The construction of the flow duration curve and the regionalization parameters analysis in the northwest of China

Jinkai Luan, Dengfeng Liu, Mu Lin and Qiang Huang

ABSTRACT

Daily runoff is the data to estimate the water resources in a river. In many catchments, the daily discharge is not well observed. Flow duration curve is an important characteristic of daily runoff, and important for the design of water conservancy projects. In the ungauged catchments, the evaluation of distribution functions and the parameters of flow duration curve is a helpful method to understand the characteristics of the flow. This study uses data from 19 hydrological stations to evaluate the applicability of 11 distribution functions to simulate flow duration curves in the northwest of China. The fitted flow duration curves are evaluated by Nash-Sutcliffe efficiency, the root mean square relative error and the coefficient of determination. The evaluation shows that, among the 11 distribution functions, the log normal model is the most suitable model to construct flow duration curves of 19 hydrological stations. Based on a multivariate linear regression model, a regional model of distribution parameters is constructed, including functions of watershed geomorphologic and climatic characteristics. The analysis of Baijiachuan hydrological station shows that the parameters *a* and *b* showed a decreasing trend. This study presents an innovative approach to evaluate regionalized parameters of flow duration curves considering the impacts of geomorphologic and climatic characteristics.

Key words | climate, distribution function, flow duration curve, geomorphology, log normal model

HIGHLIGHTS

- Daily discharge at 19 hydrological stations in the northwest of China are used to evaluate the applicability of 11 distribution functions to simulate flow duration curves.
- The log normal distribution function is the most suitable model to construct flow duration curves.
- A regional model of distribution parameters is constructed, and they are functions of watershed geomorphologic and climatic characteristics.

Jinkai Luan
Dengfeng Liu (corresponding author)
Qiang Huang
State Key Laboratory of Eco-hydraulics in Northwest Arid Region,
Xi'an University of Technology,
Xi'an 710048,
China
E-mail: *liudf@xaut.edu.cn*

Jinkai Luan
Key Laboratory of Water Cycle and Related Land Surface Processes,
Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences,
Beijing 100101,
China
and
University of Chinese Academy of Sciences,
Beijing 100049,
China

Mu Lin
School of Statistics and Mathematics,
Central University of Finance and Economics,
Beijing 100081,
China

---

doi: 10.2166/wcc.2021.324

INTRODUCTION

Water resources are one of the most important natural resources. Daily runoff series is the basic data to estimate the water resources in a river basin. However, in many catchments, the daily discharge is not well observed. Therefore, predicting daily runoff time series in ungauged catchments is both important and challenging. The flow duration curve is a characteristic curve that illustrates the relationship between frequency and magnitude of streamflow (Ganora *et al.* 2009; Cheng *et al.* 2012; Mendicino & Senatore 2013). Using the flow duration curve, it is possible to estimate the percentage of time that a specified streamflow is equaled or exceeded. The flow duration curve is an important method in hydrological calculations for the estimation of water resources and engineering designs, such as hydropower, shipping and others, because the designs of these projects depend not only on the timing of the flow, but also on the duration of the flow (Liang *et al.* 2019). There are two commonly used methods to construct flow duration curves: the total duration method and the multi-year average method (Liang *et al.* 2019). The total duration method more closely represents the change of the runoff within a given period of time (Cheng *et al.* 2012), while the multi-year average method calculates virtual curves that demonstrate typical rainfall runoffs in a certain basin based on the long-term average (Liang *et al.* 2019).

In many areas of the world there is not enough long-term observed runoff data. These areas are referred to as ungauged areas (Hrachowitz *et al.* 2013). Accurate runoff prediction in these areas is almost impossible due to the lack of data (Atieh *et al.* 2015) because hydrological models need long-term observed runoff to calibrate (Hrachowitz *et al.* 2013). The flow duration curves of gauged sub-regions can be obtained based on measured runoff data, and are used to obtain the regional flow duration curve-based parameter regionalization, which can be transformed to ungauged sub-regions to meet the needs of water resources engineering design in those areas. There are usually two ways to establish a flow duration curve in an ungauged area: a process-based modeling method and a statistical method (Blum *et al.* 2017). Although the process-based hydrological models are applied in areas with sufficient streamflow data to investigate the physical features of the basin (Liu *et al.* 2019), the application of process-based models in ungauged areas is limited due to the lack of necessary data and the uncertainty of basic runoff and climate mechanisms (Yokoo & Sivapalan 2011; Schaefli *et al.* 2013; Basso *et al.* 2015; Mueller & Thompson 2016; Reichl & Hack 2017). On the other hand, the statistical model is simple and relatively easy to implement. Many researchers have used statistical models for the development of flow duration curves in ungauged areas. Based on the assumption of logarithmic normal distribution, Li *et al.* (2010) estimated the distribution parameters of flow duration curves in southeastern Australia, and an exponential model was proposed to fit the geomorphologic and climatic features of the sub basin. Viola *et al.* (2011) distinguish dry periods and wet periods according to flows, using a three parameters power law to describe the frequency distribution of flows in wet periods in Sicily, Italy. Cheng *et al.* (2012) found that the three-parameter mixed gamma distribution can fully capture the shape of the flow duration curve and its changes between catchments and years. The possible relationships between the fitted gamma distribution parameters and the climate and physical characteristics of the basin are explored to explain and point out potential physical controls. Zhang *et al.* (2015) compared the application of the rainfall-runoff model and the flow duration curve method to predict the daily runoff in ungauged catchments of southeastern Australia, and results indicate that both methods can be further improved to simulate daily hydrographs describing the range of flow metrics in ungauged catchments. Blum *et al.* (2017) researched nearly 400 unregulated, perennial streams across the United States and found that the four-parameter kappa distribution provides a very good representation of daily streamflow across the majority of physiographic regions. Further, for some regions of the US, the three-parameter generalized Pareto and log normal distributions also provide a good approximation to flow duration curves. Mendicino & Senatore (2013) used five-parameter models and two statistical models to develop flow duration curves of 19 sub-basins in southern Italy (Calabria). Regression analysis of the model parameters as functions of the geomorphic and climatic features of the sub basin demonstrated that the statistical model was more reliable. Longobardi & Villani (2013) introduced the base flow index and estimated the distribution parameters of the flow duration curve in the Mediterranean region with the consideration of the influence of basin lithology and climate. Booker & Snelder (2012) studied the basins of the entire New Zealand. Polynomial and probability distribution functions were used to fit the flow duration curves. Stepwise regression analysis and random forest analysis were used to calculate the distribution parameters based on the characteristics of the sub-basins. Atieh *et al.* (2015) introduced an entropy-based neural network model assuming a logarithm normal distribution for the flow duration curve and predicted the distribution parameters of the flow duration curve based on the geomorphic and climatic characteristics of the basin. It was found that, compared to ordinary neural network models, the accuracy of the location and scale parameters were increased by 7 and 21%, respectively.

The above studies have provided a good theoretical and practical basis for flow duration curve research, but most of the research on flow duration curve is currently located in several countries (Fennessey & Vogel 1990; Castellarin *et al.* 2004; Booker & Snelder 2012; Boscarello *et al.* 2016; Cislaghi *et al.* 2020), such as the United States, Australia and Italy. Due to China's vast area and complex topography, there are also many ungauged catchments. As a result, water conservancy projects in these areas have no runoff data as a reference. These methods provided a new guideline for developing small hydropower in rural areas without runoff data. Northern Shaanxi in China is the main sediment source region of the Yellow River and about 20,000 check dams have been built in this region to reduce the sediment load (Fu *et al.* 2011; Wang *et al.* 2015; Fu *et al.* 2017; Li *et al.* 2017). New check dams and other projects will be built in the future. Because the catchment areas of the check dams are very small and usually are several km^2, the virgin discharge of these small catchments were not usually observed in the past.

