Assessing the impacts of reservoirs on downstream flood frequency by coupling the effect of scheduling-related multivariate rainfall into an indicator of reservoir effects

Bin Xiong[1], Lihua Xiong[1*], Jun Xia[1], Chong-Yu Xu[1, 3], Cong Jiang[2], Tao Du[4]

1. State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China

2. School of Environmental Studies, China University of Geosciences, Wuhan 430074, China

3. Department of Geosciences, University of Oslo, P.O. Box 1022 Blindern, N-0315 Oslo, Norway

4. Bureau of Hydrology, Changjiang Water Resources Commission, Wuhan 430010, China

** Corresponding author:*

Lihua Xiong, PhD, Professor

State Key Laboratory of Water Resources and Hydropower Engineering Science

Wuhan University, Wuhan 430072, China

E-mail: xionglh@whu.edu.cn

Telephone: +86-13871078660

Fax: +86-27-68773568

Abstract

Many studies have shown that downstream flood regimes have been significantly altered by upstream reservoir operation. Reservoir effects on the downstream flow regime are normally performed by comparing the pre-dam and post-dam frequencies of certain streamflow indicators, such as floods and droughts. In this study, a rainfall-reservoir composite index (RRCI) is developed to precisely quantify reservoir impacts on downstream flood frequency under a framework of a covariate-based nonstationary flood frequency analysis using the Bayesian inference method. The RRCI is derived from a combination of both a reservoir index (RI) for measuring the effects of reservoir storage capacity and a rainfall index. More precisely, the OR-joint exceedance probability (OR-JEP) of certain scheduling-related variables selected out of five variables that describe the multiday antecedent rainfall input (MARI) is used to measure the effects of antecedent rainfall on reservoir operation. Then, the RI-dependent or RRCI-dependent distribution parameters and five distributions, the gamma, Weibull, lognormal, Gumbel, and generalized extreme value, are used to analyze the annual maximum daily flow (AMDF) of the Ankang, Huangjiagang, and Huangzhuang gauging stations of the Hanjiang River, China. A phenomenon is observed that although most of the floods that peak downstream of reservoirs have been reduced in magnitude by upstream reservoirs, some relatively large flood events still have occurred, such as at the Huangzhuang station in 1983. The results of nonstationary flood frequency analysis show that, in comparison to the RI, the RRCI that combines both the RI and the OR-JEP resulted a much better explanation for such phenomena of flood occurrences downstream of reservoirs. A Bayesian inference of the 100-year return level of the AMDF shows that the optimal RRCI-dependent distribution, compared to the RI-dependent one, results in relatively smaller estimated values. However, there exist exceptions due to some low OR-JEP values. In addition, it provides a smaller uncertainty range. This study highlights the necessity of including antecedent rainfall effects, in addition to the effects of reservoir storage capacity, on reservoir operation to assess the reservoir effects on downstream flood frequency. This analysis can provide a more comprehensive approach for downstream flood risk management under the impacts of reservoirs.

Keywords: nonstationary flood frequency analysis; downstream floods; reservoir; antecedent rainfall; Bayesian inference; Hanjiang River

1 Introduction

River floods are generated by various complex nonlinear processes involving physical factors including “hydrological pre-conditions (e.g., soil saturation, snow cover), meteorological conditions (e.g., amount, intensity, and the spatial and temporal distribution of rainfall), runoff generation processes, and river routing (e.g., superposition of flood waves in the main river and its tributaries)” (Wyżga et al., 2016). In general, without reservoirs, the downstream flood extremes of most rain-dominated basins are primarily related to extreme rainfall events in the drainage area. However, with reservoirs, the downstream flood regimes should be totally different due to upstream flood control scheduling. In the literature, the significant hydrological alterations caused by reservoirs have been demonstrated in the many areas of the world. Graf (1999) showed that dams have more significant effects on streamflow in America than global climate change. Benito and Thorndycraft (2005) reported various significant changes across the United States in pre- and post-dam hydrologic regimes (e.g., minimum and maximum flows over different durations). Batalla et al. (2004) demonstrated an evident reservoir-induced hydrologic alteration in northeastern Spain. Yang et al. (2008) demonstrated the spatial variability in hydrological regimes alterations caused by the reservoirs in the middle and lower Yellow River in China. Mei et al. (2015) found that the Three Gorges Dam, the largest dam in the world, has significantly changed downstream hydrological regimes. In recent years, the cause-effect mechanisms of downstream flood peak reductions were also investigated by some researchers (Ayalew et al., 2013; Ayalew et al., 2015; Volpi et al., 2018). For example, Volpi et al. (2018) suggested that for a single reservoir, the downstream flood peak reduction was primarily dependent on its position along the river, its spillway, and its storage capacity based on a parsimonious instantaneous unit hydrograph-based model. These studies have revealed that it is crucial to assess the impacts of reservoirs on downstream flood regimes for the success of downstream flood risk management.

Flood frequency analysis is the most common technique used by hydrologists to gain knowledge of flood regimes. In conventional or stationary frequency analyses, a basic hypothesis is that hydrologic time series maintains stationarity, i.e., “free of trends, shifts, or periodicity (cyclicity)” (Salas, 1993).

However, in many cases, observations of changes in flood regimes have demonstrated that this strict assumption is invalid (Kwon et al., 2008; Milly et al., 2008). Nonstationarity in downstream flood regimes of dams makes frequency analyses more complicated. Actually, the frequency of downstream floods of dams is closely related to upstream flood operations. In recent years, there have been many attempts to link flood generating mechanisms and reservoir operations to the frequency of downstream floods (Gilroy and Mccuen, 2012; Goel et al., 1997; Lee et al., 2017; Liang et al., 2017; Su and Chen, 2018; Yan et al., 2017).

Previous studies have meaningfully increased the knowledge about reservoir-induced nonstationarity of downstream hydrological extreme frequencies (Ayalew et al., 2013; López and Francés, 2013; Liang et al., 2017; Magilligan and Nislow, 2005; Su and Chen, 2018; Wang et al., 2017; Zhang et al., 2015). There are two main approaches to incorporate reservoir effects into flood frequency analyses: the hydrological model simulation approach and the nonstationary frequency modeling approach. In the first approach, the regulated flood time series can be simulated using three model components: the stochastic rainfall generator, the rainfall-runoff model, and the reservoir flood operation module, which includes the reservoir storage capacity, the size of release structures, and the operation rules. The continuous simulation method can explicitly account for the reservoir effects on floods in the hypothetical case. However, it is difficult to apply this approach to a majority of real cases (Volpi et al., 2018) because the simplifying assumptions of this approach are only satisfied in a few of basins with single small reservoirs. Furthermore, even if the basins meet the simplifying assumptions, the detailed information required in this approach is likely unavailable. Thus, our attention is focused on the second method, the nonstationary frequency modeling approach. Nonstationary distribution models have been widely used to deal with the nonstationarity of extreme value series. In nonstationary distribution models, the distribution parameters are expressed as the functions of covariates to determine the conditional distributions of extreme value series. According to extreme value theory, the maxima series can generally be described using the generalized extreme value distribution (GEV). Thus, previous studies (El Adlouni et al., 2007; Ouarda and El‐Adlouni, 2011) have used the nonstationary generalized extreme value distribution to describe the nonstationary maxima series. Scarf (1992) modeled the changes in the location and scale parameters of the GEV over time using the power function relationship. Coles (2001) introduced several time-dependent structures (e.g., trend, quadratic, and change-point) into the location, scale, and shape parameters of the GEV. El Adlouni et al. (2007) provided a general nonstationary GEV model with an improved parameter estimate method. In recent years, "generalized additive models for location, scale, and shape" (GAMLSS) have been widely used in nonstationary hydrological frequency analyses (Du et al., 2015; Jiang et al., 2014; López and Francés, 2013; Rigby and Stasinopoulos, 2005; Villarini et al., 2009). GAMLSS provides various candidate distributions for frequency analysis, such as Weibull, gamma, Gumbel, and lognormal distributions. However, the GEV has been rarely involved in the candidate distributions of GAMLSS. In terms of a parameter estimation method for the nonstationary distribution model, the maximum likelihood (ML) method is the most common parameter estimate method. However, the ML method for a nonstationary distribution model can lead to very high quantile estimator variances when using numerical techniques to solve the likelihood function when using a small sample (El Adlouni et al., 2007). El Adlouni et al. (2007) developed the generalized maximum likelihood (GML) method and demonstrated that the GML method had better performance than the ML method in all their cases. Ouarda and El‐Adlouni (2011) introduced the Bayesian nonstationary frequency analysis. The Bayesian inference can obtain multiple estimates, forming a posterior distribution of model parameters. Thus, the Bayesian method is able to conveniently describe the uncertainty of flood estimates associated with the uncertainty of model parameters.

In the nonstationary frequency modeling approach, a dimensionless reservoir index (RI) was proposed by López and Francés (2013) as an indicator of reservoir effects, and it generally is used as a covariate for the expression of the distribution parameters (e.g., location parameter) (Jiang et al., 2014; López and Francés, 2013). Liang et al. (2017) modified the reservoir index by replacing the mean annual runoff in the expression of the RI with the annual runoff. Therefore, the modified reservoir index can reflect the impact of reservoirs on downstream flood extremes under various total inflow conditions each year. However, the precision and accuracy in the quantitative analysis of the reservoir effects on downstream floods need to be further improved. In fact, the effects of reservoirs may be closely related not only to the static reservoir storage capacity but also to the dynamic reservoir operations associated with multiple characteristics, such as the peak, the intensity, and the total volume of the multiday antecedent rainfall input (MARI), not just annual runoff.

Therefore, the aim of the study is to develop an indicator, referred to as the rainfall-reservoir composite index (RRIC), that combines the effects of reservoir storage capacity and the MARI on reservoir operation. This indicator is then used as a covariate to assess the reservoir effects on the downstream flood frequency. The specific objectives of this study are (1) to develop the RRCI; (2) to compare the RRCI with the RI using a covariate-based nonstationary flood frequency analysis; and (3) to obtain the downstream flood estimation and its uncertainty based on the optimal nonstationary distribution using the Bayesian inference.

2 Methods

To quantify the effects of reservoirs on the frequency of the annual maximum daily flow series (AMDF) downstream of reservoirs, a three-step framework (Figure 1), termed the covariate-based flood frequency analysis using the RRIC as a covariate, was established. In this section, the methods of this framework are introduced. First, a reservoir index (RI) is defined by additionally considering the effects of reservoir sediment deposition on the storage capacity. Second, the RRCI is developed by combining the RI and a rainfall index. Next, the C-vine copula model is used to construct and calculate the rainfall index. Finally, the nonstationary distribution models that utilize the Bayesian estimation are clarified.

<Figure 1>

2.1 Reservoir index (RI)

Intuitively, the larger the reservoir capacity relative to the flow of a downstream gauging station, the greater the possible effects of the reservoir on the streamflow regime. To quantify reservoir-induced alterations to the downstream streamflow regime, Batalla et al. (2004) proposed an impounded runoff index (IRI), which is a ratio of reservoir capacity (RC) to (unimpaired) mean annual runoff ($\bar{Q}$) at the gauge station, indicated as $\mathrm{IRI} = \mathrm{RC}/\bar{Q}$. For a single reservoir, the IRI is a good indicator of the extent to which a reservoir alters streamflow. To analyze the effects of a multi-reservoir system on the downsream flood frequency, López and Francés (2013) proposed a dimensionless reservoir index. In this study, we additionally considered the effects of reservoir sediment deposition on the reservoir capacity. In accordance with López and Francés (2013), the reservoir index (RI) for a downstream gauging station is defined as

$$\mathrm{RI} = \sum_{i=1}^{N} \left(\frac{A_i}{A_T} \right) \cdot \left(\frac{\left(1 - \mathrm{LR}_i\right) \cdot \mathrm{RC}_i}{\bar{Q}} \right), \tag{1}$$

where N is the total number of reservoirs upstream of the gauge station; A_i is the total basin area upstream of the i-th reservoir; A_T is the total basin area upstream of the gauge station; RC_i is the total storage capacity of the i-th reservoir; and LR_i is the loss rate (%) of RC_i due to the sediment deposition (Appendix A). Equation (1) indicates that for a reservoir system consisting of small- and middle-sized reservoirs, the RI for the downstream gauging station is generally less than one. However, for a system with some large reservoirs, such as multi-year regulating storage reservoirs, the RI of the downstream gauging station near this system may be close to one or higher.

