

Theoretical derivation for the exceedance probability of corresponding flood volume of the equivalent frequency regional composition method in hydrology

Yixin Huang, Zhongmin Liang, Yiming Hu, Binquan Li and Jun Wang

ABSTRACT

The equivalent frequency regional composition (EFRC) method is an important and commonly used tool to determine the design flood regional composition at various sub-catchments in natural conditions. One of the cases in the EFRC method assumes that the exceedance probabilities of design flood volume at upstream and downstream sites are equal, and the corresponding flood volume at intermediate catchment equals the gap between the volumes of upstream and downstream floods. However, the relationship between the exceedance probability of upstream and downstream flood volumes P and that of corresponding intermediate flood volume C has not been clarified, and whether $P > C$ or $P \leq C$ has not been theoretically proven. In this study, based on the normal, extreme value type I and Logistic distributions, the relationship between C and P is deduced via theoretical derivations, and based on the Pearson type III, two-parameter lognormal and generalized extreme value distributions, the relationship between C and P is investigated using Monte Carlo experiments. The results show that C is larger than P in the context of the design flood, whereas P is larger than C in the context of low-flow runoff. Thus, the issue of exceedance probability corresponding flood is further theoretically clarified using the EFRC method.

Key words | design flood, equivalent frequency regional composition method, flood distributions, flood regional composition, Monte Carlo experiments

Yixin Huang
Zhongmin Liang
Yiming Hu (corresponding author)
Binquan Li
Jun Wang
College of Hydrology and Water Resources,
Hohai University,
Nanjing,
China
E-mail: yiming.hu@hhu.edu.cn

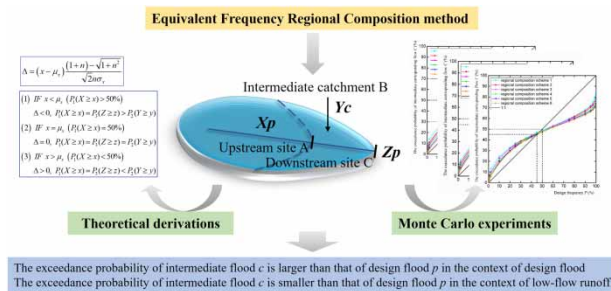
HIGHLIGHTS

- The relationship between the intermediate exceedance probability and the design exceedance frequency in the equivalent frequency regional composition method depends on the design exceedance frequency and the distribution of flood volume.
- The relationship is investigated via theoretical derivations and Monte Carlo experiments.
- The intermediate exceedance probability is larger than the design exceedance frequency in the context of design flood.
- The design exceedance frequency is larger than the intermediate exceedance probability in the context of low flow.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

doi: 10.2166/nh.2020.027

GRAPHICAL ABSTRACT



INTRODUCTION

The ability to estimate the design flood in a given return period is a fundamental issue in engineering design, as well as water resource management and planning. Hydrological frequency analyses have been used worldwide as a standard approach for estimating design flood (Kendall & Stuart 1979; Ponce 1989; Hu *et al.* 2018). The estimation of design flood generally of interest for hydrologists, engineers, and agriculturalists for the design of hydraulic structures, such as river sections and dam sites. When control engineering has not been implemented upstream of a future dam site, the design flood of the future dam can be directly calculated via a hydrological frequency analysis of the peak or duration of the flood volume series, and the flood regional composition does not need to be considered (Maidment 1992). However, for cases in which one or more control engineering features have been implemented upstream of the future dam site, such as a cascade reservoir system, the impact of the outflows at upstream site and intermediate catchment (i.e., the flood regional composition) must be considered to estimate the design flood at the future dam site. Generally, the framework for flood regional composition (Lu *et al.* 2012; Guo *et al.* 2018) consists of a first step, where a proper combination of floods at various sub-catchments is researched in natural conditions, and a second step, where the design flood (flood discharges or volumes) at the downstream site under the influence of upstream reservoir is obtained through flood routing by incorporating the reservoir operation rules. The first step of regional composition assumes that no reservoirs exist and that each sub-

catchment is in a natural condition. When a design flood event with exceedance probability p is selected from the natural flood magnitude–frequency curve at a downstream site, an appropriate combination of floods is required that occurred at upstream sites. The corresponding natural design flood hydrograph at sub-catchments of upstream sites is derived by using the same flood amplification ratio of their respective typical flood hydrograph (Yue *et al.* 2002; Xiao *et al.* 2009). Before the characteristics of the reservoir are determined, the most important step is to find an appropriate combination of floods that occurred at upstream sites (Boughton & Droop 2003).

For on-site hydrological frequency analyses that do not consider the impact of flood regional composition, the process includes the selection of distribution functions and the estimation of parameters (Rao & Hamed 2000; Badrel-din *et al.* 2012; Zeng *et al.* 2012). Based on the sample series, a design flood with a given return period is easily estimated (Kirby 1974). In different countries or regions, the recommended distribution functions used to fit extreme flood series may vary; for example, the Pearson type III (PE3) distribution has been implemented in China, the log PE3 distribution has been implemented in the United States, and the generalized extreme value (GEV) distribution has been implemented in England. In the parameter estimation of a distribution function, the method of moments (MOM) was once a widely used approach, although it has a high degree of bias (Wallis *et al.* 1974; Greenwood 1979). Thus, other methods with

less bias have been successively provided, such as the maximum-likelihood method (MLM), probability weighted moments method (PWM), weight function method, and linear moments method (Hosking 1986, 1990; Ding *et al.* 1989; Liang *et al.* 2014; Wang *et al.* 2015). Due to the impact of climate change and anthropogenic activities, many studies from the past several decades indicate that the non-stationary nature of hydrological extreme series has become increasingly significant, and the non-stationary hydrological frequency analysis method has drawn significant attention from researchers and engineering practitioners (Xiong *et al.* 2015; Hu *et al.* 2017, 2018).

Compared with on-site hydrological frequency analysis, a multi-site hydrological frequency analysis that considers the flood regional composition is more complicated and difficult. This process should consider not only the design flood of each site but also the influence of flood regional composition on the design flood (flood discharges or volumes) of the future dam site. It can be seen that the number of possible regional composition is countless and the selection of an appropriate combination is significant. In practice, several combinations, such as the best, the worst, and the most likely, are used to simulate the impact of upstream sites on design flood at downstream site (Nijssen *et al.* 2009). With respect to the design flood regional composition analysis, semi-theoretical and semi-empirical methods, such as the regional composition method, frequency combination method, and stochastic simulation method, have been widely practiced for decades. The regional composition method specifies that a flood occurs in one catchment with the same exceedance probability as in the design section, and a corresponding flood occurs in the other catchments (Ministry of Water Resources 2006). Among various possible compositions, the equivalent frequency regional composition (EFRC) method is able to select a specific one as the designed regional composition model to ensure the safety of the calculated results (Lu *et al.* 2012; Guo *et al.* 2018). The EFRC method includes two cases. In the first case, the corresponding flood at intermediate catchment is calculated if the exceedance probabilities of the design flood volume at upstream site and downstream site are equivalent. In the second case, the corresponding flood at upstream site is calculated if the exceedance probabilities of design flood volume at intermediate catchment

and downstream site are equivalent. The EFRC method is a back-calculation model based on the water balance, which is used to calculate the corresponding flood volume for an intermediate catchment without stream gauging stations. It plays an important role in multi-site hydrological frequency analysis, especially in China (Liang *et al.* 2016).

Regardless of which flood regional composition method is used, the volume of the corresponding flood is known, while the exceedance probability (or return period) of the corresponding flood volume is unknown. Using one case of the EFRC method as an example, such as one in which the equivalent exceedance probability of the design flood volume at upstream site and downstream site is used to calculate the corresponding flood at intermediate catchment, the corresponding flood volume is equivalent to the gap between the volumes of upstream and downstream floods. Whether the exceedance probability of the corresponding flood volume is larger or smaller than that of the design flood volume at downstream site has not been theoretically proven, which has introduced confusion into practical probabilistic applications.

The aim of our study is to investigate this unresolved question and to clarify it using theoretical derivations and Monte Carlo (MC) experiments. Considering that different countries or regions use various flood distributions, we selected six representative design flood volume distributions for our research. The normal, extreme value type I (EV1(2)) and logistic distributions use theoretical derivations because the joint distribution functions of them are easy to derive. The PE3, two-parameter lognormal (LN(2)) and GEV distributions are used to carry out MC experiments. Since their joint distribution functions are transcendental functions, it is difficult to calculate them directly.

METHODOLOGY

Flood regional composition

The flood regional composition is used to determine the design flood volumes at downstream sites. Many possible flood regional combinations occur in various sub-catchments. Different combinations can result in different design flood volumes at downstream site.

Figure 1 is a schematic of the natural river system, in which Z is the flood volume of downstream site C. Figure 1(a) shows a natural on-site C that does not involve the flood regional composition. The process of on-site hydrological frequency analyses includes only the selection of distribution functions and the estimation of parameters. As shown in Figure 1(b), site A was constructed upstream of site C. Therefore, the estimation of the design flood volume at site C should consider the flood regional composition impact of upstream, which means that the inflow into site C is divided into two parts: site A and intermediate catchment B between site A and site C. According to the principle of water balance, the flood volume at site C is the sum of flood volume at site A and intermediate catchment B, as shown in the following equation:

$$X + Y = Z \quad (1)$$

where random variables X , Y , and Z represent the flood volume of upstream site A, intermediate catchment B, and downstream site C, respectively.

The cascade intermediate catchments system, as shown in Figure 1(c), can be divided into many small single intermediate sub-catchments, as shown in Figure 1(b). Every single sub-catchment has a similar design flood volume composition. As the number of sub-catchments (n) increases, the number of flood regional compositions (n) increases uniformly.

EFRC method

The EFRC method is widely used to determine the design flood regional composition in China, and it is recommended by the Ministry of Water Resources of China (Ministry of Water Resources 2006; Guo et al. 2018). When there is no significant overstandard in the measured data, the design flood regional composition can be deduced by using this method. To describe the EFRC method, we consider a single intermediate sub-catchment, as shown in Figure 1(b), with one upstream site, an intermediate catchment, and a downstream site. The inflow of downstream site Z contains the outflow at upstream site X and the runoff at intermediate catchment Y . The EFRC method assumes that the exceedance probability of the sub-catchments (upstream site X or intermediate catchment Y) is equivalent to that of downstream site Z , whereas the flood volume at the other sub-catchments is obtained by back-calculation with respect to the water balance. In engineering practice, the EFRC method has two forms, as follows.