The motivation of this study is to select an optimal distribution function to describe the flow duration curve in the study area and construct the regional parameter model to calculate the parameter of flow duration curve from geomorphologic and climatic characteristics, and the spatial and temporal characteristics of the parameters will be revealed. The method in this study will provide guidance for construction and planning of soil and water conservancy projects in areas without measured flow data.

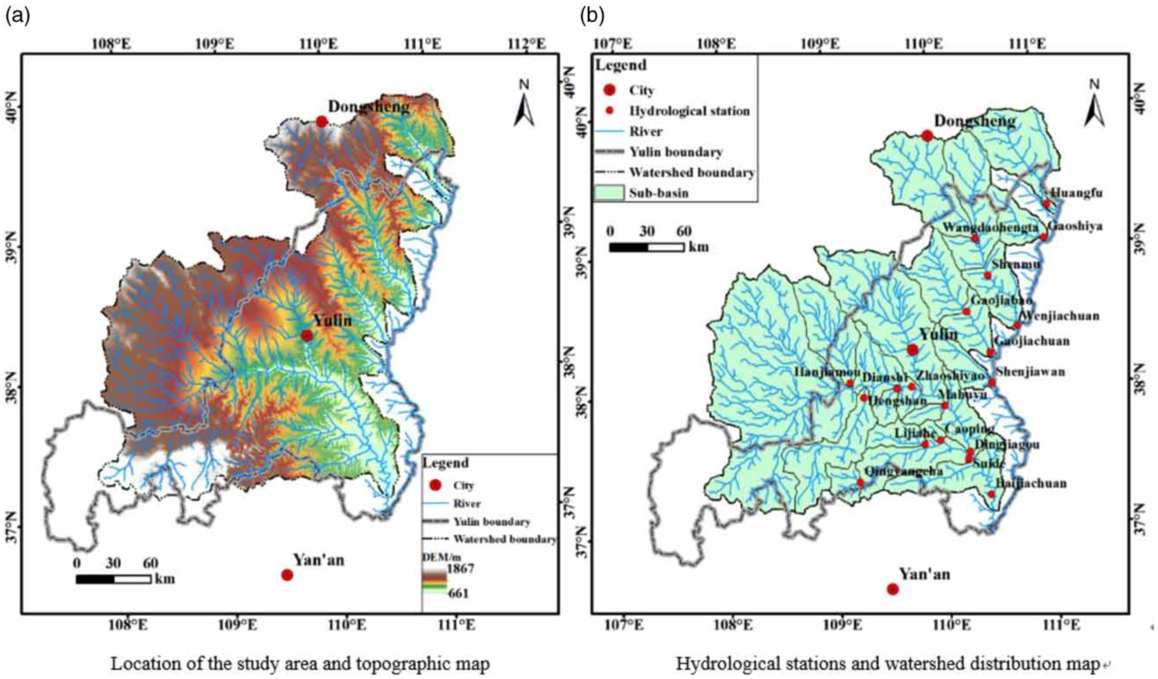
The following parts of the paper are organized as follows. The study area, the data, the distribution functions, evaluation indices and regional regression model are described in the following section. In the next section, the distribution function and their parameters are evaluated, and the multivariate linear regression model of distribution parameters is constructed. Then, the spatial distribution and temporal change of the parameters are analyzed followed by the conclusions.

MATERIALS AND METHODS

Study area

The study area is mainly located in Yulin, Shaanxi, in the northwest of China, and covers a small part of Ningxia and Inner Mongolia as well (Figure 1(a)). The study area is monitored by 19 hydrological stations, which are shown in Table 1. The study area mainly includes the Wuding River Basin (including the Hailiutou River, Heimudou River, Yuxi River, Xiaoli River and Dali River), the Jialu River Basin, the Tuwei River Basin, the Kuye River Basin, the Gushan River Basin, and the Huangfu River Basin (including Shilichang River). The 19 hydrological stations and river basins are shown in Figure 1(b) and Table 1. Table 1 shows the geomorphologic and climatic parameters of each hydrological station, including sub-basin area A (km^2) covered by hydrological stations, elevation difference ΔH (m) of sub-basins, mean elevation H (m) of sub-basins, length of main channel L, perennial flow index IPF (–) (Mendicino *et al.* 2008), average annual precipitation P (mm), average annual precipitation of dry season P_{sum} (mm), month of minimum flow, mean monthly precipitation in minimum flow month P_{mmin} (mm), and monthly precipitation of maximum flow month P_n (mm).

In this study, taking Northern Shaanxi in China as the study area, polynomial and probability distribution functions are used to construct the flow duration curve in the study area. Multiple regression analysis of the characteristic parameters of the curve as functions of the corresponding



Location of the study area and topographic map

Hydrological stations and watershed distribution map

Figure 1 | Study area and location of hydrological stations.

Table 1 | Geomorphological and climatic parameters for catchments of hydrological stations

Number	Hydrological station	A (km^2)	ΔH (m)	H (m)	L (km)	I_{PF}	P (mm)	P_{sum} (mm)	Month of minimum flow	P_{mmin} (mm)	P_n (mm)
H01	Gaoshiya station	1,274.63	580.00	1,176.28	65.14	0.83	426.27	72.72	1	2.71	148.00
H02	Gaojiachuan station	3,505.83	638.00	1,171.09	122.44	1.00	427.83	73.34	6	46.70	12.49
H03	Gaojiabao station	2,138.81	408.00	1,223.60	68.58	1.00	417.75	71.32	6	43.70	1.15
H04	Mahuyu station	376.91	399.00	1,098.68	32.64	0.96	449.85	80.01	1	2.92	95.26
H05	Hanjiamao station	2,574.20	428.00	1,271.81	109.47	1.00	349.51	63.02	6	39.94	16.70
H06	Qingyangcha station	673.27	530.00	1,378.66	26.23	1.00	495.09	90.80	6	3.23	85.02
H07	Zhaoshiyao station	22,138.10	946.00	1,329.69	234.66	0.97	368.02	67.88	6	43.28	3.08
H08	Suide station	3,902.10	901.00	1,202.32	140.25	0.99	471.71	85.08	1	3.07	29.35
H09	Shenmu station	6,930.95	628.00	1,300.30	139.39	1.00	390.07	65.30	1	2.15	19.33
H10	Huangfu station	3,279.66	628.00	1,159.31	115.08	0.58	405.03	69.66	1	2.53	48.75
H11	Baijiachuan station	36,485.42	1,206.00	1,265.55	371.98	1.00	371.12	71.00	5	33.24	187.83
H12	Shenjiawan station	1,147.43	639.00	1,125.80	63.07	0.99	444.33	77.14	1	2.93	102.54
H13	Wangdaohengta station	3,833.95	557.00	1,332.49	119.46	0.99	378.28	62.94	1	2.01	16.15
H14	Wenjiachuan station	12,517.48	818.00	1,260.10	206.07	0.98	278.69	46.88	1	1.60	72.26
H15	Dianshi station	467.09	383.00	1,172.77	32.65	1.00	435.27	78.25	1	2.71	98.93
H16	Hengshan station	2,750.67	799.00	1,385.72	130.14	1.00	433.15	79.75	6	52.81	7.95
H17	Lijiahe station	821.04	473.00	1,183.01	50.42	1.00	441.04	79.02	1	2.79	92.13
H18	Caoping station	205.52	308.00	1,075.25	17.59	0.96	450.14	80.05	1	2.93	73.93
H19	Dingjiagou station	30,123.08	1,058.00	1,291.17	312.91	1.00	347.69	67.69	6	44.05	78.02

geomorphology and climatic characteristics of the sub-basin is conducted to analyze the relationship between the parameters and the spatial features. Finally, the temporal changes of the model parameters at certain sites are studied to evaluate the impacts of land use change over time.