2.2 Rainfall-reservoir composite index (RRCI)

In addition to the reservoir capacity, the multiday antecedent rainfall input (MARI), which is an event of continuous multi-day multivariate rainfall that forms the inflow event that will be regulated by the reservoir system to become the downstream extreme flow, is a key constraint for scheduling the reservoir system. In this study, to add the antecedent rainfall effects into the new indicator of reservoir effects, five variables were used to describe the MARI: the maximum M (the maximum daily rainfall in the MARI); the intensity I (the mean daily rainfall in the MARI); the volume V (the total daily rainfall in the MARI); the timing T (the end time of MARI during that year); and the distance L (the distance between the rainfall center and the outlet). The reason that M, I, V, and L were selected is because these variables will determine the peak, the total volume, and the peak appearance time of an inflow event. The variable, T, is utilized to capture information regarding the remaining storage capacity, due to staged operation strategies during flood season used in some reservoirs. For the operation strategy that consists of increasing the flood limit water level in stages, it is expected that if the timing of the MARI is near the end of the flood season, the downstream AMDF will be less affected by reservoirs. This is because of the lesser remaining capacity during this period. The MARI variables that are selected to construct the new indicator are hereafter referred to as the scheduling-related MARI variables (denoted as $X_1, X_2, ..., X_d$). The extraction procedure of the MARI is detailed in section 3.2.

A new index is proposed in this study called the rainfall-reservoir composite index (RRIC) to more comprehensively assess the effects of reservoirs on floods by incorporating the effects of the MARI. This index is defined as

$$\mathrm{RRCI} = \begin{cases} \left(P_{\mathrm{MARI}}^{\vee}\left(\bigcup_{i=1}^{d} \left(X_i > x_i \right) \right) \right)^{(1/\mathrm{RI}-1)}, 0 < \mathrm{RI} \leq 1 \\ \mathrm{RI}, \mathrm{RI} > 1 \end{cases}, \tag{2}$$

where $P_{\mathrm{MARI}}^{\vee}$ is the OR-joint exceedance probability (OR-JEP); that is the probability that any one of the given set of values ($x_1, x_2, ..., x_d$) for the scheduling-related MARI variables will be exceeded. Here, the OR-JEP acts as a rainfall index for measuring the MARI effects. The lower this probability, the greater effects on reservoir operation the MARI has. Then, it is expected that downstream floods could possibly obtain relatively large values, and vice versa. Figure 2 illustrates the relationship in Equation (2), which shows that the RRCI is conditional on both the OR-JEP and the RI. Equation (2) can then be expressed as

$$\mathrm{RRCI} = \begin{cases} \left(1 - F\left(x_1, x_2, ...x_d\right)\right)^{(1/\mathrm{RI}-1)}, 0 < \mathrm{RI} \leq 1 \\ \mathrm{RI}, \mathrm{RI} > 1 \end{cases}, \tag{3}$$

where $F(\cdot)$ is the cumulative distribution function (CDF) that determines the dependence relationship of the variables. The expectation of the RRCI is as follows:

$$E(\mathrm{RRCI}) = \int_{\mathbb{R}^d} \left(1 - F(x_1, x_2, \ldots x_d)\right)^{(1/\mathrm{RI}-1)} dF(x_1, x_2, \ldots x_d) = \mathrm{RI} \ . \tag{4}$$

In addition, for the OR case, the following is true:

$$P_{\mathrm{MARI}}^{\vee}\left(\bigcup_{i=1}^{d}(X_i > x_i)\right) \geq P_{\mathrm{MARI}}^{\vee}(X_i > x_i) \ . \tag{5}$$

Equations (3) and (5) indicate that, in addition to the RI, the RRCI is related to the number and the dependence relationship of the scheduling-related MARI variables. To obtain a reasonable RRCI, the unrelated MARI variables should not be incorporated. In this study, the number of MARI variables that were incorporated was no more than four to avoid a “dimension disaster” in modeling their dependence. To select the scheduling-related MARI variables, a three-step selection procedure was used that included the following. (1) Selecting four variables from the five MARI variables by testing the significance of the Pearson correlation between the MARI variables and the AMDF. (2) Calculating the RRCI for all possible subsets of the four variables using the *d*-dimensional ($d = 1, 2, 3, 4$) copulas. Then finally (3) identifying the variables by using the highest rank correlation coefficient between the RRCI and the AMDF. The construction method of the *d*-dimensional ($d = 2, 3, 4$) distribution $F(x_1, x_2, \ldots, x_d)$ is described in the following subsection.

<Figure 2>

2.3 C-vine Copula model

In this subsection, a c-vine Copula model for the construction of the continuous *d*-dimensional distribution $F(x_1, x_2, ..., x_d)$ is clarified. The Sklar's theorem (Sklar, 1959) showed that for a continuous *d*-dimensional distribution, the one-dimensional marginals and dependence structure can be separated, and the dependence can be represented using a copula formula as follows:

$$F(x_1, x_2, ...x_d | \boldsymbol{\theta}) = C(u_1, u_2, ..., u_d | \boldsymbol{\theta}_c), u_i = F_{X_i}(x_i | \boldsymbol{\theta}_i) \ , \tag{6}$$

where u_i is the univariate marginal distribution of X_i; $C(\cdot)$ is the copula function; $\boldsymbol{\theta}_c$ is the copula parameter vector; $\boldsymbol{\theta}_i$ is the parameter vector of the *i*-th marginal distribution; and $\boldsymbol{\theta} = (\boldsymbol{\theta}_c, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_d)$ is the parameter vector of the entire *n*-dimensional distribution. Thus, the construction of $F(x_1, x_2, ...x_d)$ can be separated into two steps: first is the modeling of the univariate marginals; and second is the modeling of the dependence structure. For the first step, the empirical distribution is used as the univariate marginal distributions, and the change-points of the variables are tested using the Pettitt test (Pettitt, 1979). Then, if there are any, the marginal and the change-point will be addressed using the estimation method (Xiong et al., 2015). Then, for the second step, the copula construction for the dependence modeling is based on the pair-copula construction method, which has been widely used in previous research (Aas et al., 2009; Xiong et al., 2015). According to Aas et al. (2009), the joint density function $f(x_1, x_2, ..., x_d)$ is written as

$$f\left(x_1, x_2, \ldots, x_d \left|\boldsymbol{\theta}\right.\right) = c_{1\ldots n}\left(u_1, u_2, \ldots, u_d \left|\boldsymbol{\theta}_c\right.\right)\prod_{i=1}^{d} f_{X_i}\left(x_i \left|\boldsymbol{\theta}_i\right.\right), u_i = F_{X_i}\left(x_i \left|\boldsymbol{\theta}_i\right.\right) \quad . \tag{7}$$

The n-dimensional copula density $c_{1\ldots d}\left(u_1, u_2, \ldots, u_d\right)$, which can be decomposed into $d(d-1)/2$ bivariate copulas, corresponding to a c-vine structure, is given by

$$c_{1\ldots d}\left(u_1, u_2, \ldots, u_d \left|\boldsymbol{\theta}_c\right.\right) = \prod_{j=1}^{d-1}\prod_{i=1}^{d-j} c_{j,i+j|1,\ldots,j-1}\left(F\left(u_j \left|u_1, \ldots, u_{j-1}\right.\right), F\left(u_{i+j} \left|u_1, \ldots, u_{j-1}\right.\right)\left|\boldsymbol{\theta}_{j,i|1,\ldots,j-1}\right.\right), \tag{8}$$

where $c_{j,i+j|1,\ldots,j-1}$ is the density function of a bivariate pair copula, and $\boldsymbol{\theta}_{j,i|1,\ldots,j-1}$ is a parameter vector of the corresponding bivariate pair copula. Therefore, the marginal conditional distribution is

$$\begin{aligned} & F\left(u_{i+j} \left|u_1, \ldots, u_{j-1}\right.\right) = \\ & \frac{\partial C_{i+j,j-1|1,\ldots,j-2}\left(F\left(u_{i+j} \left|u_1, \ldots, u_{j-2}\right.\right), F\left(u_{j-1} \left|u_1, \ldots, u_{j-2}\right.\right)\left|\boldsymbol{\theta}_{i+j,j-1|u_1,\ldots,u_{j-2}}\right.\right)}{\partial F\left(u_{j-1} \left|u_1, \ldots, u_{j-2}\right.\right)}, \\ & j = 2, \ldots, d-1;\ i = 0, \ldots, n-j \end{aligned} \tag{9}$$

where $C_{i+j,j-1|1,\ldots,j-2}$ is a bivariate copula distribution function. The maximum dimensionality covered in this study was four. Thus for a four-dimensional copula (of which the decomposition is shown in Figure 3), the general expression of Equation (8) is

$$\begin{aligned} & c_{1234}\left(u_1, u_2, u_3, u_4 \left|\boldsymbol{\theta}_c\right.\right) = c_{12}\left(u_1, u_2 \left|\boldsymbol{\theta}_{12}\right.\right) c_{13}\left(u_1, u_3 \left|\boldsymbol{\theta}_{13}\right.\right) c_{14}\left(u_1, u_4 \left|\boldsymbol{\theta}_{14}\right.\right) \cdot \\ & c_{23|1}\left(F\left(u_2 \left|u_1\right.\right), F\left(u_2 \left|u_1\right.\right)\left|\boldsymbol{\theta}_{23|1}\right.\right) c_{24|1}\left(F\left(u_2 \left|u_1\right.\right), F\left(u_4 \left|u_1\right.\right)\left|\boldsymbol{\theta}_{24|1}\right.\right) \cdot \\ & c_{34|12}\left(F\left(u_3 \left|u_1, u_2\right.\right), F\left(u_4 \left|u_1, u_2\right.\right)\left|\boldsymbol{\theta}_{34|1}\right.\right) \end{aligned} \quad . \tag{10}$$

<Figure 3>

2.4 Covariate-based nonstationary frequency analysis using the Bayesian estimation

The covariate-based extreme frequency analysis has been widely used (Villarini et al., 2009; Ouarda and El‐Adlouni, 2011; López and Francés, 2013; Xiong et al., 2018). According to these studies, five distributions, gamma (GA), Weibull (WEI), lognormal (LOGNO), Gumbel (GU), and the generalized extreme value (GEV), were used as candidate distributions in this study. In addition, their density functions, the corresponding moments, and the used link functions are shown in Table 1. In the following, the nonstationary distribution models based on Bayesian estimation are developed for a covariate-based flood frequency analysis.

<Table 1>

Suppose that flood variable, Y_t, obeys the distribution $f_{Y_t}\left(y_t|\boldsymbol{\eta}_t\right)$ with the distribution parameters $\boldsymbol{\eta}_t=\left[\mu_t,\sigma_t,\xi\right]$. In this study, only the distribution parameters μ_t and σ_t were allowed to be dependent on covariates because the shape parameter of the GEV is sensitive to the quantile estimation of rare events. According to the linear additive formulation of the generalized additive models for location, scale, and shape (GAMLSS) (Rigby and Stasinopoulos, 2005; Villarini et al., 2009), seven nonstationary scenarios for the formulas of the two distribution parameters, μ_t and σ_t, were investigated, as shown in Table 2. The constant scenario (S0) included one scenario (both μ_t and σ_t are constants). The RI-dependent scenarios (S1) included three scenarios: S11 (μ_t is RI-dependent and σ_t is constant), S12 (μ_t is constant and σ_t is RI-dependent), and S13 (both μ_t and σ_t are RI-dependent). In addition, the RRCI-dependent scenarios (S2) including S21, S22, and S23 are similar to S11, S12, and S13, respectively.

