Water Resources Research

RESEARCH ARTICLE
10.1029/2020WR029374

Key Points:

- A river network-based hierarchical model is developed to derive flood frequency distributions in the Upper Yangtze basin
- The reservoir regulation especially that of the Three Gorges Reservoir remarkably reduces the flood magnitude in the Upper Yangtze basin
- The hierarchical model incorporating the flood dependence within river network improves estimation quality of flood frequency distributions

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

L. Xiong,
xionglh@whu.edu.cn

Citation:

Jiang, C., Xiong, L., Xu, C.-Y., & Yan, L. (2021). A river network-based hierarchical model for deriving flood frequency distributions and its application to the Upper Yangtze basin. *Water Resources Research*, *57*, e2020WR029374. https://doi.org/10.1029/2020WR029374

Received 2 DEC 2020
Accepted 13 JUL 2021

A River Network-Based Hierarchical Model for Deriving Flood Frequency Distributions and Its Application to the Upper Yangtze Basin

Cong Jiang[1], Lihua Xiong[2], Chong-Yu Xu[3], and Lei Yan[4]

[1]School of Environmental Studies, China University of Geosciences, Wuhan, China, [2]State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, China, [3]Department of Geosciences, University of Oslo, Oslo, Norway, [4]College of Water Conservancy and Hydropower, Hebei University of Engineering, Handan, China

Abstract In flood frequency analysis, expanding the additional information with hydrological reasoning beyond local at-site flood samples can be very useful for improving the accuracy of flood frequency distribution (FFD) estimation as well as reflecting a better understanding of flood characteristics. In this study, a river network-based hierarchical model is developed to estimate the FFDs in the Upper Yangtze basin by making full use of the hydrological reasoning information of both the flood dependence within the river network and reservoir regulation. Under this hierarchical model, a covariate analysis based on the generalized additive model for location, scale and shape is performed to obtain the conditional distribution of the interested flood variable given both its upstream flood variables and the reservoir index quantifying reservoir regulation; and then, the FFD of the interested flood variable is derived by combining its conditional distribution with the probability distribution of the upstream flood variables. The application to the Upper Yangtze basin indicates that the proposed hierarchical model suggests a satisfactory performance in FFD estimation. It is also found that the reservoir regulation, especially that of the Three Gorges Reservoir, is of great significance in reducing the flood magnitude in the basin. Compared to the conventional FFD estimation method that directly fits the assumed theoretical probability distributions to the at-site flood samples, the hierarchical model incorporating the flood dependence within the river network exhibits an advantage in capturing the effect of reservoir regulation on the floods as well as in reducing the uncertainty in flood quantile estimation.

1. Introduction

In the conventional practice of flood frequency analysis, the flood frequency distribution (FFD) estimation is usually based on at-site flood samples, and generally includes two standard procedures, i.e., assuming the interested flood samples follow a theoretical probability distribution; and then estimating the distribution parameters in terms of fitting the flood samples by using some statistical techniques such as the method of moments, L-moments and maximum likelihood estimation (MLE) (Hosking, 1990; Merz & Blöschl, 2008a). Hence, the FFD estimation based on at-site flood samples is more like a pure statistical issue without concerning any hydrological reasoning about the physical processes of floods. Regardless of the choice of theoretical probability distributions and the techniques of distribution parameter estimation, the reliability of FFD estimation principally depends on whether the at-site flood samples are able to be representative of the overall occurrence behavior of floods.

The FFD estimation relying only on at-site flood samples is often limited by available flood observations (Merz & Blöschl, 2008a; B. Xiong et al., 2020). In practice, continuous flood records at hydrological gauging sites often have a far less length than the design return periods of hydraulic projects (such as 100 and 1,000 years) or maybe even unavailable in many regions (Engeland et al., 2018). In addition, due to climate condition changes and intensive human activities, flood observations in numerous rivers around the world have been found to exhibit significant nonstationarity (El Adlouni et al., 2007; Jiang, Xiong, et al., 2019; Khaliq et al., 2006; López & Francés, 2013; Villarini et al., 2009; Yan et al., 2017). This suggests that the FFDs relying only on the nonstationary flood samples might fail to be accurately representative of current or future occurrence behaviors of floods.