Materials

The daily average flow data of 19 stations is derived from the Yellow River hydrological data (the upper reaches of the Yellow River middle reaches) and are assumed to be virgin flows. The longest data sequence is from 1955 to 2012, a total of 58 years. The shortest data sequence is from 1975 to 2012, a total of 38 years. The length of the data series is sufficient to meet the research requirements. Figure 2 shows the temporal availability of the data series of 19 hydrological stations. The *x*-axis is arranged according to the number in Table 1. The DEM data used in this study is from the official website of the US Geological Survey, which were extracted from NASA Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model (ASTER GDEM) with a spatial resolution of 30 m. The precipitation data (1961–2015) is from China's ground precipitation monthly value 0.5 × 0.5° grid data set (V2.0), on China Meteorological Data Sharing Service Network (http://data.cma.cn).

Methodology

Distribution functions

There are various functions available to simulate a flow duration curve, and the suitable functions are region-specific (Franchini & Suppo 1996; Ganora *et al.* 2009; Viola *et al.* 2011). The distribution functions used in this study to fit the flow duration curve are summarized in Table 2. This study uses the total duration method to construct the flow duration curve.

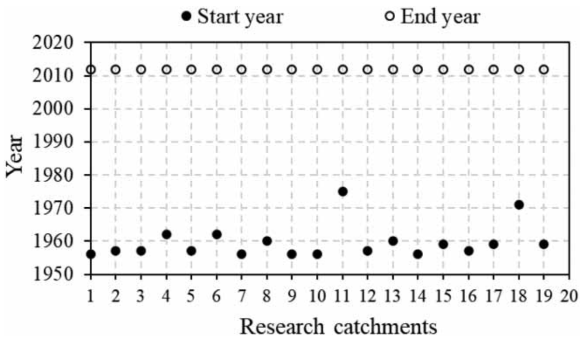


Figure 2 | Temporal availability of the data series of 19 hydrological stations.

Evaluation of the distribution functions

This study used the period from the starting year to 1999 for calibration of parameters of the distribution function, and the period of 2000–2012 for parameter verification at 19 hydrological stations. In this study, the Nash–Sutcliffe efficiency (*NSE*) (Nash & Sutcliffe 1970; Masih *et al.* 2010; Lane *et al.* 2015; Pugliese *et al.* 2016), the root mean square relative error, *RMSRE* (Nruthya & Srinivas 2015) and the coefficient of determination, R^2 (Nagelkerke 1991) are used to evaluate the applicability of different distribution functions at a certain location. *NSE* is defined as:

$$NSE = 1 - \frac{\sum_{i=1}^{N} (O_i - P_i)^2}{\sum_{i=1}^{N} (O_i - \overline{O})^2} \tag{1}$$

where O_i is the observed data on day i, $\mathrm{P_i}$ is the predicted value for day i, and $\overline{O}$ is the mean of the observed data for the entire period (day i to N), N is the total number of daily observations. *NSE* is unity minus the ratio of the mean square error to the variance in the observed data, and ranges from −∞ to 1.0. High *NSE* and low *RMSRE* indicate a suitable fit of the distribution function to the observed flow duration curve. If *NSE* is larger than 0.7, the models are considered as suitable (Nash & Sutcliffe 1970; Gupta *et al.* 2009).

The root mean square relative error, *RMSRE* is calculated as:

$$RMSRE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{O_i - P_i}{O_i}\right)^2} \tag{2}$$

The coefficient of determination R^2 is calculated as:

$$R^2 = \frac{\left[\sum_{i=1}^{N} (O_i - \overline{O})(P_i - \overline{P})\right]^2}{\sum_{i=1}^{N} (O_i - \overline{O})^2 \sum_{i=1}^{N} (P_i - \overline{P})^2} \tag{3}$$

Table 2 | Distribution functions

Number	Name	Expression	Parameter
F01	No 1 function	$y = ae^{-bx}$ (Clark 1951)	a, b
F02	No 2 function	$y = ae^{\left(\frac{b}{x}\right)}$ (Clark 1951)	a, b
F03	Log normal function	$y = ae^{b(lnx)^2}$ (Clark 1951)	a, b
F04	Gamma function	$y = ax^{-b}e^{-cx}$ (Clark 1951)	a, b, c
F05	Exponential distribution function	$F(x) = 1 - e^{-\lambda x}\ x \geq 0$ (Ahmadi *et al.* 2005)	λ
F06	Normal distribution function	$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2\sigma^2}(x-\mu)^2} dx$ (Stein 1981)	μ, σ
F07	Gamma distribution function	$F(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^x x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) dx$ (Stacy 1962)	α, β $\alpha > 0, \beta > 0$
F08	Lognormal distribution function	$F(x) = \frac{1}{x\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{\frac{-[\log x - \mu]}{2\sigma^2}\right\} dx$ $x \geq 0$ (Mosimann 1970)	μ, σ
F09	Gumbel distribution function	$F(x) = exp\left[-e^{-\left(\frac{x-\beta}{\alpha}\right)}\right]$ (Nadarajah & Kotz 2004)	α, β
F10	Generalized extreme value distribution function	$F(x) = exp\left\{-\left[1 - k\left(\frac{x-\beta}{\alpha}\right)\right]^{\frac{1}{k}}\right\}\ k \neq 0$ $F(x) = exp\left[-\exp\left(-\frac{x-\beta}{\alpha}\right)\right]\ k = 0$ (Hosking *et al.* 1985)	α, β, k
F11	Generalized Pareto distribution function	$F(x) = 1 - \left[1 - k\left(\frac{x-\beta}{\alpha}\right)\right]^{\frac{1}{k}}\ k \neq 0, \beta \leq x \leq \beta + \frac{\alpha}{k}$ (Hosking & Wallis 1987)	α, β, k

where, $\overline{P}$ is the mean of the predicted data for the entire period (day i to N).

Regional regression model

Multiple linear regression models are usually used to study the relationship between a dependent variable and multiple independent variables. If the dependence of the two can be described in a linear form, a multivariate linear model can be established for analysis. This study applies multivariate linear regression analysis to evaluate the impacts of geomorphic and climatic factors of sub-basins on the parameters of polynomial model and probability distribution function. The regression model is expressed as:

$$\theta = A_0 + A_1x_1 + A_2x_2 + \cdots\cdots + A_nx_n + \zeta \tag{4}$$

where θ are the estimated parameters, such as a, b, μ, σ, k, β in Table 2; x_i, $i = 1, 2, \ldots, n$ are the variables representing the impacts of geomorphologic and climatic characteristics in the regional model, $A_0, A_1, \ldots, A_n$ are the equation coefficients of the regional predictive models, ξ is the error term of the model. In this study, 10 variables are used to represent the geomorphologic and climatic characteristic as shown in Table 1, namely, n is equal to 10.