<Table 2>

In the following, the Bayesian inference is introduced. The GEV_S23 (representing the nonstationary GEV distribution with the S23 scenario) model was used as an example, and the model parameter vector $\boldsymbol{\theta}_{\text{GEV\_S23}} = [\alpha_0, \alpha_1, \beta_0, \beta_1, \xi]$ was used as the estimate. The Bayesian method was used to estimate $\boldsymbol{\theta}_{\text{GEV\_S23}}$. Let the prior probability distribution be $\pi(\boldsymbol{\theta}_{\text{GEV\_S23}})$, and the observations, $\boldsymbol{D}$, have the likelihood $l(\boldsymbol{D}|\boldsymbol{\theta}_{\text{GEV\_S23}})$. Then the posterior probability distribution $p(\boldsymbol{\theta}_{\text{GEV\_S23}}|\boldsymbol{D})$ can be calculated using Bayes' theorem as follows:

$$p(\boldsymbol{\theta}_{\text{GEV\_S23}}|\boldsymbol{D}) = \frac{l(\boldsymbol{D}|\boldsymbol{\theta}_{\text{GEV\_S23}})\pi(\boldsymbol{\theta}_{\text{GEV\_S23}})}{\int_{\boldsymbol{\Omega}} l(\boldsymbol{D}|\boldsymbol{\theta}_{\text{GEV\_S23}})\pi(\boldsymbol{\theta}_{\text{GEV\_S23}})d\boldsymbol{\theta}_{\text{GEV\_S23}}} \propto l(\boldsymbol{D}|\boldsymbol{\theta}_{\text{GEV\_S23}})\pi(\boldsymbol{\theta}_{\text{GEV\_S23}}), \quad (11)$$

where the integral is the normalizing constant, and $\boldsymbol{\Omega}$ is the entire parameter space. The obvious difference between the Bayesian method and the frequentist method is that the Bayesian method considers the parameters $\boldsymbol{\theta}_{\text{GEV\_S23}}$ to be random variables. In addition, the desired distribution of the random variables can be obtained using a Markov chain that can be constructed using various Markov chain Monte Carlo (MCMC) algorithms (Reis Jr and Stedinger, 2005; Ribatet et al., 2007) to process Equation (11). In addition, in this study, the Metropolis-Hastings algorithm was used (Chib and Greenberg, 1995; Viglione et al., 2013), which was done with the aid of the R package “MHadaptive”

(Chivers, 2012). A beta distribution function was used with the parameters $u=6$ and $v=9$, which were suggested by Martins and Stedinger (2000) and Martins and Stedinger (2001) as the prior distribution on the shape parameter ξ. For the other model parameters, $\alpha_0, \alpha_1, \beta_0, \beta_1$, the prior distributions were set to non-informative (flat) priors. There are two advantages of the Bayesian method. First, as noted by El Adlouni et al. (2007), this method allows the addition of other information, such as historical and regional information, by defining the prior distribution. Second, the Bayesian method can provide an explicit way to account for the uncertainty of parameters estimates. In the nonstationary case in the *t*-year, the 95% credible interval for the estimation of the flood quantile corresponding to a given probability, p, can be obtained from a set of stable parameters estimations, $\hat{\boldsymbol{\theta}}^{i}_{\text{GEV\_S23}}(i=1,2,...,M_c)$, in which M_c is the length of the Markov chain.

The procedure of model selection can identify which of the five distributions is optimal, and which of the seven nonstationary scenarios is optimal. If all the distribution parameters are identified as constants (S0), this process will be a stationary frequency analysis. To select the optimal model, the Schwarz Bayesian criterion (SBC) (Schwarz, 1978) for each fitted model object is calculated by the following:

$$\text{SBC} = -2\ln\left(\hat{l}\right) + \ln\left(n\right) * \text{df} , \tag{12}$$

where $\ln\left(\hat{l}\right)$ is the maximized log-likelihood of the model object; df is the freedom degree; and n is the number of data points. SBC has a larger penalty on the over-fitting phenomenon than the Akaike information criterion (AIC) (Akaike, 1974). The model object with the lower SBC is preferred. The worm plot and the QQ plot were employed to check whether the model represented the data well.

3 Study area and data

3.1 Study area

Hanjiang River (Figure 4), with the coordinates of 30° 30′ –34° 30′ N, 106° 00′ –114° 00′ E and a catchment area of 159,000 km^2, is the largest tributary of the Yangtze River, China. This area has a warm temperate, semi-humid, continental monsoon climate. The temperature in the basin is not much different from upstream to downstream. Although the elevation range of the study area is quite wide (13–3493 m), the study area is a rainfall-dominated area, and the snowmelt contribution is quite limited. The Ankang gauging station was used as an example. The timing of the AMDF is primarily during the major rainfall period from June to September (Figure S3a, c, and d). In addition, the winter is warm, with mean temperature values of more than 2 ℃, as shown in Figure S3b. Since 1960, many reservoirs have been completed in the Hanjiang basin. Information of the five major reservoirs is shown in Table 3, including the longitude, latitude, control area, time for completion, and capability. The Danjiangkou Reservoir in central China's Hubei province is the largest one in this basin and was completed by 1967. As a multi-purpose reservoir, it primarily aims to supply water and control floods, and it is also used for electricity generation and irrigation. The reservoir has a total storage capacity of 21.0 billion m^3, a dead storage capacity of 7.23 billion m^3, an effective storage capacity of 10.2 billion m^3, and a flood control capacity of 7.72 billion m^3. After the Danjiangkou Dam Extension Project in 2010, the Danjiangkou Reservoir gained an additional capacity of 13.0 billion m^3 and an extra flood control storage capacity of 3.3 billion m^3. In addition, this reservoir is operated using the strategy of staged increases in the flood limit water level during the flood control season (Zhang et al., 2009).

<Figure 4>

<Table 3>

3.2 Data

The assessment analysis of reservoir effects on flood frequency utilized streamflow data, reservoir data, and rainfall data. The annual maximum daily flood series (AMDF) was extracted from the daily streamflow records of the three gauges in the Hanjiang River basin; namely the Ankang (AK) station with a drainage area of 38,600 km^2, the Huangjiagang (HJG) station with a drainage area of 90,491 km^2, and the Huangzhuang (HZ) station with a drainage area of 142,056 km^2. The streamflow and reservoir data were provided by the Hydrology Bureau of the Changjiang Water Resources Commission, China (http://www.cjh.com.cn/en/index.html). The annual series of the maximum (M), the intensity (I), volume (V), the timing (T), and the distance (L) were extracted from the daily streamflow data to describe the MARI. Note that the timing of the MARI is equal to the occurrence time of the AMDF during the year. The MARI is a real-averaged event, and any two consecutive days of areal rainfall values in the MARI required more than 0.2 mm. Daily areal rainfall was calculated using the inverse distance weighting (IDW) method based on rainfall records from 16 stations (shown in Figure 4). These rainfall data were downloaded from the National Climate Center of the China Meteorological Administration (source: http://www.cma.gov.cn/). For the AK and HZ gauging stations, all the records were available from 1956 to 2015, while the HJG gauging station only had records available from 1956 to 2013.

4 Results and discussion

4.1 Identification of reservoir effects

To confirm the impact of reservoirs on the annual maximum daily flow (AMDF) in the study area, the mean and standard deviation of the AMDF before and after the construction of the two large reservoirs, the Danjiangkou reservoir (1967) upstream of the HJG and HZ stations and the Ankang reservoir (1992) upstream of the AK, HJG, and HZ stations, were compared. According to Table 4, the mean and standard deviation of the AMDF of the AK, HJG, and HZ stations were significantly reduced. By using the HJG station as an example, the mean of the AMDF (1992–2013) is 4139 m^3/s, which is only 0.28 times 14,951 m^3/s (1956–1966), and the standard deviation is 4074 m^3/s, approximately 0.52 times 7896 m^3/s (1956–1966).

<Table 4>

Figure 5 presents the linear correlation between the five MARI variables (i.e., the maximum, *M*; the intensity, *I*; volume, *V*; the timing, *T*; and the distance *L*) and the AMDF. It was found that for *M*, *I*, *V*, and *T*, except for *T* in the AK station, the Pearson correlation coefficients between these four variables and the AMDF range from 0.27 to 0.71 (p-value<0.05), indicating that these four variables are significantly related to the AMDF. However, there is a Pearson correlation coefficient of no more than 0.24 between *L* and the AMDF for each of the stations. Thus, *L* was excluded from the calculation of the RRCI. A further analysis of the reservoir effects on the downstream AMDF will be performed in the following sections.

<Figure 5>

4.2 Results for the rainfall-reservoir composite index (RRCI)

To obtain the annual values of the RRCI, the RI was estimated first. The RI was affected by the loss of the reservoir capacity, but not to a great extent (Figure S2). This happened because the main reservoirs (Dangjiangkou and Ankang reservoirs) had a small loss rate of no more than 15% (Table S1 and Figure S1).

The C-vine copula model was applied to calculate the OR-JEP of the scheduling-related MARI variables. In the modeling of the univariate marginal, the marginals of the intensity (*I*) of the AK and the HJG stations and the volume (*V*) of the HJG station were revised to deal with their significant change-points (Table S2). To identify the scheduling-related variables from *M*, *I*, *V*, and *T*, the RRCI for all the possible subsets of *M*, *I*, *V*, and *T* was calculated and compared. The Pearson, Kendall, and Spearman correlation coefficients between the RRCI and the AMDF are listed in Table 5. Note that the entire decomposition structure of the C-vine copula for each RRCI of the same station was determined by the ordering of the variables of each subset (shown in the cells of the first column in Table 5). Figure 3 shows an example for the decomposition structure of the 4-dimensional copula. As shown in the first row in Table 5, there is a negative correlation between the AMDF and the RI for each station. The values of the Pearson correlation coefficients between the AMDF and the RI for the AK, HJG, and HZ stations are -0.37, -0.55, and -0.53, respectively, demonstrating that there is a significant relation between the reservoir storage capacity and the reduction in the AMDF. For each station, with the exception of the RRCI of one-dimensional case, the values of the Pearson, Kendall, and Spearman correlation coefficients between the RRCI and the AMDF are higher than between the RI and the AMDF. According to the highest Kendall correlation, the scheduling-related variables for the AK station were *M*, *I*, *V* and *T*. Those for the HJG station were *I* and *T*, and those for the HZ station were *I*, *V*, and *T*.

<Table 5>

Table 6 shows the results of the copula modeling of the scheduling-related variables using the aid of the R package “VineCopula” (https://CRAN.R-project.org/package=VineCopula). Note that for each bivariate pair in the third column in Table 6, three one-parameter bivariate Archimedean copula families (i.e., the Gumbel, Frank, and Clayton copulas) (Nelsen, 2006) were used to select from. As shown in Table 6, the results of the Cramer-von Mises test (Genest et al., 2009) shows that all the C-vine copula models passed the test at a significance level of 0.05. This result indicated that these models were effective for simulating the joint distribution of the scheduling-related variables for the three stations. Finally, the variation in the RI and the RRCI over time is displayed in Figure 6. It can be seen that for each station, after reservoir construction, in most cases, the annual values of the RRCI are larger (close to 1) than those of the RI. In contrast, in few cases, such as in 1983 at the HZ and HJG stations, the RRCI values were lower than the RI values.

<Figure 6>

<Table 6>

4.3 Flood frequency analysis

A nonstationary flood frequency analysis using the RRCI or the RI as the covariate was performed to investigate how the reservoirs affected the downstream flood frequency. A summary of results of fitting the nonstationary models to the flood data is shown in Table 7. Based on the SBC, the lowest values indicate that the best models for the AK, HJG, and HZ stations are the nonstationary WEI distribution with S23, the nonstationary GA distribution with S21, and the nonstationary WEI distribution with S21, respectively, hereafter referred to as WEI_S23, GA_S21, and WEI_S21, respectively. Note that for any one of the five distributions (GA, WEI, LOGNO, GU, and GEV), the RRCI-dependent scenario had a lower SBC value than the RI-dependent scenario for each gauging station. Furthermore, for the RI-dependent and RRCI-dependent scenarios, using the HZ station as an example, the optimal formulas of the two distribution parameters, μ_t and σ_t, are given as follows:

(1) WEI_S11

$$\begin{aligned} \mu_t &= \exp(9.94 - 2.79\mathrm{RI}) \\ \sigma_t &= \exp(0.49) \end{aligned} \tag{13}$$

(2) WEI_S21

$$\begin{aligned} \mu_t &= \exp(9.92 - 1.42\mathrm{RRCI}) \\ \sigma_t &= \exp(0.73) \end{aligned} . \tag{14}$$

It was found that in Equations (13) and (14), there were negative estimates of -2.79 and -1.42 for α_1, respectively, revealing the decreasing degree of the frequency and magnitude of downstream floods due to the reservoir effects.