Form (1): Assuming that the probability of the design flood volume at upstream site and downstream site is P and the probability of the corresponding flood volume at intermediate catchment is C , the corresponding flood volume at intermediate catchment can be expressed as:

$$Y_C = Z_P - X_P \quad (2)$$

where X_P and Z_P are the design flood volumes at upstream site and downstream site with exceedance probability P , respectively; and Y_C is the corresponding design flood

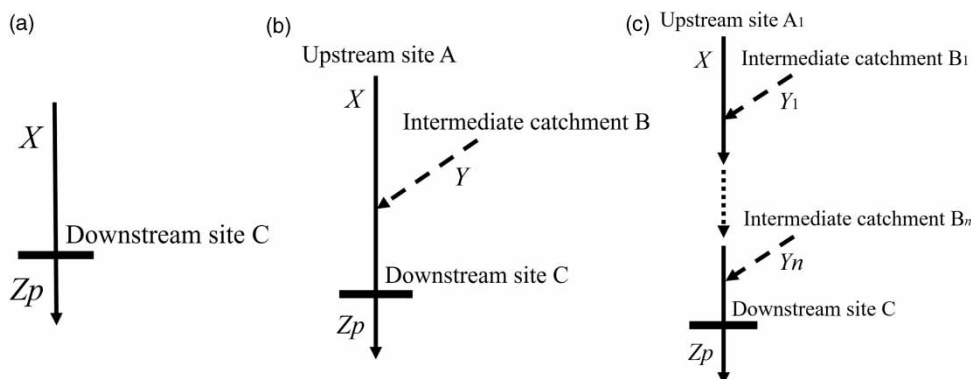


Figure 1 | Schematic of the natural river system (a–c show the natural condition, single intermediate catchment, and cascade intermediate catchments, respectively).

volume at intermediate catchment with the exceedance probability C . However, the relationship between P and C , i.e., $P > C$ or $P \leq C$, has not been theoretically proven.

Form (2): Assuming that the probability of design flood volume at intermediate catchment and downstream site is P and the probability of corresponding flood volume at upstream site is C , the corresponding flood volume at upstream site can be expressed as:

$$X_C = Z_P - Y_P \quad (3)$$

where Y_P and Z_P are the P -design probability flood volumes at intermediate catchment and downstream site, respectively, and X_C is the corresponding C -probability flood volume at upstream site. Similarly, whether $P > C$ or $P \leq C$ has not been clarified.

We take the first form of the EFRC method as an example and determine whether $P > C$ or $P \leq C$ via theoretical derivations and MC experiments in the following.

Distribution functions for the hydrological frequency analysis

In this study, the normal, EV1(2) and logistic distributions are used to investigate the relationship between the exceedance probability C of corresponding flood volume at intermediate catchment and exceedance probability P of design floods at upstream site and downstream site via theoretical derivations. For the PE3, LN(2) and GEV distributions, which are transcendental functions that are difficult to calculate, the relation between C and P is investigated via MC experiments, as shown in the following sections, and then extended to the general distribution.

Normal distribution

The probability density function (PDF) of a normally distributed variable $X \sim N(\mu, \sigma^2)$ is expressed by Equation (4).

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty \quad (4)$$

where μ and σ are the mean and the standard deviation of the random variable X , respectively. When the variable $x = \mu$, the PDF reaches a maximum value of $f(x) = \frac{1}{\sqrt{2\pi}\sigma}$, and the exceedance probability of flood $P(X \geq \mu) = 50\%$.

EV1(2) distribution

The PDF of the EV1(2) distribution (also known as Gumbel distribution) is given by Equation (5).

$$f(x) = \frac{1}{a} e^{-\left(\frac{x-m}{a} + e^{-\frac{x-m}{a}}\right)} \quad (5)$$

The variable x takes values in the range $-\infty < x < +\infty$. The distribution function of x is given by Equation (6).

$$F(x) = P(X \geq x) = 1 - e^{-e^{-\frac{x-m}{a}}} \quad (6)$$

where a and m are the scale parameter and the location parameter of the random variable X , respectively, which are estimated by the MOM (Equation (7)).

$$\begin{cases} \hat{a} = \frac{\sqrt{6}}{\pi} \sigma \\ \hat{m} = \mu - \frac{\sqrt{6}\gamma}{\pi} \sigma \end{cases} \quad (7)$$

where μ and σ are the mean and the standard deviation of the random variable X , respectively, and $\gamma = 0.57722$ is the Euler–Mascheroni constant.

Thus, Equation (6) can be transformed into Equation (8).

$$F(x) = P(X \geq x) = 1 - e^{-e^{-\frac{x - \left(\mu - \frac{\sqrt{6}\gamma}{\pi}\sigma\right)}{\frac{\sqrt{6}}{\pi}\sigma}}} \quad (8)$$

According to Equation (8), the exceedance probability of flood $P(X \geq \mu)$ is equal to 42.96%.

Logistic distribution

The PDF and distribution function of the logistic distribution are expressed in Equations (9) and (10), respectively.

$$f(x) = \frac{1}{a} e^{-\frac{x-m}{a}} \left(1 + e^{-\frac{x-m}{a}}\right)^{-2}, \quad -\infty < x < +\infty \quad (9)$$

$$F(x) = P(X \geq x) = 1 - \left(1 + e^{-\frac{x-m}{a}}\right)^{-1}, \quad -\infty < x < +\infty \quad (10)$$

where a and m are the scale parameter and the location parameter of the random variable X , respectively, which are estimated by the MOM in Equation (11).

$$\begin{cases} \hat{a} = \frac{\sqrt{3}}{\pi} \sigma \\ \hat{m} = \mu \end{cases} \quad (11)$$

where μ and σ are the mean and the standard deviation of the random variable X , respectively.

Thus, Equation (10) can be transformed into Equation (12),

$$F(x) = P(X \geq x) = 1 - \left(1 + e^{-\frac{x-\mu}{\frac{\sqrt{3}}{\pi}\sigma}}\right)^{-1} \quad (12)$$

According to Equation (12), the exceedance probability of flood $P(X \geq \mu)$ is equal to 50%.

PE3 distribution

The PDF of the PE3 distribution is expressed by Equation (13).

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} (x - a_0)^{\alpha-1} e^{-\beta(x-a_0)} \quad (13)$$

where α , β , and a_0 are the shape, scale, and location parameters, respectively. These parameters are related to the expected value (EX), coefficient of variation (Cv), and coefficient of skewness (Cs) of the distribution through the following equations.

$$\begin{cases} \hat{\alpha} = \frac{4}{Cs^2} \\ \hat{\beta} = \frac{2}{EX Cv Cs} \\ \hat{a}_0 = EX \left(1 - \frac{2Cv}{Cs}\right) \end{cases} \quad (14)$$

LN(2) distribution

The PDF of the LN(2) distribution is given by Equation (15),

$$f(x) = \frac{1}{x\sigma_y\sqrt{2\pi}} \exp\left\{-\frac{[\ln x - \mu_y]^2}{2\sigma_y^2}\right\} \quad (15)$$

where μ_y and σ_y are the mean and standard deviation of the natural logarithms of x , respectively. The variable $\ln(x)$ can be standardized, as shown in Equation (16).

$$u = \frac{\ln(x) - \mu_y}{\sigma_y} \quad (16)$$

where the standard normal variable u is obtained with the PDF given in Equation (4).

These parameters are related to the EX and Cv of the distribution through the following equations.

$$\begin{cases} \hat{\sigma}_y = \sqrt{\ln(Cv^2 + 1)} \\ \hat{\mu}_y = \ln(EX) - \frac{\hat{\sigma}_y^2}{2} \end{cases} \quad (17)$$

GEV distribution

The PDF of the GEV distribution is given by Equation (18) as follows,

$$f(x) = \frac{1}{\alpha} \left[1 - k\left(\frac{x-u}{\alpha}\right)\right]^{1/k-1} e^{-[1-k(\frac{x-u}{\alpha})]^{1/k}} \quad (18)$$

The range of the variable x depends on the sign of the parameter k . When k is negative (type II extreme value distribution EV2(3), $Cs > 1.1396$), the variable x can take on values in the range $u + \alpha/k < x < \infty$, which makes the variable suitable for flood frequency analysis. However, when k is positive (type III extreme value distribution EV3(3); $Cs < 1.1396$), x develops an upper bound and takes on values in the range $-\infty < x < u + \alpha/k$, which may not be acceptable for analyzing floods unless there is sufficient evidence that such an upper bound does exist. When $k = 0$ ($Cs = 1.1396$), the GEV distribution reduces to the EV1(2) distribution. Different value of Cs corresponds to a different approximation of \hat{k} .

$$\begin{cases} \hat{k} = 0.2858221 - 0.357983Cs + 0.116659Cs^2 - 0.022725Cs^3 \\ \quad + 0.002604Cs^4 - 0.000161Cs^5 + 0.000004Cs^6 \\ \quad \quad \quad k < 0 (1.14 < Cs < 10) \\ \hat{k} = 0.277648 - 0.322016Cs + 0.060278Cs^2 + 0.016759Cs^3 \\ \quad - 0.005873Cs^4 - 0.00244Cs^5 - 0.000050Cs^6 \\ \quad \quad \quad k > 0 (-2 < Cs < 1.14) \\ \hat{k} = -0.50405 - 0.00861Cs + 0.015497Cs^2 + 0.005613Cs^3 \\ \quad + 0.00087Cs^4 - 0.000065Cs^5 \\ \quad \quad \quad k < 0 (-10 < Cs < 0) \end{cases} \quad (19)$$

Once \hat{k} is estimated, it is substituted into Equation (20) to find $\hat{\alpha}$ and $\hat{\mu}$ based on their sample estimates of EX and Cv .

$$\begin{cases} \hat{\alpha} = [EX^2 Cv^2 \hat{k}^2 / \{\Gamma(1 + 2\hat{k}) - \Gamma^2(1 + \hat{k})\}]^{1/2} \\ \hat{\mu} = EX - \frac{\hat{\alpha}}{\hat{k}} [1 - \Gamma(1 + \hat{k})] \end{cases} \quad (20)$$

DERIVATION AND RESULTS

Derivation for normal distribution

Assume that the flood volume at upstream site, intermediate catchment, and downstream site is subject to normal distributions of $X \sim N(\mu_x, \sigma_x^2)$, $Y \sim N(\mu_y, \sigma_y^2)$, and $Z \sim N(\mu_z, \sigma_z^2)$, independently.