RESULTS

Determination of parameters for flow duration curve and evaluation of distribution function

Flow data collected at the 19 hydrological stations were used to select the best distribution function for the Northern Shaanxi Region from the 11 distribution functions listed in Table 2. Tables 3–5 show the values of *NSE*, *RMSRE* and R^2 at each station, respectively. Based on these values, the F03, i.e. the log normal function, is the best distribution function for this region, because the F03 has high *NSE*, low *RMSRE* and high R^2 for most stations. When the log normal function is used, 11 out of the 19 stations have *NSE* values higher than 0.9, and the *NSE* values at 14 out of the stations are higher than 0.8. Among the 19 stations, only two hydrological stations have *NSE* values lower than 0.7, which are Huangfu station (H10) and Wenjiachuan station (H14). At the same time, compared to the other 10

Table 3 | The *NSE* of the fitted distribution functions

Hydrological station	F01	F02	F03	F04	F05	F06	F07	F08	F09	F10	F11
H01	0.793	0.341	0.728	0.962	0.111	−0.423	−150.355	–	−188.308	−2.02E + 34	0.934
H02	0.018	0.938	0.931	0.922	0.053	−0.711	−2.865	0.036	−531.324	0.049	0.452
H03	0.010	0.987	0.957	0.957	0.026	−0.771	−2.980	0.018	−487.344	0.024	0.072
H04	0.997	–	0.990	−1.337	0.002	−0.919	−32.500	0.168	−484.422	0.002	0.436
H05	0.097	0.334	0.839	0.551	0.083	−0.293	−2.935	–	−173.360	0.179	0.422
H06	0.821	0.808	0.942	0.906	0.067	−0.610	−2.582	0.086	−442.705	0.235	−0.213
H07	0.031	0.940	0.914	0.909	0.065	−0.624	−3.299	–	−549.186	0.056	0.152
H08	0.813	0.813	0.915	0.876	0.068	−0.567	−9.675	–	−445.360	0.056	0.681
H09	0.644	0.413	0.806	0.965	0.229	−0.227	−3.299	–	−213.771	0.544	0.846
H10	0.841	0.202	0.657	0.968	0.107	−0.416	−10,383.780	–	−108.297	−14,319	−10.796
H11	0.321	0.184	0.756	0.687	0.557	0.177	−3.710	–	−55.802	0.591	0.941
H12	0.839	0.311	0.797	0.953	0.093	−0.517	−15.287	–	−181.066	0.247	−108.102
H13	0.865	0.000	0.934	0.068	0.047	−0.644	−4.851	–	−477.982	0.737	0.804
H14	0.848	0.184	0.638	0.946	0.212	−0.265	−16.404	–	−91.279	0.861	0.875
H15	0.997	–	0.998	−61.521	0.004	−0.909	−3.047	–	−567.566	0.005	0.923
H16	0.800	0.771	0.953	0.882	0.123	−0.488	−2.842	0.180	−411.640	0.228	0.166
H17	0.991	–	0.993	−2.521	0.008	−1.047	−6.918	–	−510.773	0.083	0.078
H18	0.793	0.428	0.818	0.961	0.126	−0.448	−39.466	–	−183.390	0.259	0.475
H19	0.074	0.619	0.941	0.858	0.159	−0.418	−4.078	–	−315.129	0.142	0.166

Table 4 | The *RMSRE* of the fitted distribution functions

Hydrological station	F01	F02	F03	F04	F05	F06	F07	F08	F09	F10	F11
H01	0.006	0.010	0.0064	0.0024	0.0115	0.0146	3.552	–	0.167	1.1377E + 23	0.009
H02	0.007	0.002	0.0019	0.0020	0.0071	0.0094	0.576	0.007	0.169	0.007	0.011
H03	0.008	0.001	0.0016	0.0016	0.0077	0.0104	0.576	0.008	0.172	0.008	0.016
H04	0.001	–	0.0012	0.0107	0.0130	0.0157	1.671	0.001	0.179	0.013	0.022
H05	0.013	0.011	0.0055	0.0092	0.0077	0.0107	0.573	–	0.172	0.008	0.017
H06	0.003	0.004	0.0019	0.0025	0.0078	0.0102	0.546	0.008	0.171	0.008	0.066
H07	0.007	0.002	0.0021	0.0022	0.0070	0.0091	0.599	–	0.167	0.007	0.013
H08	0.004	0.004	0.0024	0.0029	0.0078	0.0102	0.943	–	0.172	0.008	0.028
H09	0.007	0.009	0.0050	0.0021	0.0099	0.0125	0.599	–	0.170	0.097	0.043
H10	0.007	0.015	0.0097	0.0030	0.0156	0.0198	29.418	–	0.172	7.336	0.259
H11	0.020	0.022	0.0118	0.0133	0.0159	0.0218	0.627	–	0.174	0.015	0.005
H12	0.005	0.011	0.0060	0.0029	0.0128	0.0166	1.165	–	0.179	0.012	0.035
H13	0.007	0.008	0.0077	0.0075	0.0076	0.0100	0.698	–	0.171	0.005	0.015
H14	0.007	0.016	0.0105	0.0040	0.0155	0.0197	1.204	–	0.167	0.005	0.040
H15	0.000	–	0.0003	0.0567	0.0072	0.0099	0.581	–	0.169	0.007	0.030
H16	0.004	0.004	0.0018	0.0028	0.0077	0.0101	0.566	0.008	0.170	0.007	0.008
H17	0.001	–	0.0006	0.0140	0.0074	0.0107	0.812	–	0.171	0.007	0.008
H18	0.006	0.010	0.0055	0.0025	0.0120	0.0153	1.836	–	0.172	0.011	0.011
H19	0.009	0.006	0.0023	0.0036	0.0087	0.0113	0.651	–	0.168	0.009	0.010

Table 5 | The R^2 of the fitted distribution functions

Hydrological station	F01	F02	F03	F04	F05	F06	F07	F08	F09	F10	F11
H01	0.801	0.342	0.771	0.962	0.221	0.292	0.478	–	0.202	0.387	0.940
H02	0.019	0.938	0.955	0.962	0.053	0.150	0.229	0.225	0.116	0.080	0.730
H03	0.010	0.987	0.986	0.985	0.027	0.112	0.558	0.160	0.089	0.036	0.184
H04	0.997	0.000	0.990	0.900	0.009	0.040	0.124	0.423	0.024	0.071	0.944
H05	0.097	0.336	0.840	0.609	0.184	0.348	0.620	–	0.299	0.190	0.694
H06	0.842	0.809	0.942	0.914	0.089	0.194	0.337	0.421	0.143	0.612	0.705
H07	0.031	0.940	0.968	0.967	0.065	0.191	0.093	–	0.160	0.062	0.350
H08	0.830	0.813	0.916	0.882	0.099	0.212	0.035	–	0.150	0.438	0.692
H09	0.692	0.415	0.832	0.966	0.320	0.384	0.014	–	0.291	0.872	0.872
H10	0.848	0.204	0.715	0.968	0.230	0.291	0.735	–	0.196	0.247	0.651
H11	0.329	0.186	0.779	0.818	0.558	0.587	0.004	–	0.485	0.660	0.954
H12	0.854	0.314	0.831	0.959	0.142	0.238	0.055	–	0.169	0.625	0.513
H13	0.874	0.000	0.934	0.069	0.078	0.173	0.047	–	0.121	0.771	0.915
H14	0.880	0.185	0.687	0.951	0.319	0.369	0.051	–	0.269	0.897	0.929
H15	0.997	–	0.998	0.911	0.012	0.047	0.033	–	0.029	0.064	0.924
H16	0.870	0.772	0.953	0.906	0.130	0.258	0.100	0.506	0.204	0.336	0.207
H17	0.992	–	0.993	0.896	0.017	0.072	0.040	–	0.046	0.680	0.421
H18	0.811	0.430	0.847	0.965	0.197	0.272	0.184	–	0.196	0.597	0.847
H19	0.076	0.621	0.942	0.879	0.159	0.293	0.005	–	0.240	0.167	0.213

distribution functions, the log normal model has the lowest *RMSRE* values. For the log normal model, the maximum *RMSRE* is 0.0118 at Station H11, and the lowest *RMSRE* is 0.0003 at Station H15.

Table 6 shows the best-fitted parameters of the log normal distributions function (F03) for the 19 hydrological stations.