Figure 7 compares the stationary scenario (S0), the RI-dependent scenario (S1), and the RRCI-dependent scenario (S2) of the same optimal distributions that explain all the flood values and the several largest flood values for each station. The QQ plots (Figure 7a1–c1) show that overall, the RRCI-dependent scenario more adequately captured the entire empirical quantiles (particularly the smallest and largest empirical quantiles) than the two other scenarios for each station. Furthermore, as shown in

Figure 7a2–c2, for the seven largest floods (observed) of each station, the RRCI-dependent scenario produced lower quantile residuals than the two other scenarios.

<Table 7>

<Figure 7>

Figure 8 shows the performance of the best models: WEI_S23 for the AK station, GA_S21 for the HJG station, and WEI_S21 for the HZ station. The points in the worm plots in Figure 8 are within the 95% confidence interval, indicating that the selected models are reasonable. In addition, according to the centile curves plots in Figure 8, the AMFD series is well fitted by the best models. Undoubtedly, with the incorporation of the effects of the MARI, the RRCI-dependent scenario well captured the presence of nonstationarity in the downstream flood frequency. The case of the HZ station was used for the analysis (Figure 8c1). After the construction of the Danjiangkou Reservoir (1967), due to reservoir operation, most of the values of the AMDF had been reduced in magnitude by this reservoir. However, some relatively large flood events still occurred several times, such as 25,600 m^3/s in 1983 and 19,900 m^3/s in 1975. Obviously, this phenomenon of flood occurrences was well explained by the RRCI.

<Figure 8>

The 100-year return levels at a 95% credible interval from WEI_S23 and WEI_S13 for the AK station, GA_S21 and GA_S11 for the HJG station, and WEI_S21 and WEI_S11 for the HZ station are presented in Figure 9. For each station, compared to the optimal RI-dependent distribution, the optimal

RRCI-dependent distribution provided a lower 100-year return level. However, there existed exceptions. In addition, after the construction of the main reservoir, the uncertainty range of the AK station was larger than that of the HJG and HZ stations. A possible explanation for the larger uncertainty range was that the sample size (1993–2015) of the regulated floods at the AK station was smaller. Furthermore, the dependent relationship between the RRCI and the AMDF at the AK station was weaker.

<Figure 9>

4.4 Discussion

The long-term variation in the AMDF series (Figure 8) indicates that the upstream reservoirs had evidently altered the downstream flood regimes. As an example, since the completion of the Danjiangkou reservoir in 1967, the flood magnitude of the HZ station was evidently reduced overall. This is consistent with the results of the effects of reservoirs on the hydrological regime in this area found in previous studies (Cong et al., 2013; GUO et al., 2008; Jiang et al., 2014; Lu et al., 2009). In this study, it was found that there was a significant difference between downstream floods affected by the same reservoir system (with the same RI value). In most cases, relatively small downstream floods were obtained. However, it is of interest to note that there still occurred unexpected large downstream floods in a few cases, in spite of a large RI value. For example, most values of the AMDF in the HZ station have been less 10,000 m^3/s since 1967, but the values of the AMDF in 1983 and in 1975 were

25,600 m^3/s and 19,900 m^3/s, respectively. These unexpected large downstream floods were probably related to the MARI effects on reservoir operation. The five largest (unexpected) floods since 1967 and the corresponding values of the scheduling-related MARI variables in the HZ station are shown in Table 8. It was found that the largest floods from 1967 to 2015 occurred in 1983. For this flood event, the MARI was a rare event (with an OR-JEP value of 0.435 ranking the second in 1967–2015) due to the largest mean intensity ($I = 20.2$ mm) and the second latest occurrence ($T = 281$). Surprisingly, all the timing values of the MARI for these five unexpected downstream floods showed high rankings (2–9th). These timing values were near the end (approximately the 300th day of the year) of the flood control period (July–October) in this area. Actually, near the end of the major flood control period, the storage capacity should be decreased. This is because according to the operation rules of the Danjiangkou reservoir (Zhang et al., 2009), there is a staged increasing flood limit water level during the flood control season. One important cause for those unexpected large downstream floods was probably that the remaining storage capacity at the end of flood season was not sufficient to reduce some late floods. Therefore, in addition to the storage capacity of reservoirs, the MARI effects should be indispensably considered when attempting to accurately quantify the effects of the reservoir on downstream floods.

<Table 8>

With the combination of both the RI and the OR-JEP, the RRCI had a significant difference from RI (Figure 6). With a few exceptions, the RRCI values were higher than the RI values. This indicates that the real reservoir impact may be underestimated by the RI in most cases. Moreover, the RI will also probably overestimate the real reservoir impact in a few cases because of not considering special rainfall events (i.e., the MARI with low values of the OR-JEP). The results of the covariate-based nonstationary flood frequency analysis (Table 7 and Figures 7 and 8) demonstrate that, compared to the RI-dependent scenario, the RRCI-dependent scenario for the optimal nonstationary distribution more completely captured the presence of nonstationarity in the downstream flood frequency. Therefore, the RRCI might be a useful index for accessing the reservoir effects on downstream flood frequency.

Finally, the estimation errors of the OR-JEP should be noted. (1) Only those MARI samples that corresponded to the timing of the AMDF were included to estimate the OR-JEP. This means that some extreme MARI samples that corresponded to the non-maximum flow were not included, resulting in an estimation error for the OR-JEP. To reduce this error, it might be worth considering the use of the peaks-over-threshold sampling method. (2) The areal-averaged MARI was based on the records from 16 rainfall stations using the IDW method. The estimation error of the areal-averaged rainfall can be transferred to the OR-JEP estimation error. Additional rainfall site data and spatial distribution information were needed to reduce the OR-JEP estimation error. Nonetheless, the good performance of the downstream flood frequency model results demonstrated that the MARI samples still remained representative in this study.

5 Conclusions

Accurately assessing the impact of reservoirs on downstream floods is an important issue for flood risk management. In this study, to evaluate the effects of reservoirs on the downstream flood frequency of the Hanjiang River, the rainfall-reservoir composite index (RRCI) was derived from Equation (2), which considers the combination of the reservoir index (RI) and the OR-joint exceedance probability (OR-JEP) of scheduling-related rainfall variables. The main findings are summarized as follows: (1) The magnitude of the downstream flood events has been reduced by the reservoir system in the study area. However, the long-term variation in the observed AMDF series showed that despite the large reservoirs, unexpected large flood events still occurred several times, such as at the Huangzhuang station in 1983. One important cause of the unexpected large floods at the Huangzhuang station may have been related to the operation strategy of staged increases in the flood limit water level of the Danjiangkou reservoir. (2) According to the results of the covariate-based nonstationary flood frequency analysis for each station, compared to the optimal RI-dependent distribution, the optimal RRCI-dependent distribution more completely captured the presence of nonstationarity in the downstream flood frequency. (3) Furthermore, in estimating the 100-year return level for each station, the optimal RRCI-dependent distribution provided a lower 100-year return level, but there existed exceptions. In addition, it provided a smaller uncertainty range associated with the uncertainty of the model parameter.

Consequently, this study demonstrated the necessity of including the antecedent rainfall effects, in addition to the effects of reservoir storage capacity, on reservoir operation to assess the reservoir effects on downstream flood frequency. This study provides a comprehensive approach for downstream flood risk management under the impacts of reservoirs.

Acknowledgments

This research is financially supported jointly by the National Natural Science Foundation of China (NSFC Grants 41890822 and 51525902), the Research Council of Norway (FRINATEK Project 274310), and the Ministry of Education “111 Project” Fund of China (B18037), all of which are greatly appreciated. We greatly appreciate the editor and the two reviewers for their insightful comments and constructive suggestions for improving the manuscript. No conflict of interest exists in the submission of the manuscript.

References

Aas, K., Czado, C., Frigessi, A., and Bakken, H.: Pair-copula constructions of multiple dependence, Insur. Math. Econ., 44, 182-198, doi:10.1016/j.insmatheco.2007.02.001, 2009.

Akaike, H.: A new look at the statistical model identification, IEEE T. Automat. Contr., 19, 716-723, doi:10.1109/TAC.1974.1100705, 1974.

Ayalew, T. B., Krajewski, W. F., and Mantilla, R.: Exploring the effect of reservoir storage on peak discharge frequency, J. Hydrol. Eng., 18, 1697-1708, doi:10.1061/(ASCE)HE.1943-5584.0000721, 2013.

Ayalew, T. B., Krajewski, W. F., and Mantilla, R.: Insights into expected changes in regulated flood frequencies due to the spatial configuration of flood retention ponds, J. Hydrol. Eng., 20, 04015010, doi:10.1061/(ASCE)HE.1943-5584.0001173, 2015.

Batalla, R. J., Gomez, C. M., and Kondolf, G. M.: Reservoir-induced hydrological changes in the Ebro River basin (NE Spain), J. Hydrol., 290, 117-136, doi:10.1016/j.jhydrol.2003.12.002, 2004.

Benito, G. and Thorndycraft, V. R.: Palaeoflood hydrology and its role in applied hydrological sciences, J. Hydrol., 313, 3-15, doi:10.1016/j.jhydrol.2005.02.002, 2005.

Chib, S. and Greenberg, E.: Understanding the metropolis-hastings algorithm, Am. Stat., 49, 327-335, doi:10.1080/00031305.1995.10476177, 1995.

Chivers, C.: MHadaptive: General Markov chain Monte Carlo for Bayesian inference using adaptive Metropolis-Hastings sampling, 2012, available at: https://CRAN.R-project.org/package=MHadaptive, last access: 9 August 2019.

Coles, S.: An Introduction to Statistical Modeling of Extreme Values, Springer, London, doi:10.1007/978-1-4471-3675-0, 2001.

Cong, M., Chunxia, L., and Yiqiu, L.: Runoff change in the lower reaches of Ankang Reservoir and the influence of Ankang Reservoir on its downstream, Resources and Environment in the Yangtze Basin, 22, 1433-1440, 2013. (In Chinese)

Du, T., Xiong, L., Xu, C.-Y., Gippel, C. J., Guo, S., and Liu, P.: Return period and risk analysis of nonstationary low-flow series under climate change, J. Hydrol., 527, 234-250, doi:10.1016/j.jhydrol.2015.04.041, 2015.

El Adlouni, S., Ouarda, T. B. M. J., Zhang, X., Roy, R., and Bobée, B.: Generalized maximum likelihood estimators for the nonstationary generalized extreme value model, Water Resour. Res., 43, 455-456, doi:10.1029/2005WR004545, 2007.

Genest, C., Rémillard, B., and Beaudoin, D.: Goodness-of-fit tests for copulas: A review and a power study, Insur. Math. Econ., 44, 199-213, doi:10.1016/j.insmatheco.2007.10.005, 2009.

Gilroy, K. L. and Mccuen, R. H.: A nonstationary flood frequency analysis method to adjust for future climate change and urbanization, J. Hydrol., 414, 40-48, doi:10.1016/j.jhydrol.2011.10.009, 2012.

Goel, N. K., Kurothe, R. S., Mathur, B. S., and Vogel, R. M.: A derived flood frequency distribution for correlated rainfall intensity and duration, Water Resour. Res., 33, 2103–2107, doi:10.1029/97WR00812, 1997.

Graf, W. L.: Dam nation: A geographic census of American dams and their large-scale hydrologic impacts, Water Resour. Res., 35, 1305-1311, doi:10.1029/1999WR900016, 1999.

