-5.

Refsgaard, J.C., Stisen, S., Koch, J., 2022. Hydrological process knowledge in catchment modelling–Lessons and perspectives from 60 years development. Hydrol. Process. 36 (1), e14463.
Reimann, J., et al., 2018. Subcycle observation of lightwave-driven Dirac currents in a topological surface band. Nature 562 (7727), 396–400. https://doi.org/10.1038/s41586-018-0544-x.
Ribatet, M., Sedki, M., 2013. Extreme value copulas and max-stable processes. Journal De La Société Française De Statistique 154 (1), 138–150. https://sites.stat.washington.edu/peter/PASI/Docs/paper1.pdf.
Ridolfi, E., Kumar, H., Bardossy, A., 2020. A methodology to estimate flow duration curves at partially ungauged basins. Hydrol. Earth Syst. Sci. 24 (4), 2043–2060. https://doi.org/10.5194/hess-24-2043-2020.
Sadegh, M., Ragno, E., AghaKouchak, A., 2017. Multivariate Copula Analysis Toolbox (MvCAT): describing dependence and underlying uncertainty using a B ayesian framework. Water Resour. Res. 53 (6), 5166–5183. https://doi.org/10.1002/2016WR020242.
Sauquet, E., Catalogne, C., 2011. Comparison of catchment grouping methods for flow duration curve estimation at ungauged sites in France. Hydrol. Earth Syst. Sci. 15 (8), 2421–2435. https://doi.org/10.5194/hess-15-2421-2011.
Seibert, J., Staudinger, M., van Meerveld, H., 2019. Validation and over-parameterization—experiences from hydrological modeling. Computer Simulation Validation. Simulation Foundations, Methods and Applications 811–834. https://doi.org/10.1007/978-3-319-70766-2_33.
Shen, H., Tolson, B.A., Mai, J., 2022. Time to update the split-sample approach in hydrological model calibration. Water Resour. Res. 58 (3), e2021WR031523. https://doi.org/10.1029/2021WR031523.
Singh, N.M., Devi, T.T., 2022. Regionalization methods in ungauged catchments for flow prediction: review and its recent developments. Arab. J. Geosci. 15 (11), 1019. https://doi.org/10.1007/s12517-022-10287-z.
Singh, K.P., Gupta, S., Kumar, A., Shukla, S.P., 2012. Linear and nonlinear modeling approaches for urban air quality prediction. Sci. Total Environ. 426, 244–255. https://doi.org/10.1016/j.scitotenv.2012.03.076.
Sklar, A., 1973. Random variables, joint distribution functions, and copulas. Kybernetika 9 (6), (449)-460. https://dml.cz/handle/10338.dmlcz/125838.
Smith, M.S., Gan, Q., Kohn, R.J., 2012. Modelling dependence using skew t copulas: Bayesian inference and applications. J. Appl. Economet. 27 (3), 500–522. https://doi.org/10.1002/jae.1215.
Stoelzle, M., Schuetz, T., Weiler, M., Stahl, K., Tallaksen, L.M., 2020. Beyond binary baseflow separation: a delayed-flow index for multiple streamflow contributions. Hydrol. Earth Syst. Sci. 24 (2), 849–867. https://doi.org/10.5194/hess-24-849-2020.
Suwal, N., et al., 2020. Optimisation of cascade reservoir operation considering environmental flows for different environmental management classes. Renew. Energy 158, 453–464. https://doi.org/10.1016/j.renene.2020.05.161.
Tahroudi, M.N., Ramezani, Y., De Michele, C., Mirabbasi, R., 2020. Analyzing the conditional behavior of rainfall deficiency and groundwater level deficiency signatures by using copula functions. Hydrol. Res. 51 (6), 1332–1348. https://doi.org/10.2166/nh.2020.036.
Ter Braak, C.J., Vrugt, J.A., 2008. Differential evolution Markov chain with snooker updater and fewer chains. Stat. Comput. 18, 435–446. https://doi.org/10.1007/s11222-008-9104-9.
Tounsi, A., Abdelkader, M., Temimi, M., 2023. Assessing the simulation of streamflow with the LSTM model across the continental United States using the MOPEX dataset. Neural Comput. & Applic. 1–18. https://doi.org/10.1007/s00521-023-08922-1.
Vafakhah, M., Bozchaloei, S.K., 2020. Regional analysis of flow duration curves through support vector regression. Water Resour. Manag. 34 (1), 283–294. https://doi.org/10.1007/s11269-019-02445-y.
Van de Schoot, R., et al., 2021. Bayesian statistics and modelling. Nat. Rev. Methods Primers 1 (1), 1. https://doi.org/10.1038/s43586-020-00001-2.
Verma, R.K., Murthy, S., Verma, S., Mishra, S.K., 2017. Design flow duration curves for environmental flows estimation in Damodar River Basin, India. Appl Water Sci 7, 1283–1293. https://link.springer.com/article/10.1007/s13201-016-0486-0.
Vernieuwe, H., Vandenberghe, S., De Baets, B., Verhoest, N.E.C., 2015. A continuous rainfall model based on vine copulas. Hydrol. Earth Syst. Sci. 19 (6), 2685–2699. https://doi.org/10.5194/hess-19-2685-2015.