Figure 3 shows the observed flow and simulated flow duration curves using the log normal model at four typical sites, Baijiachuan station and Dingjiagou station in Wuding River Basin, Gaojiachuan station in Tuwei River Basin and Wenjiachuan station in Kuye River Basin. These four stations control the largest areas in their rivers. For all of the sites, the log normal function fits the observed flow frequency well while this model fits poorly at some stations. This 'log normal function' is listed as F03 in Table 2 and it is not the log normal distribution function as we usually know it, which is H08 in Table 2. The LN function in Blum *et al.* (2017) is the log normal distribution function. So the fitted curves are different from that in Blum *et al.* (2017).

Table 6 | The parameters of the log normal distribution function (F03) as the optimal distribution function

Hydrological station	a	b	Hydrological station	a	b
H01	4.138	0.059	H12	3.352	0.060
H02	1.675	0.084	H13	4.411	0.075
H03	2.78×10^{-6}	0.227	H14	25.171	0.055
H04	0.104	0.115	H15	1.11×10^{-4}	0.178
H05	2.379	0.050	H16	1.682	0.064
H06	0.780	0.069	H17	0.066	0.113
H07	0.266	0.104	H18	0.341	0.062
H08	4.087	0.070	H19	22.967	0.059
H09	19.226	0.057			
H10	9.096	0.058			
H11	29.287	0.049			

Impacts of geomorphological and climatic factors

Based on the best-fitted values of the model parameters (Table 6), multivariate linear regression of distribution

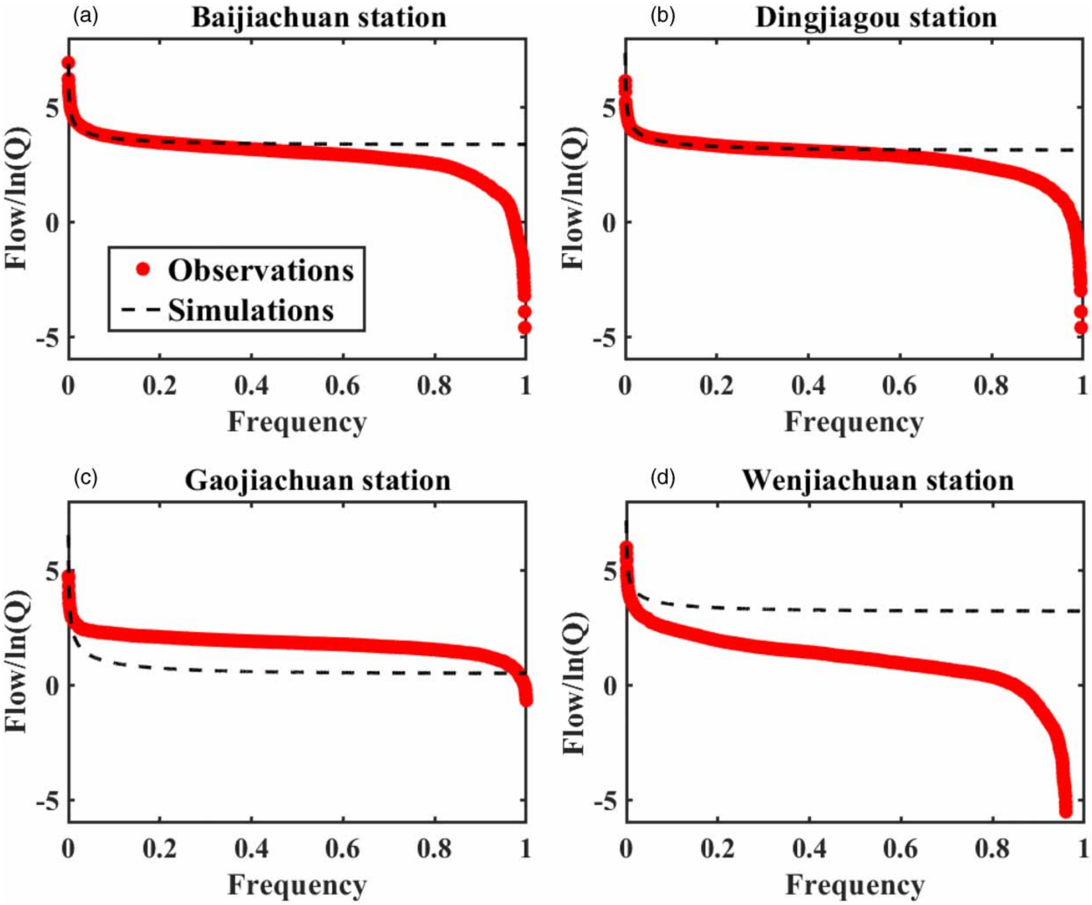


Figure 3 | Observed and simulated flow duration curves at four hydrological stations.

parameters is used to evaluate the impacts of the geomorphologic and climatic factors. A regional parameter prediction model is developed as:

$$a = -5.087 - 4.812 \times 10^{-3} A + 0.3075\, L - 6.375 \times 10^{-7} A^2 + 1.011 \times 10^{-4} AL - 3.082 \times 10^{-3}\, L^2 \quad (5)$$

For this fitting, $R^2 = 0.9085$, $F = 17.3427$, $RMSRE = 0.0873$. F is the value of joint hypotheses test:

$$b = \frac{0.2467 \times (\ln P_n)^2 - 2.288 \times (\ln P_n) + 5.307}{(\ln P_n)^3 - 8.22 \times (\ln P_n)^2 + 12.1 \times (\ln P_n) + 20.93} \quad (6)$$

For this fitting, $R^2 = 0.9351$, $F = 37.4535$, $RMSRE = 0.0140$.

The regression values of a and b parameters all pass the $F_{0.05}$ confidence test (Steiger 2004), which demonstrates that the prediction model proposed in this study is reasonable for the Northern Shaanxi Region of China. The a and b values calculated according to Equations (5) and (6) are compared with the values in Table 6 (as shown in Figure 4), and it is found that the parameter prediction model is relative reasonable.

Spatial distribution of the parameters

The spatial distributions of the parameters a and b of 19 hydrological stations in Table 6 are shown in Figures 5 and 6, respectively. As shown in Figure 5, the value of parameter a is high in the east (downstream) and low in the west (upstream). The hydrological stations in the lower reaches of the rivers (into the Yellow River) have higher a

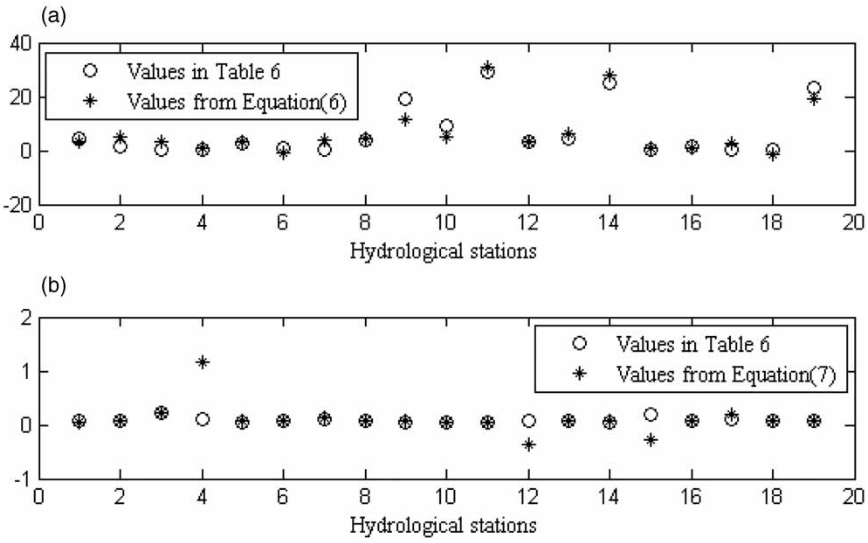


Figure 4 | Comparative analysis of Table 6 and equation of parameters (a) and (b).