Guo, W., Xia, Z., and Wang, Q.: Effects of Danjiangkou Reservoir on hydrological regimes in the middle and lower reaches of Hanjiang River, Journal of Hohai University (Natural Sciences), 36, 733-737, 2008. (In Chinese)

Jiang, C., Xiong, L., Xu, C.-Y., and Guo, S.: Bivariate frequency analysis of nonstationary low-flow series based on the time-varying copula, Hydrol. Process., 29, 1521-1534, doi:10.1002/hyp.10288, 2014.

Kwon, H.-H., Brown, C., and Lall, U.: Climate informed flood frequency analysis and prediction in Montana using hierarchical Bayesian modeling, Geophys. Res. Lett., 35, L05404, doi:10.1029/2007GL032220, 2008.

López, J. and Francés, F.: Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates, Hydrol. Earth Syst. Sci., 17, 3189-3203, doi:10.5194/hess-17-3189-2013, 2013.

Lee, J., Heo, J.-H., Lee, J., and Kim, N.: Assessment of flood frequency alteration by dam construction via SWAT simulation, Water, 9, 264, doi:10.3390/w9040264, 2017.

Liang, Z., Yang, J., Hu, Y., Wang, J., Li, B., and Zhao, J.: A sample reconstruction method based on a modified reservoir index for flood frequency analysis of non-stationary hydrological series, Stoch. Env. Res. Risk A., 32, 1561-1571, doi:10.1007/s00477-017-1465-1, 2018.

Lu, G., Liu, Y., Zou, X., Zou, Z., and Cai, T.: Impact of the Danjiangkou Reservoir on the flow regime in the middle and lower reaches of Hanjiang River, Resources and Environment in the Yangtze Basin, 18, 959-963, 2009. (In Chinese)

Magilligan, F. J. and Nislow, K. H.: Changes in hydrologic regime by dams, Geomorphology, 71, 61-78, doi:10.1016/j.geomorph.2004.08.017, 2005.

Martins, E. S. and Stedinger, J. R.: Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data, Water Resour. Res., 36, 737-744, doi:10.1029/1999WR900330, 2000.

Martins, E. S. and Stedinger, J. R.: Generalized maximum likelihood Pareto-Poisson estimators for partial duration series, Water Resour. Res., 37, 2551-2557, 2001.

Mei, X., Dai, Z., van Gelder, P. H. A. J. M., and Gao, J.: Linking Three Gorges Dam and downstream hydrological regimes along the Yangtze River, China, Earth Space Sci., 2, 94-106, doi:10.1002/2014EA000052, 2015.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., and Stouffer, R. J.: Stationarity Is Dead: Whither Water Management?, Science, 319, 573-574, doi:10.1029/2001WR000367, 2008.

Nelsen, R.: An Introduction to Copulas. Springer Science & Business Media, New York, 2007.

Ouarda, T. and El‐Adlouni, S.: Bayesian nonstationary frequency analysis of hydrological variables 1, J. Am. Water Resour. As., 47, 496-505, doi:10.1111/j.1752-1688.2011.00544.x, 2011.

Pettitt, A. N.: A non-parametric approach to the change-point problem, J. R. Stat. Soc., 28, 126, 1979.

Reis Jr, D. S. and Stedinger, J. R.: Bayesian MCMC flood frequency analysis with historical information, J. Hydrol., 313, 97-116, doi:10.1016/j.jhydrol.2005.02.028, 2005.

Ribatet, M., Sauquet, E., Grésillon, J.-M., and Ouarda, T. B.: Usefulness of the reversible jump Markov chain Monte Carlo model in regional flood frequency analysis, Water Resour. Res., 43, W08403, doi:10.1029/2006WR005525, 2007.

Rigby, R. A. and Stasinopoulos, D. M.: Generalized additive models for location, scale and shape, J. R. Stat. Soc. C.-Appl., 54, 507-554, doi:10.1111/j.1467-9876.2005.00510.x, 2005.

Salas, J. D.: Analysis and modeling of hydrologic time series, Handbook of Hydrology, 1993.

Scarf, P.: Estimation for a four parameter generalized extreme value distribution, Commun. Stat.-Theor. M., 21, 2185-2201, doi:10.1080/03610929208830906, 1992.

Schwarz, G.: Estimating the dimension of a model, Ann. Stat., 6, 461-464, 1978.

Sklar, M.: Fonctions de repartition an dimensions et leurs marges, Publications de l'Institut Statistique de l'Université de Paris, 8, 229-231, 1959.

Su, C. and Chen, X.: Assessing the effects of reservoirs on extreme flows using nonstationary flood frequency models with the modified reservoir index as a covariate, Adv. Water Resour., 124, 29-40, doi:10.1016/j.advwatres.2018.12.004, 2019.

Viglione, A., Merz, R., Salinas, J. L., and Blöschl, G.: Flood frequency hydrology: 3. A Bayesian analysis, Water Resour. Res., 49, 675-692, doi:10.1029/2011WR010782, 2013.

Villarini, G., Smith, J. A., Serinaldi, F., Bales, J., Bates, P. D., and Krajewski, W. F.: Flood frequency analysis for nonstationary annual peak records in an urban drainage basin, Adv. Water Resour., 32, 1255-1266, doi:10.1016/j.advwatres.2009.05.003, 2009.

Volpi, E., Di Lazzaro, M., Bertola, M., Viglione, A., and Fiori, A.: Reservoir effects on flood peak discharge at the catchment scale, Water Resour. Res., 54, 9623-9636, doi:10.1029/2018WR023866, 2018.

Wang, W., Li, H. Y., Leung, L. R., Yigzaw, W., Zhao, J., Lu, H., Deng, Z., Demisie, Y., and Blöschl, G.: Nonlinear filtering effects of reservoirs on flood frequency curves at the regional scale, Water Resour. Res., 53, 8277-8292, doi:10.1002/2017WR020871, 2017.

Wyżga, B., Kundzewicz, Z. W., Ruiz-Villanueva, V., and Zawiejska, J.: Flood generation mechanisms and changes in principal drivers. In: Flood Risk in the Upper Vistula Basin, Springer, Cham, 2016.

Xiong, B., Xiong, L., Chen, J., Xu, C.-Y., and Li, L.: Multiple causes of nonstationarity in the Weihe annual low-flow series, Hydrol. Earth Syst. Sci., 22, 1525-1542, doi:10.5194/hess-22-1525-2018, 2018.

Xiong, L., Jiang, C., Xu, C.-Y., Yu, K. X., and Guo, S.: A framework of change‐point detection for multivariate hydrological series, Water Resour. Res., 51, 8198-8217, doi:10.1002/2015WR017677, 2015.

Yan, L., Xiong, L., Liu, D., Hu, T., and Xu, C. Y.: Frequency analysis of nonstationary annual maximum flood series using the time‐varying two‐component mixture distributions, Hydrol. Process., 31, 69-89, doi:10.1002/hyp.10965, 2017.

Yang, T., Zhang, Q., Chen, Y. D., Tao, X., Xu, C. Y., and Chen, X.: A spatial assessment of hydrologic alteration caused by dam construction in the middle and lower Yellow River, China, Hydrol. Process., 22, 3829-3843, doi:10.1002/hyp.6993, 2008.

Zhang, L., Xu, J., Huo, J., and Chen, J.: Study on stage flood control water level of Danjiangkou Reservoir, Journal of Yangtze River Scientific Research Institute, 26, 13-16, 2009. (In Chinese)

Zhang, Q., Gu, X., Singh, V. P., Xiao, M., and Chen, X.: Evaluation of flood frequency under non-stationarity resulting from climate indices and reservoir indices in the East River basin, China, J. Hydrol., 527, 565-575, doi:10.1016/j.jhydrol.2015.05.029, 2015.

Tables

Table 1: Summary of the probability density functions, the corresponding moments, and the used link functions for nonstationary flood frequency analysis

Distributions	Probability density functions	Moments	Link functions
Gamma (GA)	$f_Y\left(y\middle|\mu_t,\sigma_t\right)=\dfrac{\left(y\right)^{1/\sigma_t^2-1}}{\Gamma\left(1/\sigma_t^2\right)\left(\mu\sigma_t^2\right)^{1/\sigma_t^2}}\exp\left(-\dfrac{y}{\mu_t\sigma_t^2}\right)$ $y>0,\mu_t>0,\sigma_t>0$	$E\left(Y\right)=\mu_t$ $Var\left(Y\right)=\mu_t^2\sigma_t^2$	$g_1\left(\mu_t\right)=\ln\left(\mu_t\right)$ $g_2\left(\sigma_t\right)=\ln\left(\sigma_t\right)$
Weibull (WEI)	$f_Y\left(y\middle|\mu_t,\sigma_t\right)=\left(\dfrac{\sigma_t}{\mu_t}\right)\left(\dfrac{y}{\mu_t}\right)^{\sigma_t-1}\exp\left(-\left(\dfrac{y}{\mu_t}\right)^{\sigma_t}\right)$ $y>0,\mu_t>0,\sigma_t>0$	$E\left(Y\right)=\mu_t\Gamma\left(1+1/\sigma_t\right)$ $Var\left(Y\right)=\mu_t^2\left[\Gamma\left(1+2/\sigma_t\right)-\Gamma^2\left(1+1/\sigma_t\right)\right]$	$g_1\left(\mu_t\right)=\ln\left(\mu_t\right)$ $g_2\left(\sigma_t\right)=\ln\left(\sigma_t\right)$
Lognormal (LOGNO)	$f_Y\left(y\middle|\mu_t,\sigma_t\right)=\dfrac{1}{y\sigma_t\sqrt{2\pi}}\exp\left\{-\dfrac{\left[\log\left(y\right)-\mu_t\right]^2}{2\sigma_t^2}\right\}$ $y>0,-\infty<\mu_t<\infty,\sigma_t>0$	$E\left(Y\right)=w^{1/2}\exp(\mu_t)$ $Var\left(Y\right)=w(w-1)\exp(2\mu_t)$ $w=\exp(\sigma_t^2)$	$g_1\left(\mu_t\right)=\ln\left(\mu_t\right)$ $g_2\left(\sigma_t\right)=\ln\left(\sigma_t\right)$
Gumbel (GU)	$f_Y\left(y\middle|\mu_t,\sigma_t\right)=\dfrac{1}{\sigma_t}\exp\left\{\left(\dfrac{y-\mu_t}{\sigma_t}\right)-\exp\left(\dfrac{y-\mu_t}{\sigma_t}\right)\right\}$ $-\infty<y<\infty,-\infty<\mu_t<\infty,\sigma_t>0$	$E\left(Y\right)=\mu_t-0.57722\sigma_t$ $Var\left(Y\right)=\left(\pi^2/6\right)\sigma_t^2$	$g_1\left(\mu_t\right)=\mu_t$ $g_2\left(\sigma_t\right)=\ln\left(\sigma_t\right)$
Generalized extreme value (GEV)	$f_Y\left(y\middle|\mu_t,\sigma_t,\xi\right)=\dfrac{1}{\sigma_t}\left[1+\xi\left(\dfrac{y-\mu_t}{\sigma_t}\right)\right]^{-1/\xi-1}\exp\left\{-\left[1+\xi\left(\dfrac{y-\mu_t}{\sigma_t}\right)\right]^{-1/\xi}\right\}$ $y>\mu_t-\sigma_t/\xi,-\infty<\mu_t<\infty,\sigma_t>0,-\infty<\xi<\infty$	$E\left(Y\right)=\mu_t-\dfrac{\sigma_t}{\xi}+\dfrac{\sigma_t}{\xi}\eta_1$ $Var\left(Y\right)=\sigma_t^2\left(\eta_2-\eta_1^2\right)/\xi$ $\eta_m=\Gamma\left(1-m\xi\right)$	$g_1\left(\mu_t\right)=\mu_t$ $g_2\left(\sigma_t\right)=\ln\left(\sigma_t\right)$

Table 2: Seven nonstationary scenarios for the formulas of the two distribution parameters (i.e., μ_t and σ_t)

Scenario classification	Scenario codes	The formula of distribution parameters	
		$g_1(\mu_t)$	$g_2(\sigma_t)$
Stationary (S0)	S0	α_0	β_0
RI-dependent (S1)	S11	$\alpha_0 + \alpha_1$RI	β_0
	S12	α_0	$\beta_0 + \beta_1$RI
	S13	$\alpha_0 + \alpha_1$RI	$\beta_0 + \beta_1$RI
RRCI-dependent (S2)	S21	$\alpha_0 + \alpha_1$RRCI	β_0
	S22	α_0	$\beta_0 + \beta_1$RRCI
	S23	$\alpha_0 + \alpha_1$RRCI	$\beta_0 + \beta_1$RRCI

Table 3: Information of the five major reservoirs in the Hanjiang River basin.