Vogel, R.M., Fennessey, N.M., 1995. Flow duration curves II: a review of applications in water resources planning 1. J. Am. Water Resour. Assoc. 31 (6), 1029–1039. https://doi.org/10.1111/j.1752-1688.1995.tb03419.x.
Vorogushyn, S., et al., 2018. Evolutionary leap in large-scale flood risk assessment needed. Wiley Interdisciplinary Reviews-Water 5 (2), 7. https://doi.org/10.1002/wat2.1266.
Wang, W., et al., 2023. Research on parameter regionalization of distributed hydrological model based on machine learning. Water 15 (3), 518. https://doi.org/10.3390/w15030518.
Wang, C., Gomez-Velez, J.D., Wilson, J.L., 2022. Dynamic coevolution of baseflow and multiscale groundwater flow system during prolonged droughts. J. Hydrol. 609, 127657. https://doi.org/10.1016/j.jhydrol.2022.127657.
Waseem, M., Shin, J.Y., Kim, T.W., 2015. Comparing spatial interpolation schemes for constructing a flow duration curve in an ungauged basin. Water Resource Manage 29 (7), 2249–2265. https://doi.org/10.1007/s11269-015-0939-5.
Wolff, W., Duarte, S.N., 2021. Toward geostatistical unbiased predictions of flow duration curves at ungauged basins. Adv. Water Resour. 152, 13. https://doi.org/10.1016/j.advwatres.2021.103915.
Wright, K.A., Sendek, K.H., Rice, R.M., Thomas, R.B., 1990. Logging effects on streamflow: storm runoff at Caspar Creek in northwestern California. Water Resour. Res. 26 (7), 1657–1667. https://doi.org/10.1029/WR026i007p01657.
Wu, H.J., Su, X.L., Singh, V.P., Feng, K., Niu, J.P., 2021. Agricultural drought prediction based on conditional distributions of Vine copulas. Water Resour. Res. 57 (8), 23. https://doi.org/10.1029/2021WR029562.
Xin, Y., Lu, N., Jiang, H., Liu, Y., Yao, L., 2021. Performance of ERA5 reanalysis precipitation products in the Guangdong-Hong Kong-Macao greater Bay Area. China. Journal of Hydrology 602, 126791. https://doi.org/10.1016/j.jhydrol.2021.126791.
Xu, T., Liang, F., 2021. Machine learning for hydrologic sciences: An introductory overview. Wiley Interdiscip. Rev. Water 8 (5), e1533.
Yaeger, M., et al., 2012. Exploring the physical controls of regional patterns of flow duration curves–Part 4: a synthesis of empirical analysis, process modeling and catchment classification. Hydrol. Earth Syst. Sci. 16 (11), 4483–4498. https://doi.org/10.5194/hess-16-4435-2012.
Ye, S., Yaeger, M., Coopersmith, E., Cheng, L., Sivapalan, M., 2012. Exploring the physical controls of regional patterns of flow duration curves–Part 2: role of seasonality, the regime curve, and associated process controls. Hydrol. Earth Syst. Sci. 16 (11), 4447–4465. https://doi.org/10.5194/hess-16-4435-2012.
Yeh, H.-F., Chen, H.-Y., 2022. Assessing the long-term hydrologic responses of river catchments in Taiwan using a multiple-component hydrograph approach. J. Hydrol. 610, 127916. https://doi.org/10.1016/j.jhydrol.2022.127916.
Yi, X., Zhou, F., Wang, X., Yang, Y., Guo, H., 2014. Classification and runoff simulation of data-scarce basins based on self-organizing maps. Prog. Geogr. 33, 1109–1116. https://www.progressingeography.com/CN/10.11820/dlkxjz.2014.08.011.
Yilmaz, K.K., Gupta, H.V., Wagener, T., 2008. A process-based diagnostic approach to model evaluation: application to the NWS distributed hydrologic model. Water Resour. Res. 44 (9). https://doi.org/10.1029/2007WR006716.
Yu, Q., Jiang, L., Wang, Y., Liu, J., 2023. Enhancing streamflow simulation using hybridized machine learning models in a semi-arid basin of the Chinese loess Plateau. J. Hydrol. 617, 129115. https://doi.org/10.1016/j.jhydrol.2023.129115.
Zhang, K., Luhar, M., Brunner, M.I., Parolari, A.J., 2023. Streamflow prediction in poorly gauged watersheds in the United States through data-driven sparse sensing. Water Resour. Res. 59 (4), e2022WR034092. https://doi.org/10.1029/2022WR034092.
Zhang, Y., Vaze, J., Chiew, F.H., Li, M., 2015. Comparing flow duration curve and rainfall–runoff modelling for predicting daily runoff in ungauged catchments. J. Hydrol. 525, 72–86. https://doi.org/10.1016/j.jhydrol.2015.03.043.
Zhou, S., Wang, Y., Li, Z., Chang, J., Guo, A., 2021. Quantifying the uncertainty interaction between the model input and structure on hydrological processes. Water Resour. Manag. 35, 3915–3935. https://doi.org/10.1007/s11269-021-02954-9.
Zounemat-Kermani, M., Batelaan, O., Fadaee, M., Hinkelmann, R., 2021. Ensemble machine learning paradigms in hydrology: a review. J. Hydrol. 598, 13. https://doi.org/10.1016/j.jhydrol.2021.126266.