Let $\mu_y = m\mu_x$, ($m > 0$) and $\sigma_y^2 = n^2\sigma_x^2$, ($n > 0$). As shown in Figure 1(b), the downstream flood volume Z equals the sum of upstream flood volume X and intermediate flood volume Y . Thus, $\mu_z = \mu_x + \mu_y = \mu_x + m\mu_x$ and $\sigma_z^2 = \sigma_x^2 + \sigma_y^2 = \sigma_x^2 + n^2\sigma_x^2$, which means that the distribution parameters of the random variables Y and Z can be represented by the statistics of X , $Y \sim N(m\mu_x, n^2\sigma_x^2)$, and $Z \sim N(\mu_x + m\mu_x, \sigma_x^2 + n^2\sigma_x^2)$.

With additional calculations, we can identify the exceedance probability of the volumes of upstream, intermediate, and downstream floods. The calculation of upstream flood volume X is used as an example, and the exceedance probability in the context of $\{X \geq x\}$ is obtained as follows.

$$P_1(X \geq x) = \frac{1}{\sqrt{2\pi}\sigma_x} \int_x^{+\infty} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} dx \quad (21)$$

Let $\frac{x-\mu_x}{\sqrt{2}\sigma_x} = t$,

$x = \sqrt{2}\sigma_x \cdot t + \mu_x$, $dx = \sqrt{2}\sigma_x \cdot dt$,

where $x \in (x, +\infty)$ and $t \in (\frac{x-\mu_x}{\sqrt{2}\sigma_x}, +\infty)$.

Thus, Equation (21) can be transformed into Equation (22).

$$P_1(X \geq x) = \frac{1}{\sqrt{2\pi}\sigma_x} \int_{\frac{x-\mu_x}{\sqrt{2}\sigma_x}}^{+\infty} e^{-t^2} \sqrt{2}\sigma_x dt = \frac{1}{\sqrt{\pi}} \int_{\frac{x-\mu_x}{\sqrt{2}\sigma_x}}^{+\infty} e^{-t^2} dt \quad (22)$$

In the same way, the exceedance probabilities of the volumes of intermediate flood and downstream flood are obtained.

$$P_2(Y \geq y) = \frac{1}{\sqrt{\pi}} \int_{\frac{y-\mu_y}{\sqrt{2}\sigma_y}}^{+\infty} e^{-t^2} dt = \frac{1}{\sqrt{\pi}} \int_{\frac{y-m\mu_x}{\sqrt{2n}\sigma_x}}^{+\infty} e^{-t^2} dt \quad (23)$$

$$P_3(Z \geq z) = \frac{1}{\sqrt{\pi}} \int_{\frac{z-\mu_z}{\sqrt{2}\sigma_z}}^{+\infty} e^{-t^2} dt = \frac{1}{\sqrt{\pi}} \int_{\frac{z-(1+m)\mu_x}{\sqrt{2(1+n^2)}\sigma_x}}^{+\infty} e^{-t^2} dt \quad (24)$$

For the EFRC method, the exceedance probabilities of X and Z are both $P = P_1(X \geq x) = P_3(Z \geq z)$, and the exceedance probability of Y calculated by Equation (1) is $C = P_2(Y \geq y)$. Thus, Equation (22) is equal to Equation (24).

$$\frac{1}{\sqrt{\pi}} \int_{\frac{x-\mu_x}{\sqrt{2}\sigma_x}}^{+\infty} e^{-t^2} dt = \frac{1}{\sqrt{\pi}} \int_{\frac{z-(1+m)\mu_x}{\sqrt{2(1+n^2)}\sigma_x}}^{+\infty} e^{-t^2} dt \quad (25)$$

$$\frac{x - \mu_x}{\sqrt{2}\sigma_x} = \frac{z - (1 + m)\mu_x}{\sqrt{2(1 + n^2)}\sigma_x} \quad (26)$$

$$z = \sqrt{1 + n^2}(x - \mu_x) + (1 + m)\mu_x \quad (27)$$

Similarly, we can compare Equations (22) and (23) to obtain the relationship between P and C , which indicates that Equations (22) and (23) are different in their integral lower bound. We define Δ as the difference between the lower bounds.

$$\Delta = \frac{x - \mu_x}{\sqrt{2}\sigma_x} - \frac{y - \mu_y}{\sqrt{2}\sigma_y} \quad (28)$$

By plugging Equations (2) and (27) into Equation (28), Equation (28) can be easily transformed into the expression in Equation (29).

$$\Delta = (x - \mu_x) \frac{(1 + n) - \sqrt{1 + n^2}}{\sqrt{2n}\sigma_x} \quad (29)$$

where $\frac{(1 + n) - \sqrt{1 + n^2}}{\sqrt{2n}\sigma_x} > 0$.

1. If $x < \mu_x$, i.e., $P_1(X \geq x) > 50\%$, $\Delta < 0$, $P_1(X \geq x) = P_3(Z \geq z) > P_2(Y \geq y)$, then P is larger than C .

- 2. If $x = \mu_x$, i.e., $P_1(X \geq x) = 50\%$,
 $\Delta = 0$, $P_1(X \geq x) = P_3(Z \geq z) = P_2(Y \geq y)$,
then P equals C .
- 3. If $x > \mu_x$, i.e., $P_1(X \geq x) < 50\%$,
 $\Delta > 0$, $P_1(X \geq x) = P_3(Z \geq z) < P_2(Y \geq y)$,
then P is less than C .

Overall, in the context of using the first EFRC method for normal distribution floods, the relationship between C and P depends on whether or not P is larger than 50%. Thus, for a design flood (volume) whose exceedance probability is generally less than 50%, C is greater than P ; however, for low-flow condition whose exceedance probability is generally larger than 50%, C is less than P . For example, if the return periods of the volumes of upstream and downstream floods are both 1,000 years, then the return period of intermediate flood volume could be 100 years, and if the guarantee rates of upstream and downstream flows are both 90%, then the guarantee rate of intermediate flow volume could be 70%.

In the context of other EFRC methods, similar conclusions can be drawn for a normal distribution flood.

Derivation for EV1(2) distribution

Assume that the floods at upstream site, intermediate catchment, and downstream site are subject to the EV1(2) distribution as follows,

$$f(x) = \frac{1}{\frac{\sqrt{6}}{\pi} \sigma_x} e^{-\left[\frac{x - \left(\mu_x - \frac{\sqrt{6}\gamma}{\pi} \sigma_x \right)}{\frac{\sqrt{6}}{\pi} \sigma_x} + e^{-\frac{x - \left(\mu_x - \frac{\sqrt{6}\gamma}{\pi} \sigma_x \right)}{\frac{\sqrt{6}}{\pi} \sigma_x}} \right]} \tag{30}$$

$$f(y) = \frac{1}{\frac{\sqrt{6}}{\pi} \sigma_y} e^{-\left[\frac{y - \left(\mu_y - \frac{\sqrt{6}\gamma}{\pi} \sigma_y \right)}{\frac{\sqrt{6}}{\pi} \sigma_y} + e^{-\frac{y - \left(\mu_y - \frac{\sqrt{6}\gamma}{\pi} \sigma_y \right)}{\frac{\sqrt{6}}{\pi} \sigma_y}} \right]} \tag{31}$$

$$f(z) = \frac{1}{\frac{\sqrt{6}}{\pi} \sigma_z} e^{-\left[\frac{z - \left(\mu_z - \frac{\sqrt{6}\gamma}{\pi} \sigma_z \right)}{\frac{\sqrt{6}}{\pi} \sigma_z} + e^{-\frac{z - \left(\mu_z - \frac{\sqrt{6}\gamma}{\pi} \sigma_z \right)}{\frac{\sqrt{6}}{\pi} \sigma_z}} \right]} \tag{32}$$

Let $\mu_y = m\mu_x$ ($m > 0$) and $\sigma_y^2 = n^2\sigma_x^2$ ($n > 0$). As shown in Figure 1(b), the downstream flood volume Z equals the sum of upstream flood volume X and intermediate flood volume Y . Then $\mu_z = \mu_x + \mu_y = \mu_x + m\mu_x$ and $\sigma_z^2 = \sigma_x^2 + \sigma_y^2 = \sigma_x^2 + n\sigma_x^2$; thus, the distribution parameters of the random variables Y and Z can be represented by the X statistic.

$$f(y) = \frac{1}{\frac{\sqrt{6}}{\pi} n\sigma_x} e^{-\left[\frac{y - \left(m\mu_x - \frac{\sqrt{6}\gamma}{\pi} n\sigma_x \right)}{\frac{\sqrt{6}}{\pi} n\sigma_x} + e^{-\frac{y - \left(m\mu_x - \frac{\sqrt{6}\gamma}{\pi} n\sigma_x \right)}{\frac{\sqrt{6}}{\pi} n\sigma_x}} \right]} \tag{33}$$

$$f(z) = \frac{1}{\frac{\sqrt{6}}{\pi} (\sqrt{1+n^2})\sigma_x} e^{-\left[\frac{z - \left((1+m)\mu_x - \frac{\sqrt{6}\gamma}{\pi} (\sqrt{1+n^2})\sigma_x \right)}{\frac{\sqrt{6}}{\pi} (\sqrt{1+n^2})\sigma_x} + e^{-\frac{z - \left((1+m)\mu_x - \frac{\sqrt{6}\gamma}{\pi} (\sqrt{1+n^2})\sigma_x \right)}{\frac{\sqrt{6}}{\pi} (\sqrt{1+n^2})\sigma_x}} \right]} \tag{34}$$

With additional calculations, we can identify the exceedance probability of flood at upstream site, intermediate catchment, and downstream site. Using the calculation of upstream flood volume X as an example, the exceedance probability of upstream flood volume is obtained as follows.

$$P_1(X \geq x) = \frac{1}{\frac{\sqrt{6}}{\pi} \sigma_x} \int_x^{+\infty} e^{-\left[\frac{x - \left(\mu_x - \frac{\sqrt{6}\gamma}{\pi} \sigma_x \right)}{\frac{\sqrt{6}}{\pi} \sigma_x} + e^{-\frac{x - \left(\mu_x - \frac{\sqrt{6}\gamma}{\pi} \sigma_x \right)}{\frac{\sqrt{6}}{\pi} \sigma_x}} \right]} dx \tag{35}$$