Reservoirs	Longitude	Latitude	Area (km^2)	Year	Capacity (10^9 m^3)
Shiquan	108.05	33.04	23,400	1974	0.566
Ankang	108.83	32.54	35,700	1992	3.21
Huanglongtan	110.53	32.68	10,688	1978	1.17
Dangjiangkou	111.51	32.54	95,220	1967	34.0
Yahekou	112.49	33.38	3030	1960	1.32

Table 4: Change in the mean and standard deviation of the AMDF after the construction of the two large reservoirs (Danjiangkou reservoir completed by 1967, and the Ankang reservoir built by 1992).

Stations	Mean (m^3/s)			Standard deviation (m^3/s)		
	1956–1966	1967–1991	1992–2015	1956–1966	1967–1991	1992–2015
AK	9451	10,468	6506	4341	4623	4454
HJG	14,951	7524	4139	7896	5482	4074
HZ	16,603	10,120	5958	8833	5420	4721

Table 5: Correlation coefficients between the RRCI and the AMDF.

Subset of rainfall variables	AK			HJG			HZ		
	Pearson	Kendall	Spearman	Pearson	Kendall	Spearman	Pearson	Kendall	Spearman
-*	-0.37	-0.18	-0.28	-0.55	-0.37	-0.54	-0.53	-0.38	-0.55
M	-0.27	-0.27	-0.37	-0.67	-0.53	-0.74	-0.45	-0.37	-0.51
I	-0.26	-0.25	-0.34	**-0.74**	-0.57	-0.79	-0.54	-0.41	-0.56
V	-0.32	-0.28	-0.39	-0.63	-0.49	-0.69	-0.57	-0.48	-0.65
T	-0.11	-0.17	-0.24	-0.68	-0.55	-0.73	-0.48	-0.40	-0.57
M, *I*	-0.37	-0.28	-0.38	-0.70	-0.56	-0.77	-0.56	-0.43	-0.58
M, *V*	-0.42	-0.29	-0.40	-0.64	-0.50	-0.71	-0.56	-0.45	-0.60
M, *T*	-0.37	-0.26	-0.36	-0.69	-0.57	-0.77	-0.64	-0.46	-0.63
I, *V*	-0.46	-0.31	-0.42	-0.71	-0.54	-0.76	-0.65	-0.50	-0.67
I, *T*	-0.34	-0.22	-0.31	-0.73	**-0.60**	**-0.80**	-0.68	-0.50	-0.66
V, *T*	-0.43	-0.28	-0.39	-0.68	-0.55	-0.75	-0.69	-0.52	-0.71
M, *I*, *V*	-0.49	-0.31	-0.42	-0.65	-0.53	-0.74	-0.63	-0.47	-0.63
M, *I*, *T*	-0.41	-0.27	-0.37	-0.68	-0.57	-0.78	-0.67	-0.49	-0.66
M, *V*, *T*	-0.50	-0.29	-0.40	-0.65	-0.56	-0.76	-0.67	-0.49	-0.67
I, *V*, *T*	-0.51	-0.31	-0.41	-0.67	-0.58	-0.78	**-0.71**	**-0.53**	**-0.70**
M, *I*, *V*, *T*	**-0.53**	**-0.31**	**-0.42**	-0.65	-0.57	-0.77	-0.69	-0.52	-0.69

*The values in the first row are the correlation coefficients between RI and flood series

Table 6: Results of the copula models for scheduling-related rainfall variables

Stations	Scheduling-related variables	Pairs	Copula type	Parameters θ_c	Kendall's tau	Goodness-of-fit test based on the empirical copula	
						CvM*	p-value
AK	*M*, *I*, *V*, *T*	14	Clayton	0.16	0.08	0.169	0.860
		13	Clayton	1.28	0.39		
		12	Clayton	1.01	0.33		
		24|1	Frank	1.21	0.17		
		23|1	Frank	-2.24	-0.24		
		34|12	Clayton	0.96	0.11		
HJG	*I*, *T*	24	Clayton	1.37	0.41	0.473	0.425
HZ	*I*, *V*, *T*	24	Gumbel	1.12	0.11	0.181	0.820
		23	Clayton	1.31	0.40		
		34|2	Clayton	0.49	0.20		

* CvM is the statistic of the Cramer-von Mises test. If the p-value of the C-vine copula model is less than the significance level of 0.05, the model is considered to be not consistent with the empirical copula.

Table 7: Summary of the results of the nonstationary flood distribution models

Stations	Covariates	Distributions	The optimal formulas* of distribution parameters				AIC	SBC
			Selected models	μ_t	σ_t	ξ		
AK	RI	GA	WEI_S23	exp(9.24-2.64RI)	exp(-0.769+2.9RI)	-	1177.2	1185.5
	RI	WEI		exp(9.36-2.83RI)	exp(0.882-3.18RI)	-	1176.9	1185.3
	RI	LOGNO		exp(9.14-3.86RI)	exp(-0.716+3.28RI)	-	1180.4	1188.8
	RI	GU		11875-13093RI	exp(8.5)	-	1199.6	1205.9
	RI	GEV		7685-15252RI	exp(8.3)	-0.043	1182.3	1190.6
	RRCI	GA		exp(9.28-1.11RRCI)	exp(-0.825+0.689RRCI)	-	1165.3	1173.7
	RRCI	WEI		exp(9.4-1.17RRCI)	exp(0.982-0.884RRCI)	-	1163.8	**1172.2**
	RRCI	LOGNO		exp(9.19-1.33RRCI)	exp(-0.749+0.677RRCI)	-	1168.0	1176.4
	RRCI	GU		12555-7535RRCI	exp(8.4)	-	1188.0	1194.2
	RRCI	GEV		8460-6722RRCI	exp(8.2)	-0.096	1172.1	1180.5
HJG	RI	GA	GA_S21	exp(9.7-1.62RI)	exp(-0.25)	-	1139.9	1146.0
	RI	WEI		exp(9.75-1.56RI)	exp(0.27)	-	1141.4	1147.5
	RI	LOGNO		exp(9.47-1.8RI)	exp(-0.17)	-	1140.9	1147.1
	RI	GU		17955-14399RI	exp(8.8)	-	1189.5	1195.7
	RI	GEV		6976-5930RI	exp(8.79-1.49RI)	0.43	1149.9	1160.2
	RRCI	GA		exp(9.99-1.99RRCI)	exp(-0.45)	-	1112.5	**1118.6**
	RRCI	WEI		exp(10.1-1.97RRCI)	exp(0.53)	-	1113.2	1119.4
	RRCI	LOGNO		exp(9.75-1.94RRCI)	exp(-0.38)	-	1113.9	1120.1
	RRCI	GU		23067-20871RRCI	exp(9.2-1.7RRCI)	-	1121.3	1129.6
	RRCI	GEV		12113-10683RRCI	exp(9.2-2.01RRCI)	0.051	1112.5	1122.8
HZ	RI	GA	WEI_S21	exp(9.85-2.87RI)	exp(-0.42)	-	1198.3	1204.9
	RI	WEI		exp(9.94-2.79RI)	exp(0.49)	-	1198.6	1204.9
	RI	LOGNO		exp(9.63-2.93RI)	exp(-0.33)	-	1201.1	1207.4
	RI	GU		18661-23706RI	exp(8.8)	-	1237.5	1243.7
	RI	GEV		9605-13545RI	exp(9.03-2.56RI)	0.099	1207.8	1218.3
	RRCI	GA		exp(9.85-1.52RRCI)	exp(-0.61)	-	1173.1	1179.4
	RRCI	WEI		exp(9.92-1.42RRCI)	exp(0.73)	-	1171.2	**1177.5**
	RRCI	LOGNO		exp(9.72-1.55RRCI)	exp(-0.51)	-	1178.7	1185.0
	RRCI	GU		19214-14344RRCI	exp(8.86-0.881RRCI)	-	1189.7	1198.1

RRCI	GEV	12502-9911RRCI	exp(8.96-1.37RRCI)	-0.068	1176.0	1186.4

*The model parameters in the optimal formulas are the posterior mean from the Bayesian inference.

Table 8: Summary of the rainfall information for the five largest floods after the construction (1967) of the Danjiangkou reservoir in the HZ station

Year	Values (Ranking in 1967-2015)				
	AMDF [m^3/s]	OR_JEP [-]	I [mm]	V [mm]	T [day of the year]
1983	25,600 (1)	0.435 (2)	20.2 (1)	121.4 (19)	281 (2)
1975	19,900 (2)	0.557 (7)	9.6 (18)	163.6 (13)	277 (6)
1974	18,200 (3)	0.506 (4)	12.0 (7)	120.4 (20)	278 (4)
2005	16,800 (4)	0.651 (11)	8.2 (27)	179.7 (10)	278 (4)
1984	16,100 (5)	0.461 (3)	9.9 (15)	256.3 (4)	273 (9)

Figures

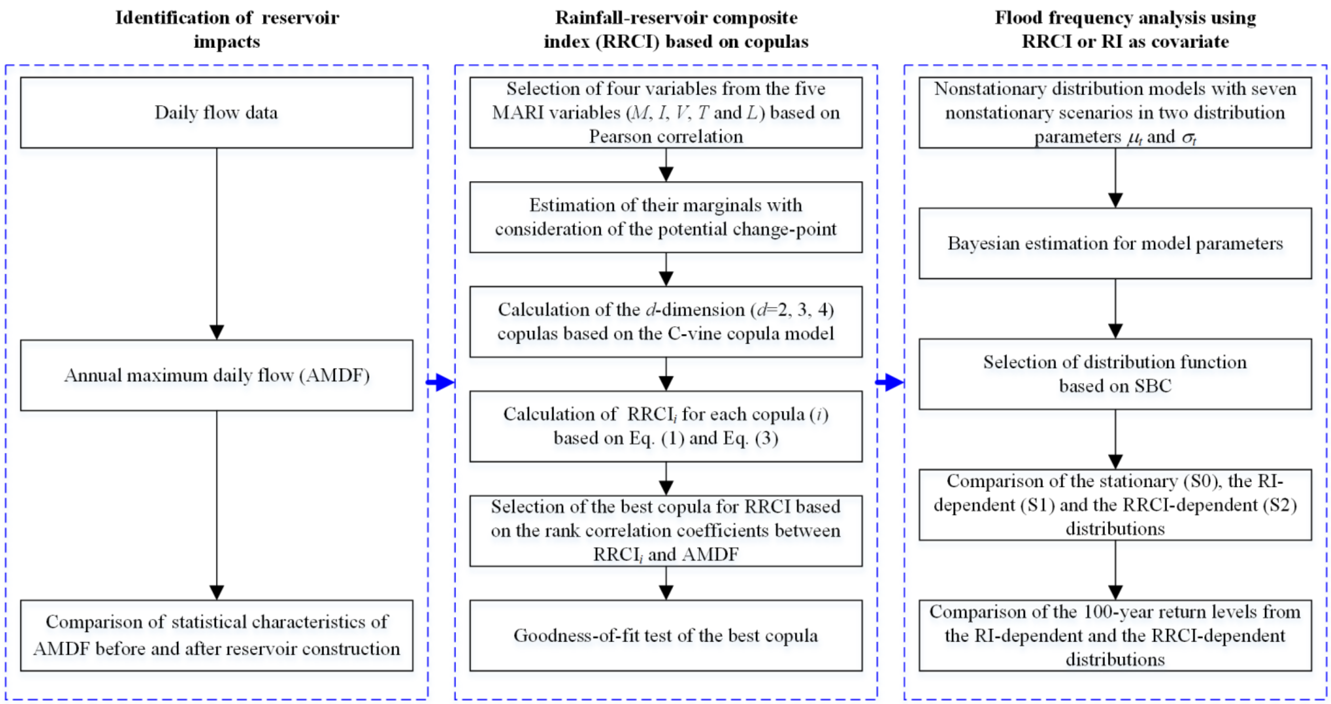

Figure 1: Flowchart of the nonstationary covariate-based flood frequency analysis using the rainfall-reservoir composite index (RRCI)