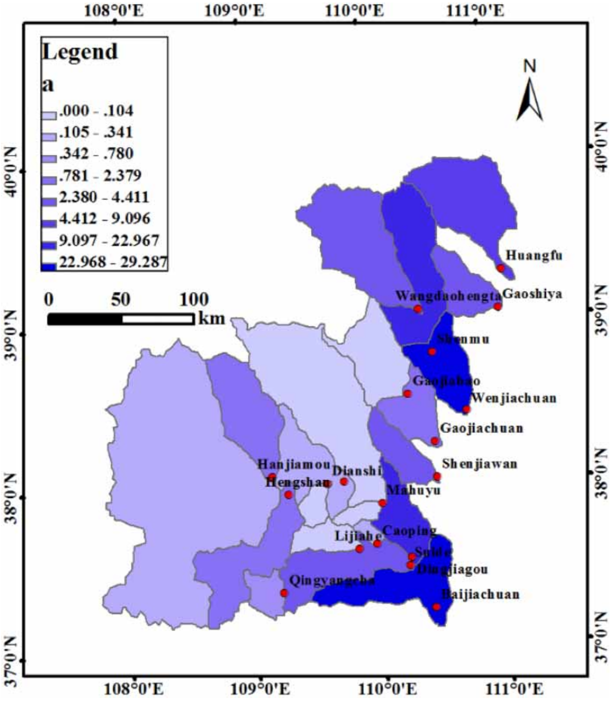


Figure 5 | Spatial distribution of parameter (a).

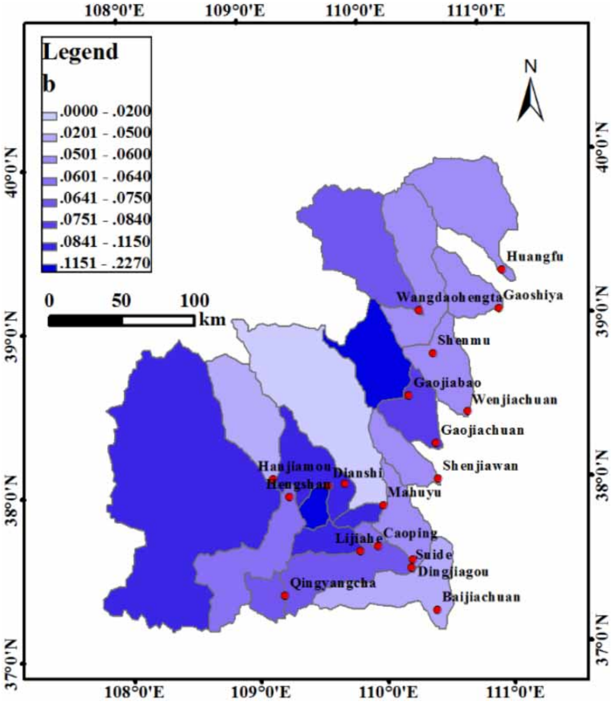


Figure 6 | Spatial distribution of parameter (b).

value, and the hydrological stations in the middle and western rivers show lower a value. Figure 6 shows that parameter b and parameter a basically show the opposite spatial distribution, which is, the b value is relatively large in the west and relatively small in the east. The changes of parameters a and b are caused by changes in regional climate, geomorphology and sub-catchment area. This can be explained by Equations (5) and (6). Based on the spatial distribution of the parameters *a* and *b*, the values of parameters *a* and *b* in different sub-basin ranges can be roughly estimated, so that the flow duration curve of any river section in the region can be effectively estimated.

Temporal change of the parameters

Due to the impacts of human activities on the underlying surface and climate, the geomorphologic and climatic conditions in the area have changed over time. These changes also affect the distribution parameters of the flow duration curves. This section reveals the temporal changes of distribution parameters in the Northern Shaanxi Region. Baijiachuan hydrological station, which covers the entire Wuding River Basin, is used as an example to present the temporal changes of the parameters. The flow duration curves of different periods are analyzed and the values of the parameters are shown in Table 7. Before 1978, the land use was relatively unchanged; after 1978, agricultural area increased quickly. Therefore, the parameters of the flow duration curve are evaluated in different periods as shown in Table 7.

Table 7 demonstrates decreasing trends for both parameters *a* and *b* over time and the flow duration curves are shown in Figure 7. These changes may be caused by the change of land use and vegetation status, and the interaction between humans and water (Liu *et al.* 2015) in the basin, especially reservoir operation (Zhang *et al.* 2021). Based on the Shaanxi Provincial Statistical Yearbook (1984–2012), it was found that the irrigation water consumption and the construction of reservoirs in the Wuding River Basin showed a gradually increasing and then decreasing trend, peaking in the 1990s. This also led to the longest duration of the low-value flow in Figure 7 from 1995 to 2004. In this process, the hydrological conditions are also changing (Huang *et al.* 2015).

Table 7 | Temporal changes of the distribution parameters at Baijiachuan hydrological station

Time interval	*a*	*b*	*NSE*	R^2
1975–1978	33.4168	0.0752	0.9135	0.9201
1979–2012	27.6823	0.0484	0.7725	0.7923
1975–1984	31.4797	0.0604	0.8980	0.9052
1985–1994	28.5699	0.0609	0.9008	0.9078
1995–2004	25.1597	0.0590	0.8111	0.8272
2005–2012	22.0017	0.0541	0.7612	0.7767

The simulated flow duration curves by log normal function (F03) at Baijiachuan are plotted in Figure 7. At the medium flow part and high flow part, simulated flow duration curves fit well in the periods of 1995–2004 and 2005–2012, but the low flow part is not fitted satisfactorily. At Baijiachuan (H11 in Figure 3(a)), the best function is F11 (Generalized Pareto distribution function). At Baijiachuan (H11, upper-left figure in Figure 3), the best function is F11 (Generalized Pareto distribution function).

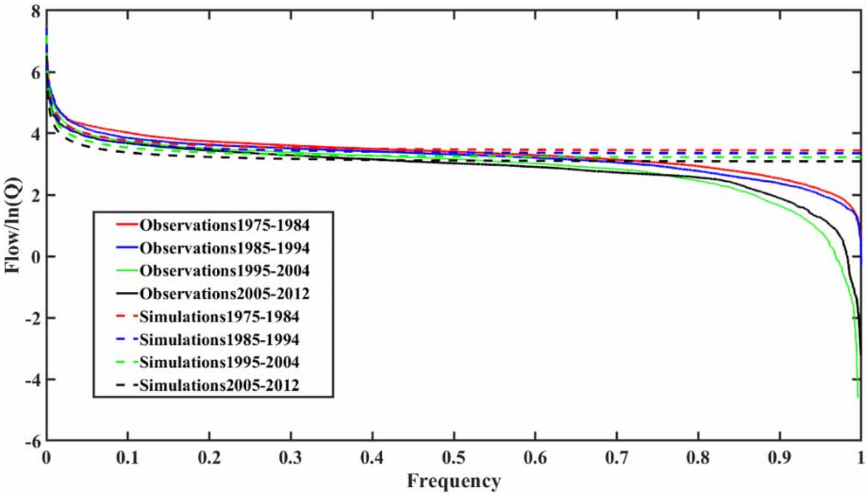


Figure 7 | The flow duration curves of 4 periods at Baijiachuan hydrological station.