Let $\frac{x - \left(\mu_x - \frac{\sqrt{6}\gamma}{\pi} \sigma_x \right)}{\frac{\sqrt{6}}{\pi} \sigma_x} = t$,
 $x = \frac{\sqrt{6}}{\pi} \sigma_x \cdot t + \left(\mu_x - \frac{\sqrt{6}\gamma}{\pi} \sigma_x \right)$, $dx = \frac{\sqrt{6}}{\pi} \sigma_x \cdot dt$,
where $x \in (x, +\infty)$ and $t \in \left(\frac{x - \left(\mu_x - \frac{\sqrt{6}\gamma}{\pi} \sigma_x \right)}{\frac{\sqrt{6}}{\pi} \sigma_x}, +\infty \right)$
Then,

$$P_1(X \geq x) = \int_{\frac{x - \left(\mu_x - \frac{\sqrt{6}\gamma}{\pi} \sigma_x \right)}{\frac{\sqrt{6}}{\pi} \sigma_x}^{+\infty} e^{-(t+e^{-t})} dt \tag{36}$$

Similarly, the exceedance probability of the volumes of intermediate flood and downstream flood are obtained.

$$P_2(Y \geq y) = \int_{\frac{y - (\mu_y - \frac{\sqrt{6}\gamma}{\pi}\sigma_y)}{\frac{\sqrt{6}}{\pi}\sigma_y}}^{+\infty} e^{-(t+e^{-t})} dt$$

$$= \int_{\frac{y - (m\mu_x - \frac{\sqrt{6}\gamma}{\pi}n\sigma_x)}{\frac{\sqrt{6}}{\pi}n\sigma_x}}^{+\infty} e^{-(t+e^{-t})} dt \tag{37}$$

$$P_3(Z \geq z) = \int_{\frac{z - (\mu_z - \frac{\sqrt{6}\gamma}{\pi}\sigma_z)}{\frac{\sqrt{6}}{\pi}\sigma_z}}^{+\infty} e^{-(t+e^{-t})} dt$$

$$= \int_{\frac{z - [(1+m)\mu_x - \frac{\sqrt{6}\gamma}{\pi}\sqrt{1+n^2}\sigma_x]}{\frac{\sqrt{6}}{\pi}\sqrt{1+n^2}\sigma_x}}^{+\infty} e^{-(t+e^{-t})} dt \tag{38}$$

For the EFRC method, the exceedance probabilities of the volumes of upstream and downstream floods are both $P = P_1(X \geq x) = P_3(Z \geq z)$ and that of the corresponding flood volume calculated by Equation (2) is $C = P_2(Y \geq y)$. Thus, Equation (36) is equal to Equation (38),

$$\int_{\frac{x - (\mu_x - \frac{\sqrt{6}\gamma}{\pi}\sigma_x)}{\frac{\sqrt{6}}{\pi}\sigma_x}}^{+\infty} e^{-(t+e^{-t})} dt = \int_{\frac{z - [(1+m)\mu_x - \frac{\sqrt{6}\gamma}{\pi}\sqrt{1+n^2}\sigma_x]}{\frac{\sqrt{6}}{\pi}\sqrt{1+n^2}\sigma_x}}^{+\infty} e^{-(t+e^{-t})} dt \tag{39}$$

$$\frac{x - (\mu_x - \frac{\sqrt{6}\gamma}{\pi}\sigma_x)}{\frac{\sqrt{6}}{\pi}\sigma_x} = \frac{z - [(1+m)\mu_x - \frac{\sqrt{6}\gamma}{\pi}\sqrt{1+n^2}\sigma_x]}{\frac{\sqrt{6}}{\pi}\sqrt{1+n^2}\sigma_x} \tag{40}$$

$$z = \sqrt{1+n^2} \left[x - \left(\mu_x - \frac{\sqrt{6}\gamma}{\pi}\sigma_x \right) \right]$$

$$+ \left[(1+m)\mu_x - \frac{\sqrt{6}\gamma}{\pi}\sqrt{1+n^2}\sigma_x \right] \tag{41}$$

Similarly, we can compare Equations (36) and (37) to obtain the relationship between P and C . After comparing the equations, we found that Equations (36) and (37) are different in their integral lower bounds. We defined Δ as the difference between these bounds.

$$\Delta = \frac{x - (\mu_x - \frac{\sqrt{6}\gamma}{\pi}\sigma_x)}{\frac{\sqrt{6}}{\pi}\sigma_x} - \frac{y - (m\mu_x - \frac{\sqrt{6}\gamma}{\pi}n\sigma_x)}{\frac{\sqrt{6}}{\pi}n\sigma_x} \tag{42}$$

By plugging Equations (2) and (41) into Equation (42), Equation (42) can be easily transformed into Equation (43).

$$\Delta = (x - \mu_x) \frac{[(1+n) - \sqrt{1+n^2}]}{\frac{\sqrt{6}}{\pi}n\sigma_x} \tag{43}$$

where $\frac{[(1+n) - \sqrt{1+n^2}]}{\frac{\sqrt{6}}{\pi}n\sigma_x} > 0$.

1. If $x < \mu_x$, i.e., $P_1(X \geq x) > 42.96\%$, $\Delta < 0$, $P_1(X \geq x) = P_3(Z \geq z) > P_2(Y \geq y)$, then P is larger than C .
2. If $x = \mu_x$, i.e., $P_1(X \geq x) = 42.96\%$, $\Delta = 0$, $P_1(X \geq x) = P_3(Z \geq z) = P_2(Y \geq y)$, then P equals C .
3. If $x > \mu_x$, i.e., $P_1(X \geq x) < 42.96\%$, $\Delta > 0$, $P_1(X \geq x) = P_3(Z \geq z) < P_2(Y \geq y)$, then P is less than C .

Thus, when using the first EFRC method for the EV1(2) distribution flood, the relationship between C and P depends on whether P is larger than 42.96%. For a design flood (volume) whose exceedance probability is generally less than 42.96%, C is larger than P , whereas for the design low flow whose exceedance probability is generally larger than 42.96%, C is less than P .

Derivation for logistic distribution

Assume that the floods at upstream site, intermediate catchment, and downstream site are independently subject to a logistic distribution.

$$f(x) = \frac{1}{\frac{\sqrt{3}}{\pi}\sigma_x} e^{-\frac{x-\mu_x}{\frac{\sqrt{3}}{\pi}\sigma_x}} \left(1 + e^{-\frac{x-\mu_x}{\frac{\sqrt{3}}{\pi}\sigma_x}} \right)^{-2} \tag{44}$$

$$f(y) = \frac{1}{\frac{\sqrt{3}}{\pi}\sigma_y} e^{-\frac{y-\mu_y}{\frac{\sqrt{3}}{\pi}\sigma_y}} \left(1 + e^{-\frac{y-\mu_y}{\frac{\sqrt{3}}{\pi}\sigma_y}} \right)^{-2} \tag{45}$$

$$f(z) = \frac{1}{\frac{\sqrt{3}}{\pi}\sigma_z} e^{-\frac{z-\mu_z}{\frac{\sqrt{3}}{\pi}\sigma_z}} \left(1 + e^{-\frac{z-\mu_z}{\frac{\sqrt{3}}{\pi}\sigma_z}} \right)^{-2} \tag{46}$$

Let $\mu_y = m\mu_x$, ($m > 0$) and $\sigma_y^2 = n^2\sigma_x^2$, ($n > 0$). As shown in Figure 1(b), the downstream flood volume Z equals the sum of upstream flood volume X and intermediate flood volume Y . Then, $\mu_z = \mu_x + \mu_y = \mu_x + m\mu_x$ and $\sigma_z^2 = \sigma_x^2 + \sigma_y^2 = \sigma_x^2 + n\sigma_x^2$; thus, the distribution parameters of random variables Y and Z can be represented by the statistics of X .

$$f(y) = \frac{1}{\frac{\sqrt{3}}{\pi}n\sigma_y} e^{-\frac{y-m\mu_y}{\frac{\sqrt{3}}{\pi}n\sigma_y}} \left(1 + e^{-\frac{y-m\mu_y}{\frac{\sqrt{3}}{\pi}n\sigma_y}} \right)^{-2} \tag{47}$$

$$f(z) = \frac{1}{\frac{\sqrt{3}}{\pi}(\sqrt{1+n^2})\sigma_z} e^{-\frac{z-(1+m)\mu_x}{\frac{\sqrt{3}}{\pi}(\sqrt{1+n^2})\sigma_z}} \left[1 + e^{-\frac{z-(1+m)\mu_x}{\frac{\sqrt{3}}{\pi}(\sqrt{1+n^2})\sigma_z}} \right]^{-2} \tag{48}$$

With additional calculations, we can identify the exceedance probability of the volumes of upstream, intermediate, and downstream floods. The calculation of upstream flood volume X is used as an example, and the exceedance probability of upstream flood volume is obtained as follows.

$$P_1(X \geq x) = \frac{1}{\frac{\sqrt{3}}{\pi}\sigma_x} \int_x^{+\infty} e^{-\frac{x-\mu_x}{\frac{\sqrt{3}}{\pi}\sigma_x}} \left[1 + e^{-\frac{x-\mu_x}{\frac{\sqrt{3}}{\pi}\sigma_x}} \right]^{-2} dx \tag{49}$$

Let $\frac{x-\mu_x}{\frac{\sqrt{3}}{\pi}\sigma_x} = t$,

$x = \frac{\sqrt{3}}{\pi}\sigma_x \cdot t + \mu_x$, $dx = \frac{\sqrt{3}}{\pi}\sigma_x \cdot dt$,

where $x \in (x, +\infty)$ and $t \in \left(\frac{x-\mu_x}{\frac{\sqrt{3}}{\pi}\sigma_x}, +\infty \right)$

Then,

$$P_1(X \geq x) = \int_{\frac{x-\mu_x}{\frac{\sqrt{3}}{\pi}\sigma_x}^{+\infty} e^{-t}(1+e^{-t})^{-2} dt \tag{50}$$

Similarly, the exceedance probabilities of the volumes of intermediate and downstream floods are obtained.