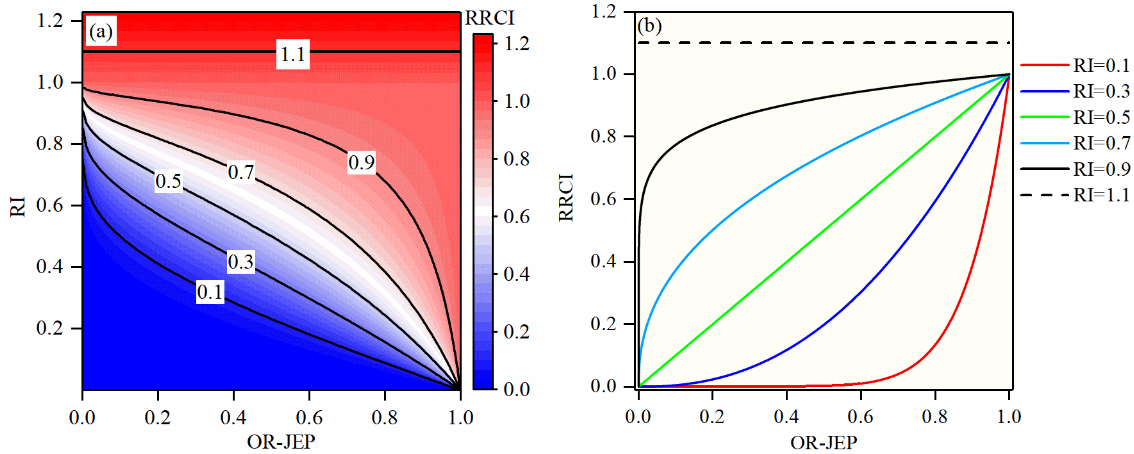

Figure 2: Relationship in Equation (2). (a) The contour plot of the RRCI against both the RI and the OR-JEP; and (b) is the function curves of the RRCI against the OR-JEP under different values of RI

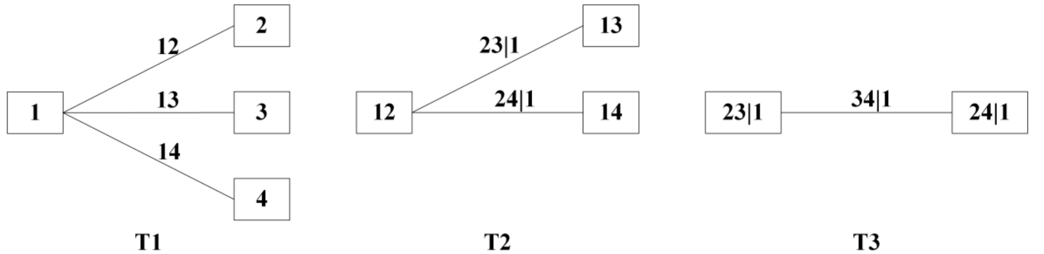

Figure 3: Decomposition of a C-vine copula using four variables and three trees (denoted by T1, T2, and T3)

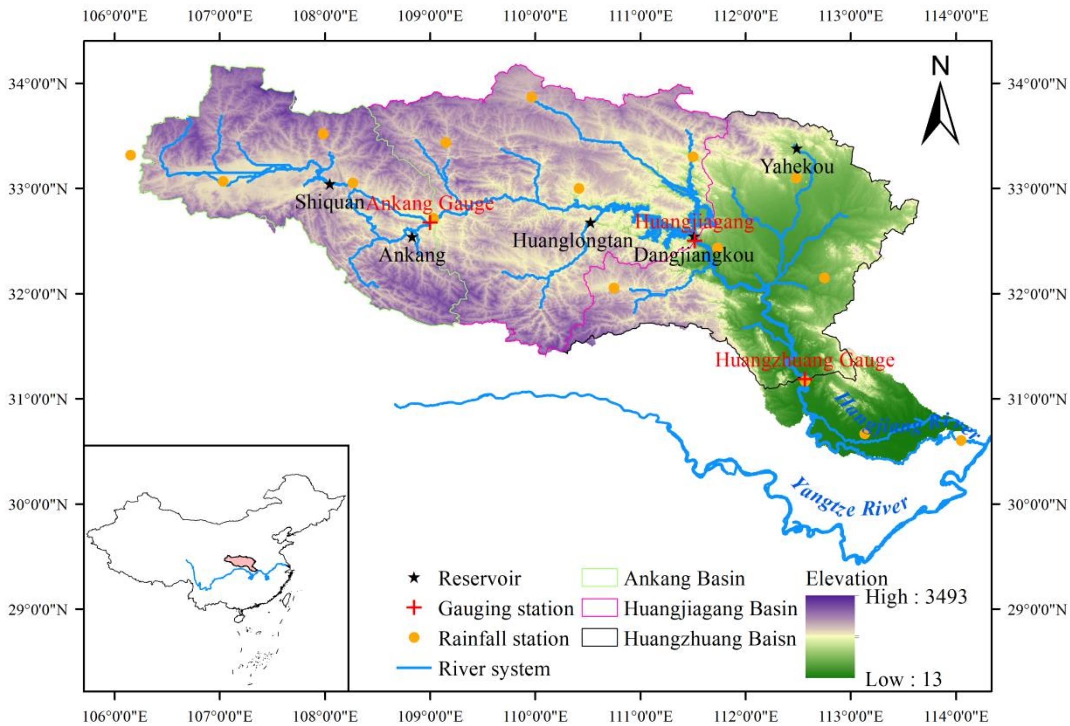

Figure 4: Geographic location of the reservoirs, gauging stations, and rainfall stations along the Hanjiang River.

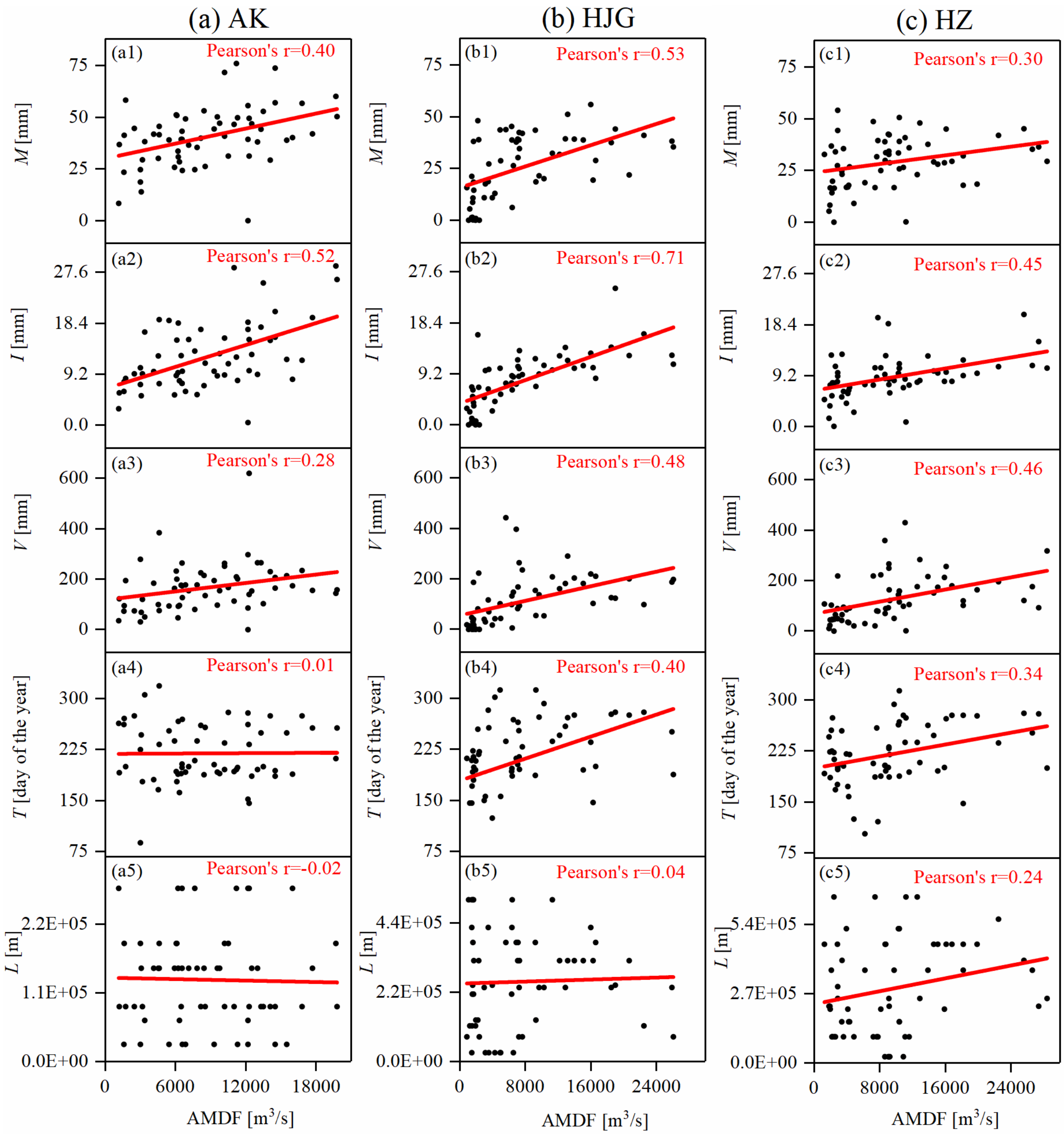

Figure 5: Linear correlation between the five MARI variables and the AMDF for (a) the AK station, (b) the HJG station, and (c) the HZ station

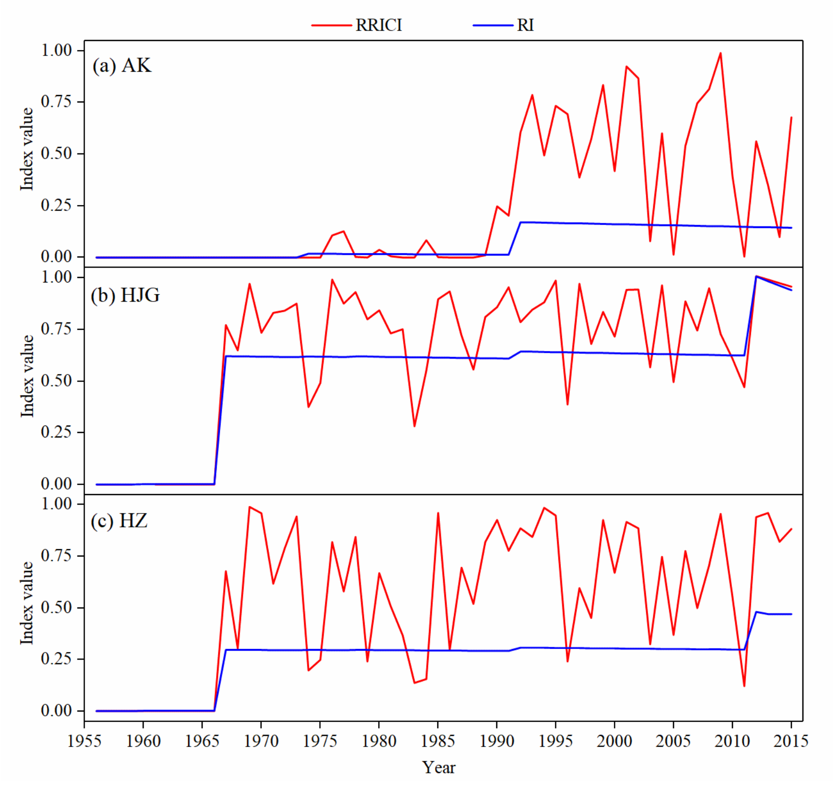

Figure 6: Variation of the RI and the RRCI for (a) the AK station, (b) the HJG station, and (c) the HZ station