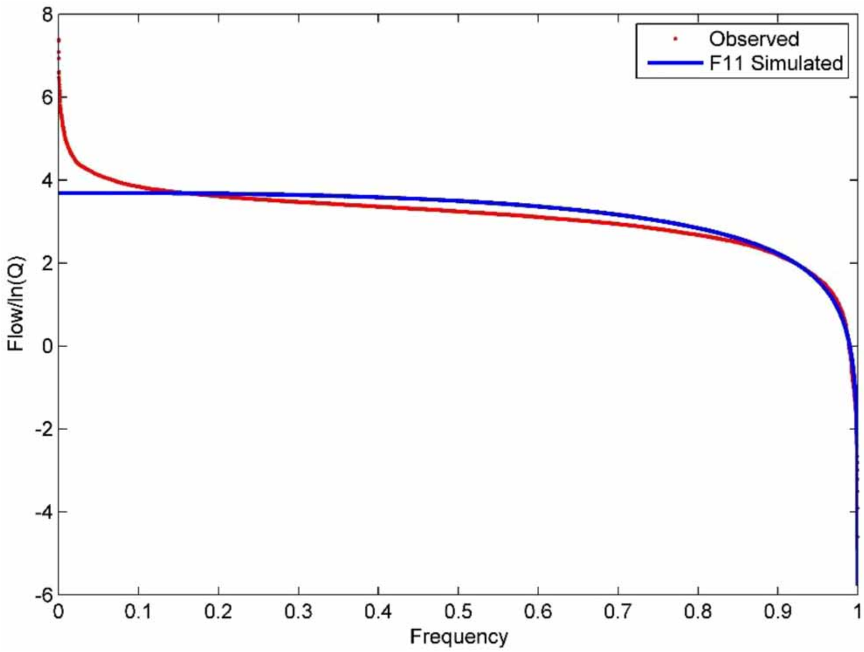


Figure 8 | The simulated flow duration curves by F11 at Baijiachuan in 1975–2012.

The simulated flow duration curves by F11 at Baijiachuan are plotted in Figure 8. At the medium flow part and low flow part, simulated flow duration curves fit well, but the high flow part is not fitted satisfactorily. So perhaps the performance of the current functions for the flow duration curve is not good enough in this region. New functions for the flow duration curve should be tested to find a better function to simulate the flow duration curve in the region.

CONCLUSIONS

In this study, the flow duration curve of 19 hydrological stations in Northern Shaanxi in China was studied, and it was found that the log normal function is the optimal model among a total of 11 function models to simulate the flow duration curves in Northern Shaanxi. Then the two parameters of the log normal function are studied in parameter regionalization, and it is found that the two parameters have a strong regression relationship with regional climate and geomorphologic characteristics. The spatial distribution of parameter a and b of the log normal model is also evaluated for the entire region, which shows high values of parameter a in the east and low values of parameter a in the west. Parameter b and parameter a showed the opposite spatial distribution, which is that b value is relatively large in the west and relatively small in the east.

Taking Baijiachuan hydrological station as an example, the temporal changes of distribution parameters are analyzed. The study shows that, for the Wuding River Basin, which is covered by the Baijiachuan station, both parameters a and b show a decreasing trend over the study period. This change may be related to the land use, vegetation status, reservoir operation and water abstractions for irrigation.

This study presents an innovative approach to evaluate regionalized parameters of flow duration curves based on geomorphologic and climatic characteristics, which can be used to estimate the water resources and provide guidance for construction and planning of water conservancy projects in the catchments without measured runoff data in China. New functions for the flow duration curve should be tested to find a better function to simulate the flow duration curve in the region, and attention should be paid to the fact that due to the change of land use/land cover, and reservoir operation, the parameters of the flow duration curve will change. Therefore, in the water resources management of the river basin, the construction of the flow duration curve is an on-going task as the evolution of the underlying conditions in a river basin.

ACKNOWLEDGEMENTS

This work was partially supported by the National Natural Science Foundation of China (NSFC) (Grant No. 51779203, 51609270, 51939009), Basic Research Plan of Natural Science of Shaanxi Province (2016JQ5105) and Planning project of science and technology of water resources of Shaanxi (2014slkj-04). The authors would like to thank the editor and the anonymous reviewers for their constructive comments.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

Ahmadi, J., Doostparast, M. & Parsian, A. 2005 Estimation and prediction in a two-parameter exponential distribution based on k-record values under LINEX loss function. *Communications in Statistics – Theory and Methods* **34** (4), 795–805. https://doi.org/10.1081/STA-200054393.

Atieh, M., Gharabaghi, B. & Rudra, R. 2015 Entropy-based neural networks model for flow duration curves at ungauged sites. *Journal of Hydrology* **529**, 1007–1020. https://doi.org/10.1016/j.jhydrol.2015.08.068.

Basso, S., Schirmer, M. & Botter, G. 2015 On the emergence of heavy tailed streamflow distributions. *Advances in Water Resources* **82**, 98–105. https://doi.org/10.1016/j.advwatres.2015.04.013.

Blum, A. G., Archfield, S. A. & Vogel, R. M. 2017 On the probability distribution of daily streamflow in the United States. *Hydrology and Earth System Sciences* **21** (6), 3093–3103. https://doi.org/10.5194/hess-21-3093-2017.

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

Boscarello, L., Ravazzani, G., Cislaghi, A. & Mancini, M. 2016 Regionalization of flow-duration curves through catchment classification with streamflow signatures and physiographic-climate indices. *Journal of Hydrologic Engineering* **21** (3), 05015027. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001307.

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

Cheng, L., Yaeger, M., Viglione, A., Coopersmith, E., Ye, S. & Sivapalan, M. 2012 Exploring the physical controls of regional patterns of flow duration curves – part 1: insights from statistical analyses. *Hydrology and Earth System Sciences* **16** (11), 4435–4446. https://doi.org/10.5194/hess-16-4435-2012.

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 Journal* **65** (3), 348–370. https://doi.org/10.1016/j.advwatres.2004.08.005.

Clark, C. 1951 Urban population densities. *Journal of Royal Statistical Society* **114** (4), 490–496.

Fennessey, N. M. & Vogel, R. M. 1990 Regional flow-duration curves for ungauged sites in Massachusetts. *Journal of Water Resources Planning and Management* **116** (4), 530–549. https://doi.org/10.1061/(ASCE)0733-9496(1990)116:4(530).

Franchini, M. & Suppo, M. 1996 Regional analysis of flow duration curves for a limestone region. *Water Resources Management* **10** (3), 199–218. https://doi.org/10.1007/BF00424203.

Fu, B., Liu, Y., He, C., Zeng, Y. & Wu, B. 2011 Assessing the soil erosion control service of ecosystems change in the Loess Plateau of China. *Ecological Complexity* **8** (4), 284–293. https://doi.org/10.1016/j.ecocom.2011.07.003.

Fu, B., Wang, S., Liu, Y., Liu, J., Liang, W. & Miao, C. 2017 Hydrogeomorphic ecosystem responses to natural and anthropogenic changes in the Loess Plateau of China. *Annual Review of Earth and Planetary Sciences* **45**, 223–243. https://doi.org/10.1146/annurev-earth-063016-020552.

Ganora, D., Claps, P., Laio, F. & Viglione, A. 2009 An approach to estimate nonparametric flow duration curves in ungauged basins. *Water Resources Research* **45** (10), W10418. https://doi.org/10.1029/2008WR007472.

Gupta, H. V., Kling, H., Yilmaz, K. K. & Martinez, G. F. 2009 Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *Journal of Hydrology* **377** (1), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

Hosking, J. R. M. & Wallis, J. R. 1987 Parameter and quantile estimation for the Generalized Pareto distribution. *Technometrics* **29** (3), 339–349. https://doi.org/10.2307/1269343.