$$P_2(Y \geq y) = \int_{\frac{y-m\mu_y}{\frac{\sqrt{3}}{\pi}n\sigma_y}^{+\infty} e^{-t}(1+e^{-t})^{-2} dt = \int_{\frac{y-(m\mu_x)}{\frac{\sqrt{3}}{\pi}n\sigma_x}^{+\infty} e^{-t}(1+e^{-t})^{-2} dt \tag{51}$$

$$P_3(Z \geq z) = \int_{\frac{z-(1+m)\mu_x}{\frac{\sqrt{3}}{\pi}(\sqrt{1+n^2})\sigma_x}^{+\infty} e^{-t}(1+e^{-t})^{-2} dt = \int_{\frac{z-[(1+m)\mu_x]}{\frac{\sqrt{3}}{\pi}(\sqrt{1+n^2})\sigma_x}^{+\infty} e^{-t}(1+e^{-t})^{-2} dt \tag{52}$$

For the EFRC method, $P = P_1(X \geq x) = P_3(Z \geq z)$ and $C = P_2(Y \geq y)$. Thus, Equation (50) is equal to Equation (52).

$$\int_{\frac{x-\mu_x}{\frac{\sqrt{3}}{\pi}\sigma_x}^{+\infty} e^{-t}(1+e^{-t})^{-2} dt = \int_{\frac{z-[(1+m)\mu_x]}{\frac{\sqrt{3}}{\pi}(\sqrt{1+n^2})\sigma_x}^{+\infty} e^{-t}(1+e^{-t})^{-2} dt \tag{53}$$

$$\frac{x-\mu_x}{\frac{\sqrt{3}}{\pi}\sigma_x} = \frac{z-[(1+m)\mu_x]}{\frac{\sqrt{3}}{\pi}(\sqrt{1+n^2})\sigma_x} \tag{54}$$

$$z = (\sqrt{1+n^2})(x-\mu_x) + (1+m)\mu_x \tag{55}$$

Similarly, we can compare Equations (50) and (51) to obtain the relationship between P and C . After comparing the two equations, we found that Equations (50) and (51) are different in terms of their integral lower bounds. We defined Δ as the difference between the bounds.

$$\Delta = \frac{x-\mu_x}{\frac{\sqrt{3}}{\pi}\sigma_x} - \frac{y-(m\mu_x)}{\frac{\sqrt{3}}{\pi}n\sigma_x} \tag{56}$$

By plugging Equations (2) and (55) into Equation (56), Equation (56) can be easily transformed into Equation (57).

$$\Delta = (x-\mu_x) \frac{[(1+n)-\sqrt{1+n^2}]}{\frac{\sqrt{3}}{\pi}n\sigma_x} \tag{57}$$

where $\frac{[(1+n)-\sqrt{1+n^2}]}{\frac{\sqrt{3}}{\pi}n\sigma_x} > 0$.

1. If $x < \mu_x$, i.e., $P_1(X \geq x) > 50\%$, $\Delta < 0$, $P_1(X \geq x) = P_3(Z \geq z) > P_2(Y \geq y)$, then P is larger than C .
2. If $x = \mu_x$, i.e., $P_1(X \geq x) = 50\%$, $\Delta = 0$, $P_1(X \geq x) = P_3(Z \geq z) = P_2(Y \geq y)$, then P is equal to C .
3. If $x > \mu_x$, i.e., $P_1(X \geq x) < 50\%$, $\Delta > 0$, $P_1(X \geq x) = P_3(Z \geq z) < P_2(Y \geq y)$, then P is less than C .

When using the first EFRC method for flood regional composition with logistic distributions, the relationship between C and P depends on whether P is larger than

50% or not. Thus, for a design flood (volume), C is greater than P , whereas for a design low flow, C is less than P .

Experiment analysis for PE3 distribution

For a PE3 distribution as shown in Equation (13), the multivariate PDF is a transcendental equation. Therefore, a theoretical form of derivation for the relationship between P and C may not be obtained. Then, MC experiments are used to infer whether $P > C$ or $P \leq C$.

MC experiments are performed to produce and utilize random numbers to solve complex calculations (Denny & Yevjevich 1972; Christiane 2009; Xing et al. 2019). Specifically, for the problem to be solved, a random variable is constructed, a large number of random numbers are sampled according to the variable's numerical characteristics, and the corresponding parameter values are calculated from these samples as the solution to the problem. Figure 2 shows a flow chart that clearly illustrates the procedure of MC experiment. In the figure, the steps of MC experiment are as follows.

Step 1: Randomly generate upstream flood volume X_{rng} and downstream flood volume Z_{rng} according to their numerical characteristics (EX , EZ , C_s , and C_v). Z_{rng} minus X_{rng} is internal flood volume Y .

Step 2: Repeat Step 1 100,000 times. Obtain 100,000 random internal floods Y at the same time.

Step 3: Plot the hydrologic frequency analysis curve of Y through 100,000 random numbers.

Step 4: Plot hydrologic frequency analysis curves of Z and X through their numerical characteristics.

Step 5: Based on design probability P , design floods Z_p and X_p are obtained from the hydrologic frequency analysis curves of Z and X , respectively.

Step 6: Z_p minus X_p is the corresponding flood volume Y_c .

Step 7: Based on the Y_c , exceedance probability C is obtained from the hydrological frequency analysis curve of Y .

Step 8: Repeat Steps 5–7 for design probability P of 0.01, 0.1, 1, 2, 10, 20, 25, 30, 35, 40, 45, 50, 60, 75, 80, 90, 95, 97, 99, and 99.9%. Compare the relationship between P and C . Find critical point P_0 .

For the PE3 distribution, we performed 100,000 random trials for the flood regional composition. In each of these trials, we produced the random upstream flood volume X and downstream flood volume Z according to their numerical characteristics (EX , C_v , and C_s), and we identified the corresponding flood volume at intermediate catchment Y based on the difference between the volumes of downstream and upstream floods. Then, we obtained the cumulative

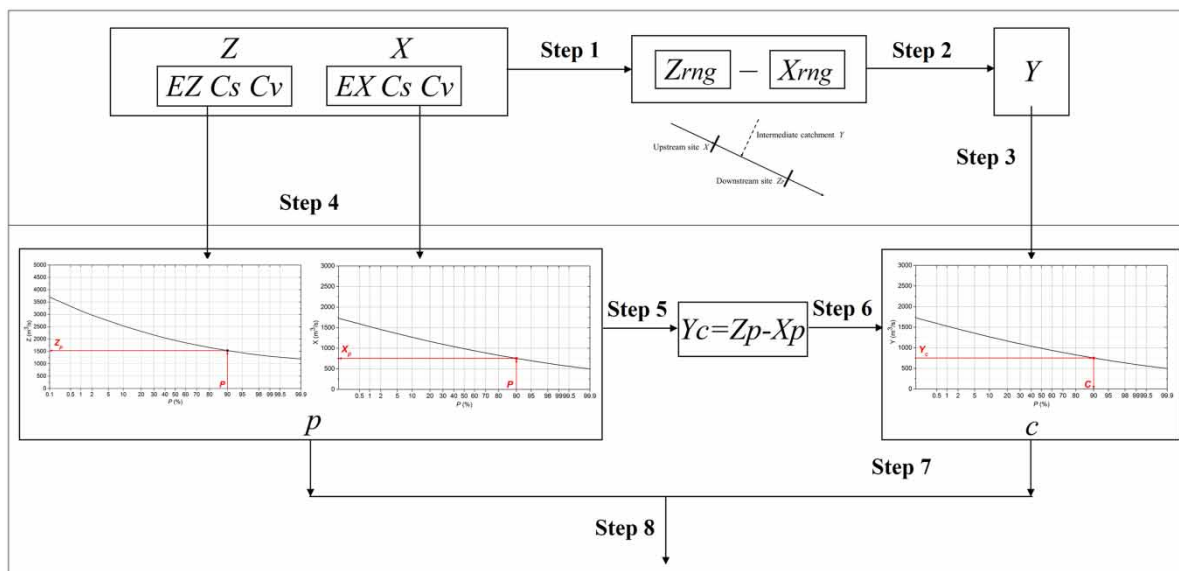


Figure 2 | The flow chart of MC experiment.

distribution of Y through 100,000 corresponding intermediate floods volumes. Considering the universality and extensive suitability of the EFRC method, six regional composition schemes were constructed as representatives. To ensure that the downstream flood volume was always greater than the upstream flood volume, the EX of upstream site was 1,000 and the EZ of downstream site was 2,000. For Cv and Cs , both upstream and downstream presented values of $Cv = 0.2, Cs = 0.8, Cv = 0.4, Cs = 1.6, Cv = 0.5, Cs = 2.0, Cv = 0.7, Cs = 2.8, Cv = 1.0, Cs = 4.0,$ and $Cv = 1.5, Cs = 6.0$. Therefore, six regional composition schemes are obtained in total. Design probabilities P of each regional composition scheme are 0.01, 0.1, 1, 2, 10, 20, 25, 30, 35, 40, 45, 50, 60, 75, 80, 90, 95, 97, 99, and 99.9%.

A specific regional composition scheme is used as an example, such as when the parameters of downstream floods (volume) are $EZ = 2,000, Cv = 0.4,$ and $Cs/Cv = 4$ and those of upstream floods (volume) are $EX = 1,000, Cv = 0.4,$ and $Cs/Cv = 4$. The results of the statistical

experiments are shown in Table 1, and the relationship between the exceedance probabilities of C and P is presented in Figure 3 (the same as the regional composition scheme 2 in Figure 4).