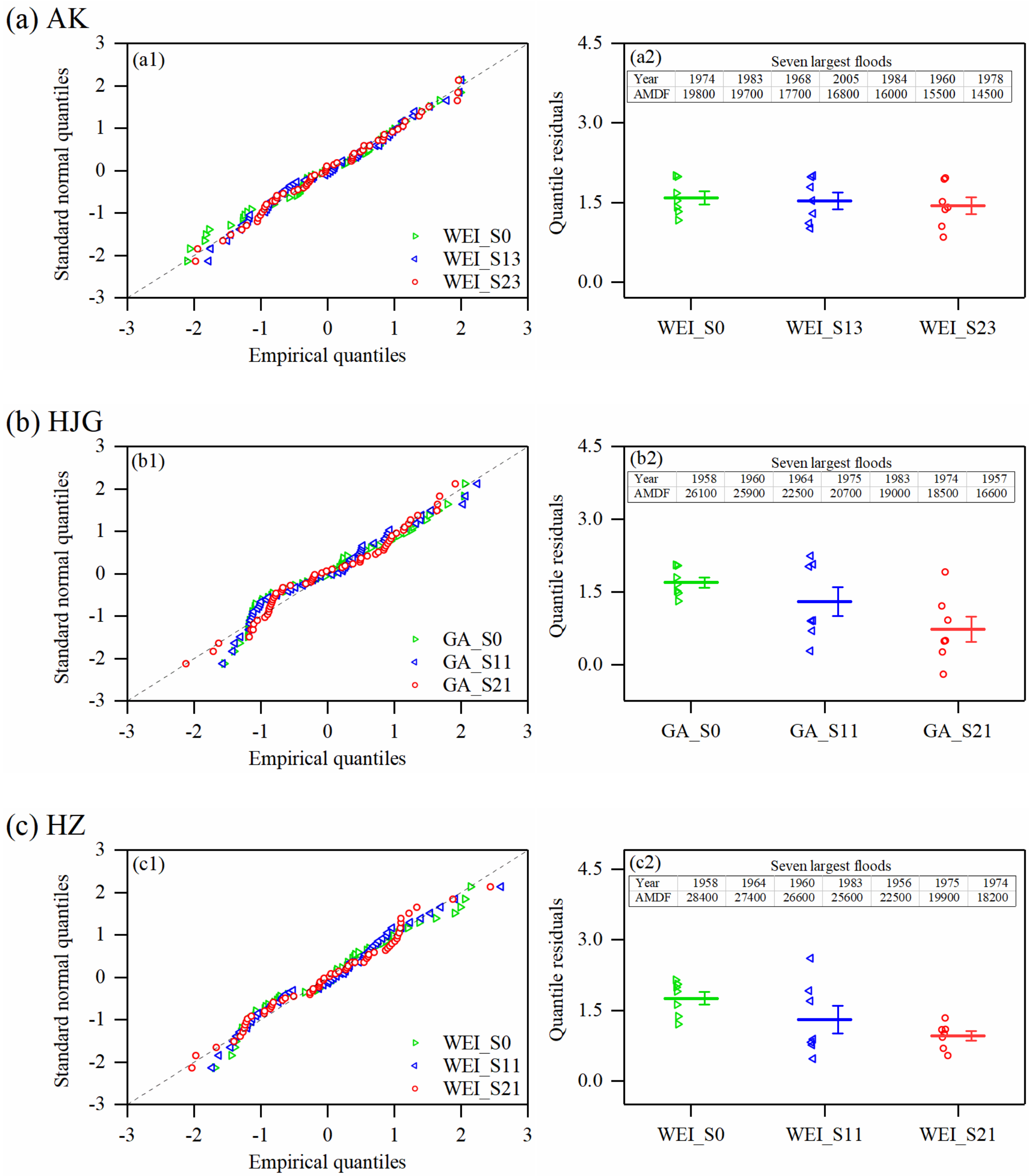

Figure 7: Comparison of the stationary (S0), the RI-dependent (S1), and the RRCI-dependent (S2) scenarios of the same optimal distributions for (a) the AK station, (b) the HJG station, and (c) the HZ station. The left panels (a1, b1, and c1) are the QQ plots for the entire AMDF series in each station.

The right panels (a2, b2, and c2) are the plots of the quantile residuals for the seven largest floods (their values and occurrence years have been listed) in each station, and the means of their quantile residuals (points) and the corresponding standard errors are indicated by the lines

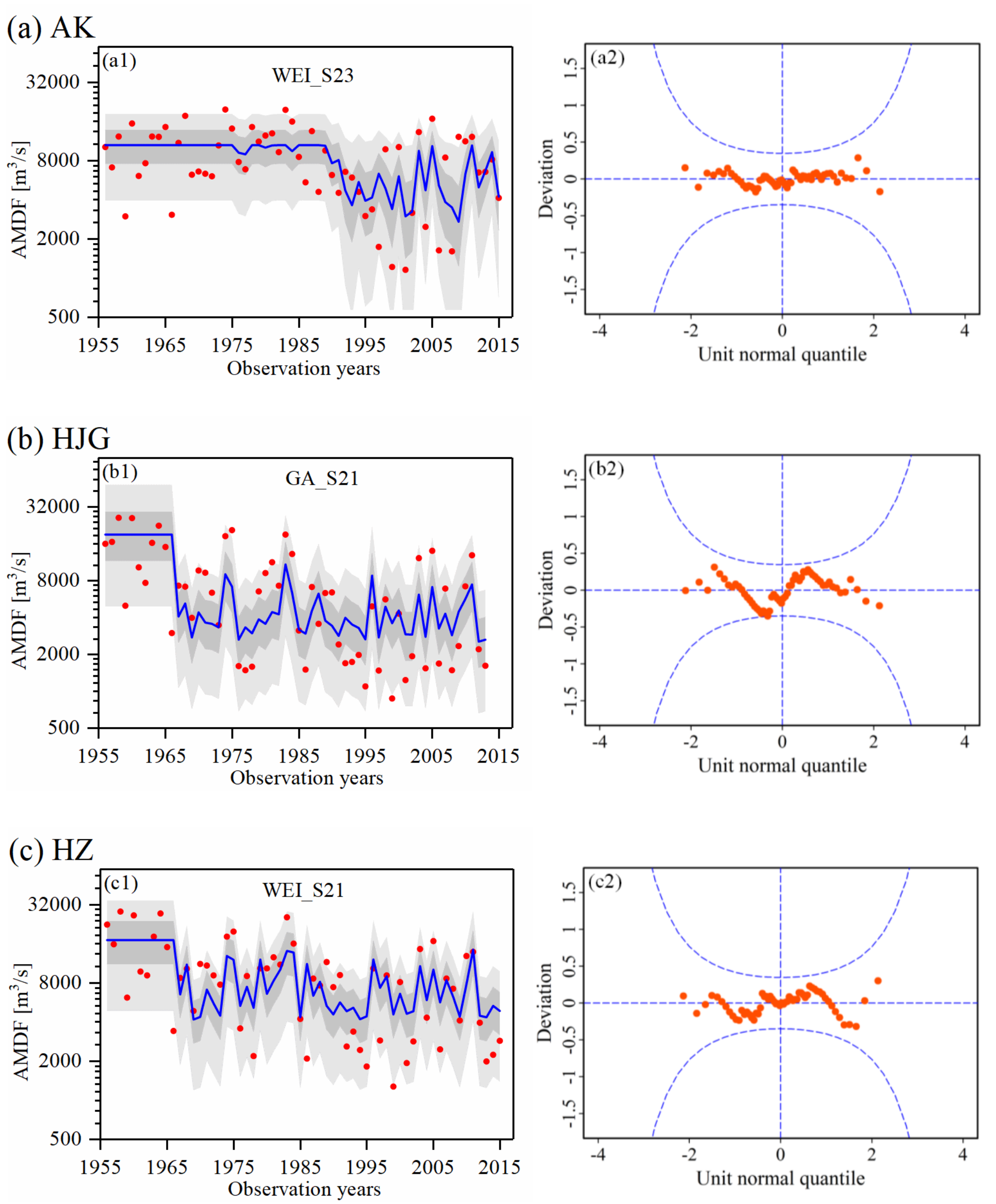

Figure 8: Performance of (a) WEI_S23 for the AK station, (b) GA_S21 for the HJG station, and (c) WEI_S21 for the HZ station. The left panels (a1, b1, and c1) are the centile curves plots in each station (the 50th centile curves are indicated by the thick blue lines; the light gray-filled areas are

between the 5th and 95th centile curves; the dark grey-filled areas are between the 25th and 75th centile curves; and the filled red points indicate the observed series). The right panels (a2, b2, and c2) are the worm plots. A reasonable model should have the plotted points within the 95% confidence intervals (between the two blue dashed curves)

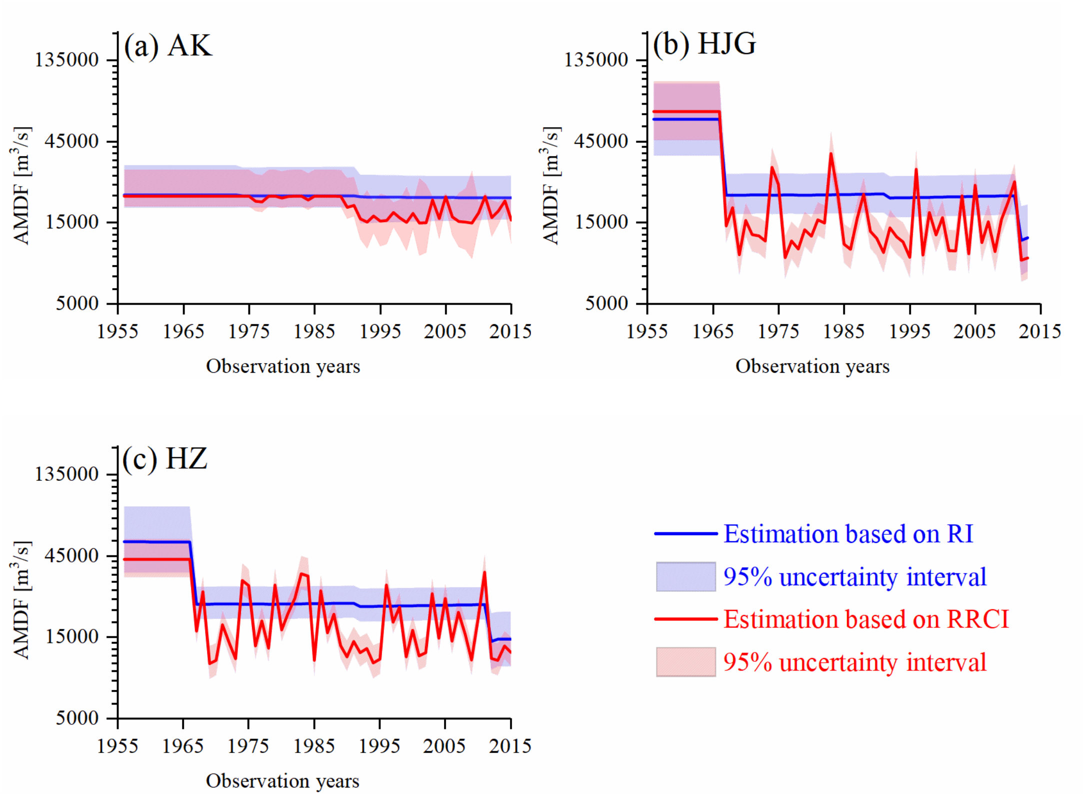

Figure 9: Statistical inference of the 100-year return levels with a 95% uncertainty interval using the optimal RI-dependent and the RRCI-dependent distributions: (a) WEI_S13 and WEI_S23 for the AK station, (b) GA_11 and GA_S21 for the HJG station, and (c) WEI_S11 and WEI_S21 for the HZ station. In nonstationary case, the 95% credible interval in the t-year is calculated by a set of the (99th) percentile estimations which are obtained by the flood distribution functions determined by the values of both covariate in that year and posterior parameter samples.