Hosking, J. R. M., Wallis, J. R. & Wood, E. F. 1985 Estimation of the Generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* **27** (3), 251–261. https://doi.org/10.2307/1269706.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., Mcdonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E. & Cudennec, C. 2013 A decade of predictions in ungauged basins – a review. *Hydrological Sciences Journal* **58** (6), 1198–1255. https://doi.org/10.1080/02626667.2013.803183.

Huang, S. Z., Chang, J. X., Leng, G. Y. & Huang, Q. 2015 Integrated index for drought assessment based on variable fuzzy set theory: a case study in the Yellow River basin, China. *Journal of Hydrology* **527**, 608–618. https://doi.org/10.1016/j.jhydrol.2015.05.032.

Lane, P. N. J., Best, A. E., Hickel, K. & Zhang, L. 2015 The response of flow duration curves to afforestation. *Journal of Hydrology* **310** (1–4), 253–265. https://doi.org/10.1016/j.jhydrol.2005.01.006.

Li, M., Shao, Q. X., Zhang, L. & Chiew, F. H. S. 2010 A new regionalization approach and its application to predict flow duration curve in ungauged basins. *Journal of Hydrology* **389**, 137–145. https://doi.org/10.1016/j.jhydrol.2010.05.039.

Li, P., Mu, X., Holden, J., Wu, Y., Irvine, B., Wang, F., Gao, P., Zhao, G. & Sun, W. 2017 Comparison of soil erosion models used to study the Chinese Loess Plateau. *Earth-Science Reviews* **170**, 17–30. https://doi.org/10.1016/j.earscirev.2017.05.005.

Liang, Z. M., Li, G. F. & Wang, J. 2019 *Hydrological Analysis and Calculation*. China Water & Power Press, Beijing, China.

Liu, D. F., Tian, F. Q., Lin, M. & Sivapalan, M. 2015 A conceptual socio-hydrological model of the co-evolution of humans and water: case study of the Tarim River basin, western China. *Hydrology and Earth System Sciences* **19** (2), 1035–1054. https://doi.org/10.5194/hess-19-1035-2015.

Liu, D. F., Tian, F. Q., Li, H. Y., Lu, H., Lin, M. & Sivapalan, M. 2019 Temporal and spatial signatures of sediment transport at the watershed scale: an approach to understand the behavior of the watershed. *Tecnología y ciencias del agua* **10** (4), 18–45. https://doi.org/10.24850/j-tyca-2019-04-02.

Longobardi, A. & Villani, P. 2013 A statistical, parsimonious, empirical framework for regional flow duration curve shape prediction in high permeability Mediterranean region. *Journal of Hydrology* **507**, 174–185. https://doi.org/10.1016/j.jhydrol.2013.10.019.

Masih, I., Uhlenbrook, S., Maskey, S. & Ahmad M, D. 2010 Regionalization of a conceptual rainfall–runoff model based on similarity of the flow duration curve: a case study from the semi-arid Karkheh basin, Iran. *Journal of Hydrology* **391** (1–2), 188–201. https://doi.org/10.1111/febs.12708.

Mendicino, G. & Senatore, A. 2013 Evaluation of parametric and statistical approaches for the regionalization of flow duration curves in intermittent regimes. *Journal of Hydrology* **480**, 19–32. https://doi.org/10.1016/j.jhydrol.2012.12.017.

Mendicino, G., Senatore, A. & Versace, P. 2008 A Groundwater Resource Index (GRI) for drought monitoring and forecasting in a Mediterranean climate. *Journal of Hydrology* **357** (3–4), 282–302. https://doi.org/10.1016/j.jhydrol.2008.05.005.

Mosimann, J. E. 1970 Size allometry-size and shape variables with characterizations of Lognormal and generalized Gamma distributions. *Journal of the American Statistical Association* **65** (330), 930–945. https://doi.org/10.2307/2284599.

Mueller, M. F. & Thompson, S. E. 2016 Comparing statistical and process-based flow duration curve models in ungauged basins and changing rain regimes. *Hydrology and Earth System Sciences* **20**, 669–683. https://doi.org/10.5194/hess-20-669-2016.

Nadarajah, S. & Kotz, S. 2004 The beta Gumbel distribution. *Mathematical Problems in Engineering* **4**, 323–332. https://doi.org/10.1155/S1024123X04403068.

Nagelkerke, N. J. D. 1991 A note on a general definition of the coefficient of determination. *Biometrika* **78** (3), 691–692. https://doi.org/10.1093/biomet/78.3.691.

Nash, J. E. & Sutcliffe, J. V. 1970 River forcasting using conceptual models. Part I: a discussion of principles. *Journal of Hydrology* **10**, 280–290. https://doi.org/10.1016/0022-4(70)90255-6.

Nruthya, K. & Srinivas, V. V. 2015 Evaluating methods to predict streamflow at ungauged sites using regional flow duration curves: a case study. *Aquatic Procedia* **4**, 641–648. https://doi.org/10.1016/j.aqpro.2015.02.083.

Pugliese, A., Farmer, W. H., Castellarin, A., Archfield, S. A. & Vogel, R. M. 2016 Regional flow duration curves: geostatistical techniques versus multivariate regression. *Advances in Water Resources* **96**, 11–22. https://doi.org/10.1016/j.advwatres.2016.06.008.

Reichl, F. & Hack, J. 2017 Derivation of flow duration curves to estimate hydropower generation potential in data-scarce regions. *Water* **9**, 572–586. https://doi.org/10.3390/w9080572.

Schaefli, B., Rinaldo, A. & Botter, G. 2013 Analytic probability distributions for snow-dominated streamflow. *Water Resources Research* **49**, 2701–2713. https://doi.org/10.1002/wrcr.20234.

Stacy, E. W. 1962 Generalization of Gamma distribution. *Annals of Mathematical Statistics* **33** (3), 1187–1188. https://doi.org/10.1214/aoms/1177704481.

Steiger, J. H. 2004 Beyond the F test: effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods* **9** (2), 164–182. https://doi.org/10.1037/1082-989X.9.2.164.

Stein, C. M. 1981 Estimation of the mean of a multivariate normal-distribution. *Annals of Statistics* **9** (6), 1135–1151. https://doi.org/10.1214/aos/1176345632.

Viola, F., Noto, L. V., Cannarozzo, M. & La Loggia, G. 2011 Regional flow duration curves for ungauged sites in Sicily. *Hydrology and Earth System Sciences* **15** (1), 323–331. https://doi.org/10.5194/hess-15-323-2011.

Wang, S., Fu, B., Piao, S., Lu, Y., Ciais, P., Feng, X. & Wang, Y. 2015 Reduced sediment transport in the Yellow River due to anthropogenic changes. *Nature Geoscience* **9** (1), 38–41. https://doi.org/10.1038/ngeo2602.

Yokoo, Y. & Sivapalan, M. 2011 Towards reconstruction of the flow duration curve: development of a conceptual framework with a physical basis. *Hydrology and Earth System Sciences* **15**, 2805–2819. https://doi.org/10.5194/hess-15-2805-2011.

Zhang, Y., Vaze, J., Chiew, F. H. S. & Li, M. 2015 Comparing flow duration curve and rainfall-runoff modelling for predicting daily runoff in ungauged catchments. *Journal of Hydrology* **525**, 72–86. https://doi.org/10.1016/j.jhydrol.2015.03.043.

Zhang, L. P., Huang, Q., Liu, D. F., Deng, M. J., Zhang, H. X., Pan, B. Z. & Zhang, H. G. 2021 Long-term and mid-term ecological operation of cascade hydropower plants considering ecological water demands in arid region. *Journal of Cleaner Production* **279**, 123599. https://doi.org/10.1016/j.jclepro.2020.123599.

First received 23 November 2020; accepted in revised form 11 April 2021. Available online 13 May 2021