As illustrated in Figure 3, a critical point P_0 is presented, which is the exceedance probability corresponding to the intersection of two lines. When the exceedance probability of design flood P is greater than the critical point P_0 , then the exceedance probability of corresponding flood at intermediate catchment C is less than P ($C < P$). When P is exactly equal to P_0 , then $C = P$. When P is smaller than P_0 , then $C > P$. All of these results show that in the context of using the first EFRC method for PE3 distribution, the relationship between C and P depends on the design probabilities of the volumes of upstream and downstream floods. This conclusion is consistent with that drawn from the theoretical derivation for normal, EV1(2) and logistic distributions. More specifically, the P_0 of normal, EV1(2) and logistic distributions are 50, 42.96, and 50%,

Table 1 | Typical composition results of the Pearson type III distributions based on Monte Carlo experiments

	Exceedance probability $P/\%$	Downstream flood volume $EZ = 2,000$ $Cv = 0.4$ $Cs/Cv = 4$ Zp	Upstream flood volume $EX = 1,000$ $Cv = 0.4$ $Cs/Cv = 4$ Xp	Corresponding intermediate flood volume $Yc = Zp - Xp$ Yc	Exceedance probability of intermediate flood volume $C/\%$	$(C - P)/C$
Design exceedance probability	0.01	8,337	3,806	4,530	0.3	0.97
	0.1	6,224	3,170	3,055	3	0.96
	1	4,682	2,350	2,332	8	0.87
	2	4,213	2,117	2,096	11	0.81
	10	3,056	1,538	1,518	22	0.56
	20	2,531	1,277	1,254	31	0.36
	25	2,361	1,188	1,173	34	0.27
	30	2,216	1,115	1,101	37	0.20
	35	2,094	1,052	1,042	40	0.13
	40	1,982	995	987	43	0.07
Design guarantee rate	45	1,883	945	939	45	0.00
	50	1,791	900	891	48	- 0.05
	60	1,630	818	812	52	- 0.15
	75	1,414	708	705	58	- 0.28
	80	1,345	674	671	61	- 0.32
	90	1,205	602	603	65	- 0.39
	95	1,126	563	564	67	- 0.41
	97	1,088	544	544	68	- 0.42
	99	1,043	521	522	70	- 0.42
	99.9	1,010	505	505	71	- 0.41

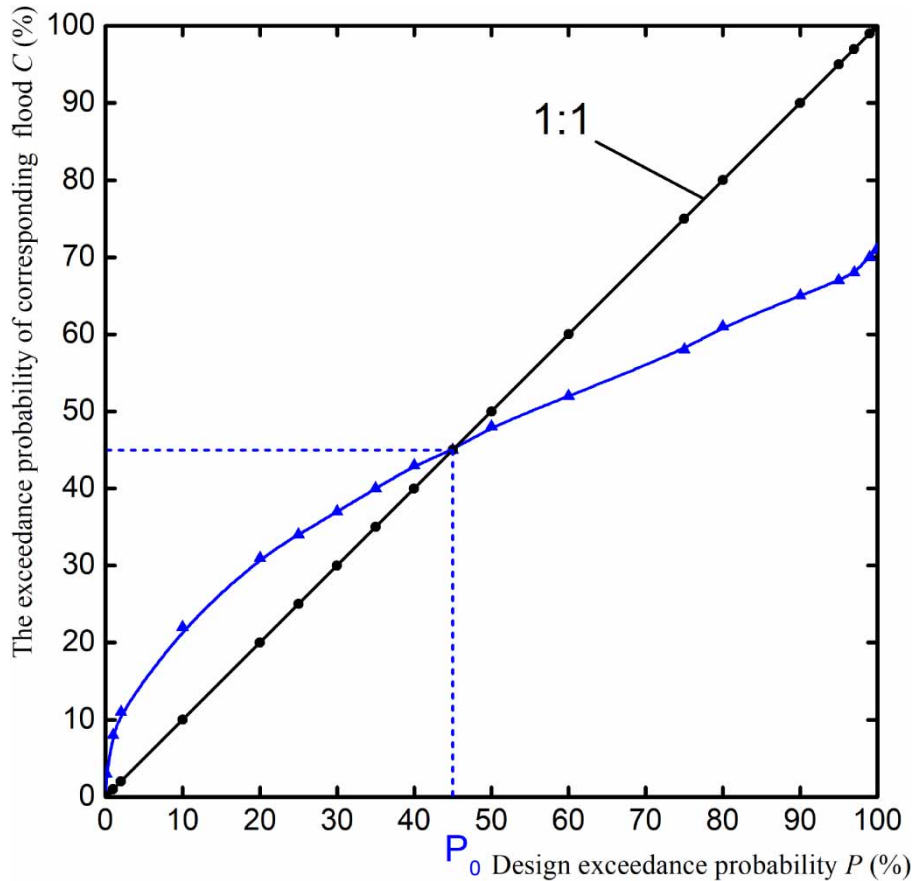


Figure 3 | The result of a statistical experiment for the Pearson type III distributions.

respectively, whereas the P_0 of PE3 distribution is not fixed and may be greater than, less than, or equal to 50%.

P_0 of the typical regional composition in Figure 3 is 45%, which is less than 50%. Similarly, P_0 of the other five regional compositions could be obtained. All of the results are shown in Figure 4 and Table 2, the maximum of P_0 is 50% and the minimum of P_0 is 30%; namely, the critical point P_0 of PE3 distribution is between 30 and 50%. In addition, because of the various C_s values in the experiments, the range of P_0 of PE3 distribution is broad (from 30 to 50%). When the C_s is larger than 2.0, the curve of the PDF of PE3 distribution looks like an 'L' shape, and when C_s is less than 2.0, the curve is shaped like a bell. The 'L' shape means that the value of a variable has the greatest likelihood near its minimum value, which does not conform to the hydrological phenomenon. For the hydrological variables, the chances of occurrence of extremely large value and extremely small values are very

low, while the chances of occurrence of intermediate values are higher; that is, the curve should be shaped like a bell. Therefore, the PE3 distribution of $C_s > 2.0$ is not suitable in hydrology. However, in most cases, the value of C_s is not restricted in practice. Therefore, we constructed six regional composition schemes with six different C_s values, including $C_s > 2.0$. If we do not consider the case of $C_s > 2.0$, the range of critical point P_0 of the PE3 distribution will be between 45 and 50% (as shown in Table 2).

Taking into account the actual situation of design flood regional composition, the design probabilities of the volumes of upstream and downstream floods are generally small, such as $P = 0.01$, 0.1, or 1%; and these values are far less than 30%. Therefore, the exceedance probability of corresponding intermediate flood volume is greater than that of the volumes of upstream and downstream floods. For example, if a 100-year design flood ($P = 1\%$) occurs in an upstream site and a downstream site, then the

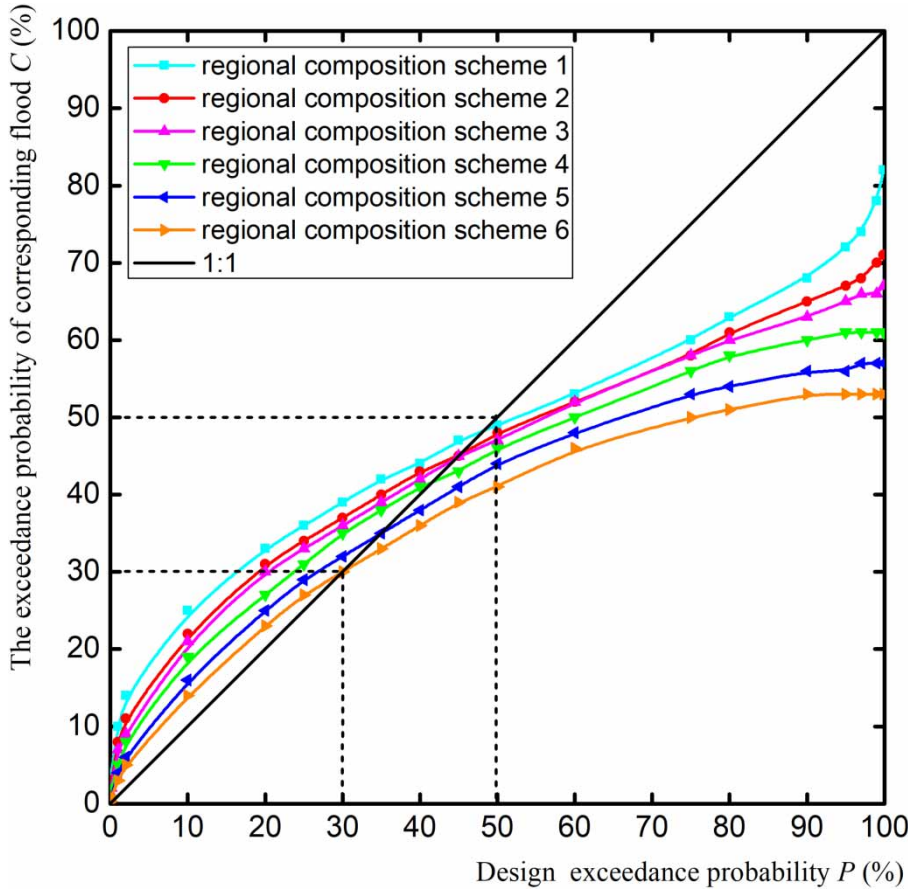


Figure 4 | The results of all statistical experiments for the Pearson type III distributions.

Table 2 | The results of P_0 of six regional composition schemes for three distributions

Regional composition scheme	1	2	3	4	5	6
PE3	$EZ = 2,000$	$EZ = 2,000$	$EZ = 2,000$	$EZ = 2,000$	$EZ = 2,000$	$EZ = 2,000$
	$Cv = 0.2$	$Cv = 0.4$	$Cv = 0.5$	$Cv = 0.7$	$Cv = 1.0$	$Cv = 1.5$
	$Cs = 0.8$	$Cs = 1.6$	$Cs = 2.0$	$Cs = 2.8$	$Cs = 4.0$	$Cs = 6.0$
	$EX = 1,000$	$EX = 1,000$	$EX = 1,000$	$EX = 1,000$	$EX = 1,000$	$EX = 1,000$
	$Cv = 0.2$	$Cv = 0.4$	$Cv = 0.5$	$Cv = 0.7$	$Cv = 1.0$	$Cv = 1.5$
	$Cs = 0.8$	$Cs = 1.6$	$Cs = 2.0$	$Cs = 2.8$	$Cs = 4.0$	$Cs = 6.0$
	0.50	0.45	0.45	0.40	0.35	0.30
LN(2)	$EZ = 2,000$	$EZ = 2,000$	$EZ = 2,000$	$EZ = 2,000$	$EZ = 2,000$	$EZ = 2,000$
	$Cv = 0.2$	$Cv = 0.4$	$Cv = 0.5$	$Cv = 0.7$	$Cv = 1.0$	$Cv = 1.5$
	$EX = 1,000$	$EX = 1,000$	$EX = 1,000$	$EX = 1,000$	$EX = 1,000$	$EX = 1,000$
	$Cv = 0.2$	$Cv = 0.4$	$Cv = 0.5$	$Cv = 0.7$	$Cv = 1.0$	$Cv = 1.5$
	0.50	0.50	0.45	0.45	0.45	0.45
GEV	$EZ = 2,000$	$EZ = 2,000$	$EZ = 2,000$	$EZ = 2,000$	$EZ = 2,000$	$EZ = 2,000$
	$Cv = 0.2$	$Cv = 0.4$	$Cv = 0.5$	$Cv = 0.7$	$Cv = 1.0$	$Cv = 1.5$
	$Cs = 0.8$	$Cs = 1.6$	$Cs = 2.0$	$Cs = 4.5$	$Cs = 9.0$	$Cs = 10.0$
	$EX = 1,000$	$EX = 1,000$	$EX = 1,000$	$EX = 1,000$	$EX = 1,000$	$EX = 1,000$
	$Cv = 0.2$	$Cv = 0.4$	$Cv = 0.5$	$Cv = 0.7$	$Cv = 1.0$	$Cv = 1.5$
	$Cs = 0.8$	$Cs = 1.6$	$Cs = 2.0$	$Cs = 4.5$	$Cs = 9.0$	$Cs = 10.0$
	0.50	0.45	0.45	0.45	0.45	0.45

corresponding flood volume at intermediate catchment will be less than the 100-year intermediate flood volume. For low-flow regional compositions, the design exceedance probability (or design guarantee rate) of upstream and downstream design flows is generally greater than 50%, such as $P = 75, 90,$ or 99% . Therefore, the exceedance probability of corresponding intermediate flow volume is less than that of upstream and downstream flows. For example, if a 90% guarantee-rate design flow occurs in upstream site and downstream site, then the guarantee rate of corresponding flow at intermediate catchment will be less than 90%.

Experiment analysis for LN(2) distribution

For the flood regional composition of LN(2) distribution (Equation (15)), we performed 100,000 randomized trials,

which is similar to the experiment analysis of the PE3 distribution. Six regional composition schemes were constructed as representatives. For the downstream site, EZ is 2,000, and for upstream site, EX is 1,000. The Cv values of both upstream and downstream floods (volume) are 0.2, 0.4, 0.5, 0.7, 1.0, and 1.5.

We then used a specific regional composition scheme as an example, such as one in which the parameters of downstream flood (volume) are $EZ = 2,000$ and $Cv = 0.7$, and the parameters of upstream flood (volume) are $EX = 1,000$ and $Cv = 0.7$. The results of the statistical experiment and the relationship between the exceedance probabilities of C and P are shown in Figure 5 (the regional composition scheme 4). For this scheme, the critical point P_0 of the LN(2) distribution is 45%, which is less than 50%.

Similarly, the critical point P_0 of the other five regional compositions could be obtained. All of the results are shown

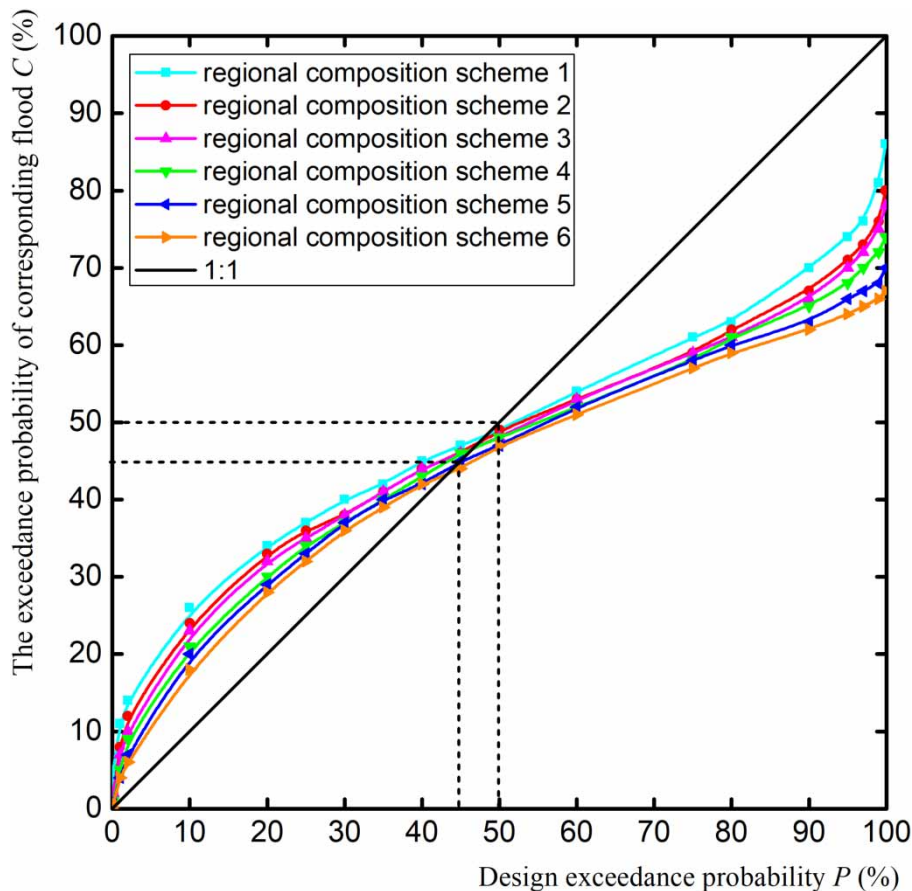


Figure 5 | The results of all statistical experiments for the two-parameter lognormal distributions.

in Figure 5 and Table 2. The critical point P_0 of LN(2) distribution is between 45 and 50%.

Therefore, for the design flood regional composition of LN(2) distribution, the exceedance probability of corresponding intermediate flood volume is greater than that of the volumes of upstream and downstream floods. For low-flow regional composition, the exceedance probabilities of corresponding intermediate flow volume are less than the design probabilities (or design guarantee rates) of upstream and downstream flows.

Experiment analysis for GEV distribution

For GEV distribution (Equation (18)), we also performed 100,000 randomized trials. In the six regional composition schemes, the EZ of downstream site was 2,000 and that of upstream site was 1,000. For the Cv and Cs , both upstream

and downstream presented values of $Cv = 0.2, Cs = 0.8, Cv = 0.4, Cs = 1.6, Cv = 0.5, Cs = 2.0, Cv = 0.7, Cs = 4.5, Cv = 1.0, Cs = 9.0,$ and $Cv = 1.5, Cs = 10$.

A specific regional composition scheme is used as an example, such as one in which the parameters of downstream floods (volume) are $EZ = 2,000, Cs = 0.8,$ and $Cv = 0.2,$ and the parameters of upstream floods (volume) are $EX = 1,000, Cs = 0.8,$ and $Cv = 0.2.$ The results of the statistical experiment and the relationship between the exceedance probabilities of C and P are shown in Figure 6 (the regional composition scheme 1). For this scheme, the critical point P_0 of a typical GEV distribution is approximately 50%, which is similar to the values of the PE3 and LN(2) distributions.

In the same way, the critical point P_0 of the other five regional compositions could be obtained. All of the results are shown in Figure 6 and Table 2, the critical point P_0 of the GEV distribution is between 45 and 50%.

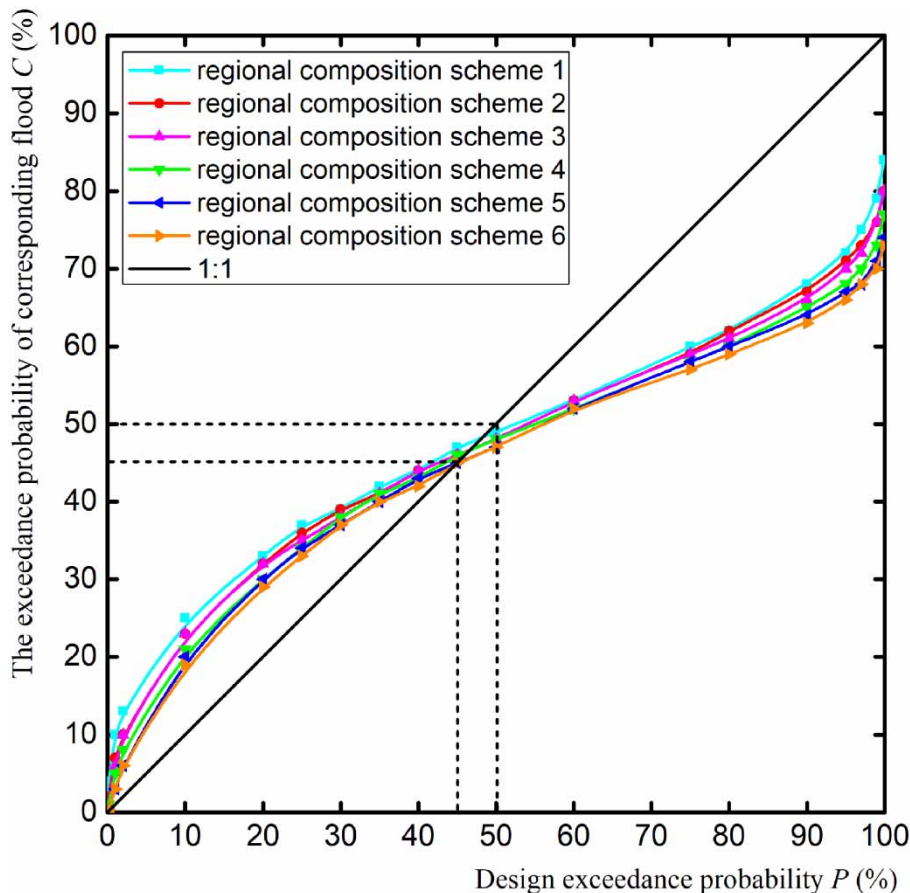


Figure 6 | The results of all statistical experiments for the generalized extreme value distributions.

Table 3 | The results of P_0 of six regional composition schemes for three distributions

Distribution	Normal	EV1(2)	Logistic	PE3	LN(2)	GEV
Critical point P_0 (%)	50	42.96	50	30–50	45–50	45–50

In the actual design flood regional composition, the intermediate exceedance probability is greater than the design exceedance probability. For the design low-flow regional composition, the exceedance probability is less than the design exceedance probability (or design guarantee rate).

Table 3 presents all the results of derivation and experiment analysis. In summary, for the EFRC method with different flood distribution types, the relationship between the intermediate exceedance probability C and design exceedance probability P can be determined through theoretical derivations or statistical experiments to achieve consistent conclusions: For design flood regional composition, the exceedance probability of the corresponding intermediate flood volume is greater than the design probabilities of the volumes of upstream and downstream floods. For example, if a 100-year design flood (volume) occurs in upstream site and downstream site, then the return period of corresponding flood volume at intermediate catchment is less than 100 years. For low-flow conditions, the exceedance probability of the corresponding intermediate flow volume is less than the design probabilities of upstream and downstream flows. Thus, if a 90% guarantee-rate design flood (volume) occurs in upstream site and downstream site, the guarantee rate of corresponding flood volume at intermediate catchment is less than 90%.

CONCLUSIONS

In the EFRC method, which is commonly used to resolve the regional composition of the design flood volume at various sub-catchments in natural conditions, the exceedance probability (or return period) of the estimated corresponding flood volume is unknown. This study performs theoretical derivations and MC experiments to investigate the relationship between the probability of the corresponding flood

volume and that of the design flood volume at the dam site. The following conclusions are obtained.

1. Critical probability value P_0 exists in the EFRC method. When the exceedance probability of downstream design flood volume P is greater than P_0 , the exceedance probability of corresponding flood volume C is less than P , i.e., ($C < P$); however, when P is equal to P_0 , $C = P$, and when P is less than P_0 , $C > P$. The value of P_0 is related to the distribution function of hydrological variables. For normal distribution, EV1(2) distribution and logistic distribution, P_0 equals 50, 42.96, and 50%, respectively. For PE3 distribution, LN(2) distribution and GEV distribution, P_0 is not fixed. For the PE3 distribution, P_0 ranges from approximately 30–50%, whereas for both LN (2) and GEV distributions, P_0 ranges from approximately 45 to 50% based on the MC experiments.
2. In terms of a design flood event, the design exceedance probability P is generally far less than 30% (e.g., $P = 0.01, 0.1$, or 1%); thus, the corresponding exceedance probability C is greater than P . For example, if 100-year design floods ($P = 1\%$, i.e., return period $T = 100$) occur in upstream site and downstream site, the return period of corresponding flood volume at intermediate catchment is less than that of a 100-year event. Therefore, in practical application, the upstream and downstream catchments are very similar in terms of response time, while the intermediate catchment is very small compared to the others, that is the intermediate catchment volume is not as ‘critical’ as those of the other sub-catchments for large T (return period). Besides, when the flood control target at downstream site has been set, the flood control standard of intermediate catchment or upstream site is not as high as that of downstream site, which leads to the prevention of the excessive capacity of flood diversion and peak cutting at upstream reservoirs.
3. In terms of a design low-flow event, the design exceedance probability (or design guarantee rate) P is generally greater than 60% (e.g., $P = 75, 90$, or 99%); thus, the corresponding exceedance probability C is less than P . For example, if 90% guarantee-rate design flows occur in upstream site and downstream site, the guarantee rate of corresponding low-flow at intermediate

catchment is less than 90%, meaning a lower guarantee rate standard at intermediate catchment.

- The above conclusions also apply to other regional composition analyses with the EFRC method. For example, if the exceedance probability of design floods at intermediate catchment and downstream site are equivalent, then the corresponding flood volume at upstream site can be calculated. In addition, for other statistical distribution models, the critical probability P_0 would be different, and the corresponding conclusions would be different, although the method provided in this study can be used for reference.

DATA AVAILABILITY

The data in this study were randomly generated through statistical experiments.

ACKNOWLEDGEMENTS

This study was supported by the Key Special Project of the National Key Research and Development Program of China (2016YFC0402709, 2016YFC0402706), the Major Program of the National Natural Science Foundation of China (41730750), and the National Natural Science Foundation of China (51709073). The authors extend their sincere thanks to all who were involved in this paper.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

- Badreldin, G. H. H., Isameldin, A. A., Li, J. & Feng, P. 2012 At site and regional frequency analysis for Sudan annual rainfall by using the L-moments and nonlinear regression techniques. *International Journal of Engineering Research and Development* **3** (6), 13–19.
- Boughton, W. & Droop, O. 2003 Continuous simulation for design flood estimation – a review. *Environmental Modelling and Software* **18** (4), 309–318. [https://doi.org/10.1016/s1364-8152\(03\)00004-5](https://doi.org/10.1016/s1364-8152(03)00004-5).
- Christiane, L. 2009 *Monte Carlo and Quasi-Monte Carlo Sampling*. Springer, New York.
- Denny, J. L. & Yevjevich, V. 1972 Probability and statistics in hydrology. *Journal of the American Statistical Association* **68** (343), 755. <https://doi.org/10.2307/2284828>.
- Ding, J., Song, D. & Yang, R. 1989 Further research on application of probability weighted moments in estimating parameters of the Pearson type three distribution. *Journal of Hydrology* **110**, 239–257. [https://doi.org/10.1016/0022-1694\(89\)90190-X](https://doi.org/10.1016/0022-1694(89)90190-X).
- Greenwood, J. A. 1979 Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research* **15**. <https://doi.org/10.1029/WR015i005p01049>
- Guo, S., Muhammad, R., Liu, Z., Xiong, F. & Yin, J. 2018 Design flood estimation methods for cascade reservoirs based on copulas. *Water* **10** (5), 560. <https://doi.org/10.3390/w10050560>.
- Hosking, J. R. M. 1986 *The Theory of Probability Weighted Moments*. IBM Research Report, IBM.
- Hosking, J. R. M. 1990 L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society* **52** (1), 105–124. <https://doi.org/10.2307/2345653>.
- Hu, Y., Liang, Z., Chen, X., Liu, Y., Wang, H., Yang, J., Wang, J. & Li, B. 2017 Estimation of design flood using EWT and ENE metrics and uncertainty analysis under non-stationary conditions. *Stochastic Environmental Research and Risk Assessment* (24), 1–10. <https://doi.org/10.1007/s00477-017-1404-1>.
- Hu, Y., Liang, Z., Singh, V. P., Zhang, X., Wang, J., Li, B. & Wang, H. 2018 Concept of equivalent reliability for estimating the design flood under non-stationary conditions. *Water Resources Management* **32** (3), 1–15. <https://doi.org/10.1007/s11269-017-1851-y>.
- Kendall, M. G. & Stuart, A. 1979 *The Advanced Theory of Statistics*. Charles Griffin, London.
- Kirby, W. 1974 Algebraic boundedness of sample statistics. *Water Resources Research* **10** (2), 220–222. <https://doi.org/10.1029/WR010i002p00220>.
- Liang, Z., Hu, Y., Li, B. & Yu, Z. 2014 A modified weighted function method for parameter estimation of Pearson type three distribution. *Water Resources Research* **50** (4), 3216–3228. <https://doi.org/10.1002/2013WR013653>.
- Liang, Z., Huang, H., Cheng, L., Hu, Y., Yang, J. & Tang, T. 2016 Safety assessment for dams of the cascade reservoirs system of Lancang River in extreme situations. *Stochastic Environmental Research and Risk Assessment* **31** (9), 2459–2469. <https://doi.org/10.1007/s00477-016-1331-6>.
- Lu, B., Gu, H., Xie, Z., Liu, J., Ma, L. & Lu, W. 2012 Stochastic simulation for determining the design flood of cascade

- reservoir systems. *Hydrology Research* **43**, 54–63. <https://doi.org/10.2166/nh.2011.002>.
- Maidment, D. R. 1992 *Handbook of hydrology*. *Earth-Science Reviews* **24**, 227–229. https://doi.org/10.1007/978-3-642-41714-6_80111.
- Ministry of Water Resources (MWR) 2006 *Guidelines for Calculating Design Flood of Water Resources and Hydropower Projects*. Chinese Water Resources and Hydropower Press, Beijing. (In Chinese).
- Nijssen, D., Schumann, A., Pahlow, M. & Klein, B. 2009 *Planning of technical flood retention measures in large river basins under consideration of imprecise probabilities of multivariate hydrological loads*. *Natural Hazards and Earth System Sciences* **9**, 1349–1363. <https://doi.org/10.5194/nhess-9-1349-2009>.
- Ponce, V. M. 1989 *Engineering Hydrology: Principles and Practices*. Prentice Hall, Upper Saddle River.
- Rao, A. R. & Hamed, K. H. 2000 *Flood Frequency Analysis*. CRC Press LLC, Boca Raton, CA.
- Wallis, J. R., Matalas, N. C. & Slack, J. R. 1974 *Just a moment!*. *Water Resources Research* **10** (2), 211–219. <https://doi.org/10.1029/WR010i002p00211>.
- Wang, J., Liang, Z., Hu, Y. & Wang, D. 2015 *Modified weighted function method with the incorporation of historical floods into systematic sample for parameter estimation of Pearson type three distribution*. *Journal of Hydrology* **527** (3), 958–966. <https://doi.org/10.1016/j.jhydrol.2015.05.023>.
- Xiao, Y., Guo, S., Liu, P., Yan, B. & Chen, L. 2009 *Design flood hydrograph based on multicharacteristic synthesis index method*. *Journal of Hydrologic Engineering* **14** (12), 1359–1364. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2009\)14:12\(1359\)](https://doi.org/10.1061/(ASCE)1084-0699(2009)14:12(1359)).
- Xing, Z., Zhang, H., Ji, Y., Xinglong, G., Fu, Q. & Li, H. 2019 *Markov chain Monte Carlo based on adaptive metropolis algorithm applied in combined model to deal with the uncertainty of weights of single models*. *Water Science and Technology: Water Supply* **19** (4), 1129–1136.
- Xiong, L., Du, T., Xu, C. Y., Guo, S., Jiang, C. & Gippel, C. J. 2015 *Non-stationary annual maximum flood frequency analysis using the norming constants method to consider non-stationarity in the annual daily flow series*. *Water Resources Management* **29** (10), 3615–3633. <https://doi.org/10.1007/s11269-015-1019-6>.
- Yue, S., Ouarda, T. B., Bobée, B., Legendre, P. & Bruneau, P. 2002 *Approach for describing statistical properties of flood hydrograph*. *Journal of Hydrologic Engineering* **7** (2), 147–153. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2002\)7:2\(147\)](https://doi.org/10.1061/(ASCE)1084-0699(2002)7:2(147)).
- Zeng, X., Wang, D. & Wu, J. 2012 *Sensitivity analysis of the probability distribution of groundwater level series based on information entropy*. *Stochastic Environmental Research and Risk Assessment* **26** (3), 345–356. <https://doi.org/10.1007/s00477-012-0556-2>.

First received 26 February 2020; accepted in revised form 2 July 2020. Available online 31 July 2020