10.1029/2020WR029374

Beyond local at-site flood samples, incorporating the additional information relevant to flood processes into FFD estimation appears to be the only way to deal with the issues of both the lack of long-term flood observations and flood nonstationarity. Historical information from palaeoflood surveys can expand flood records dating back to a much earlier period than gauging site observations (Engeland et al., 2018; Guo & Cunnane, 1991). By utilizing more evidence of large flood records, the FFD involving historical flood information is able to enhance the reliability of flood quantiles, especially these at the tail (Toonen, 2015; B. Xiong, et al., 2020). Unlike the FFD estimation relying on the flood samples at a single gauging site, the regional flood frequency analysis utilizes the flood samples of a group of neighboring catchments, which contain the information about the regional homogeneity of catchment characteristics controlling flood generations. The methods of regional flood frequency analysis have been shown to be able to improve the accuracy of FFD estimation and apply to ungauged regions (Dawdy et al., 2012; Hosking & Wallis, 1997; Kroll & Stedinger, 1998; Thorarinsdottir et al., 2018). Eagleson (1972) outlined a process-based approach for FFD estimation, in which FFD is no longer directly estimated from at-site flood samples based on an assumed theoretical probability distribution, but is derived by the probability distribution of hydrological input variables (such as rainfall intensity, rainfall duration and soil moisture) via a hydrological model (Sivapalan et al., 2005; Yu et al., 2019). Obviously, the process-based approach concerning the hydrological reasoning information of both hydrological input conditions and flood generation processes can provide a practicable alternative to estimate FFDs at ungauged regions. From a cause-effect perspective, flood nonstationarity can be certainly attributed to some specific driving forces, such as climatic condition changes and human activities (Jiang, Xiong, et al., 2019; López & Francés, 2013; Villarini et al., 2009; Vogel et al., 2011; Volpi et al., 2018; B. Xiong et al., 2020; L. Xiong et al., 2015). In recent decades, the generalized additive model for location, scale and shape (GAMLSS) (Rigby & Stasinopoulos, 2005) has been widely employed to capture flood nonstationarity by linking FFD parameters to the covariates indicating nonstationarity driving forces (Debele et al., 2017; Jiang, Xiong, et al., 2019; López & Francés, 2013; Villarini et al., 2009, 2010; B. Xiong et al., 2019; Yan et al., 2017). The FFDs involving the additional information of nonstationarity driving forces are more adaptable to a changing environment. Given the above, the expansion of the additional information with hydrological reasoning is of great significance in improving the accuracy of FFD estimation as well as providing a better understanding of flood characteristics (Khaliq et al., 2006; Merz & Blöschl, 2008a, 2008b; Naghettini et al., 1996; Toonen, 2015).

Due to the hydraulic connection of river channel, there is an inherent linkage between the flood processes at different locations in the river network (Dingman, 2015; Ravindranath et al., 2019). In other words, the flood variable at a gauging site of interest should be dependent on its upstream flood variables. It can be expected that the flood dependence within the river network can provide a more complete understanding of the flood characteristics such as the regional composition of rive floods and the transmission of flood characteristics along river channel. However, to our knowledge, only very few studies have took account of the flood dependence within the river network in flood frequency analysis. Merz and Blöschl (2008b) presented a relevant study, in which the flood moments at a river confluence were directly calculated by simply summing the flood moments at two immediately upstream sites. Apparently, the aforementioned approach only applies to a particular situation, in which the flood at the site of interest should be the linearly additive sum of its upstream floods. The methods of regional flood frequency analysis are able to utilize the flood information at the gauging sites in neighboring catchments, but do not involve the flood dependence within the river network.

In this study, a river network-based hierarchical model is developed as an approach that can make full use of the hydrological reasoning information of the flood dependence within the river network in FFD estimation. To illustrate the development of the hierarchical model, we present a case study for the Upper Yangtze River basin, where the flood processes are being significantly regulated by reservoirs. A covariate analysis based on the GAMLSS model is carried out to build the conditional distribution of the interested flood variable given both its upstream flood variables and the reservoir index (RI) quantifying reservoir regulation. Under the framework of the hierarchical model, the FFD at a gauging site of interest is no longer characterized by a theoretical distribution relying only on the local at-site flood samples, but is derived by combining the conditional distribution of the interested flood variable with the probability distribution of its upstream flood variables.

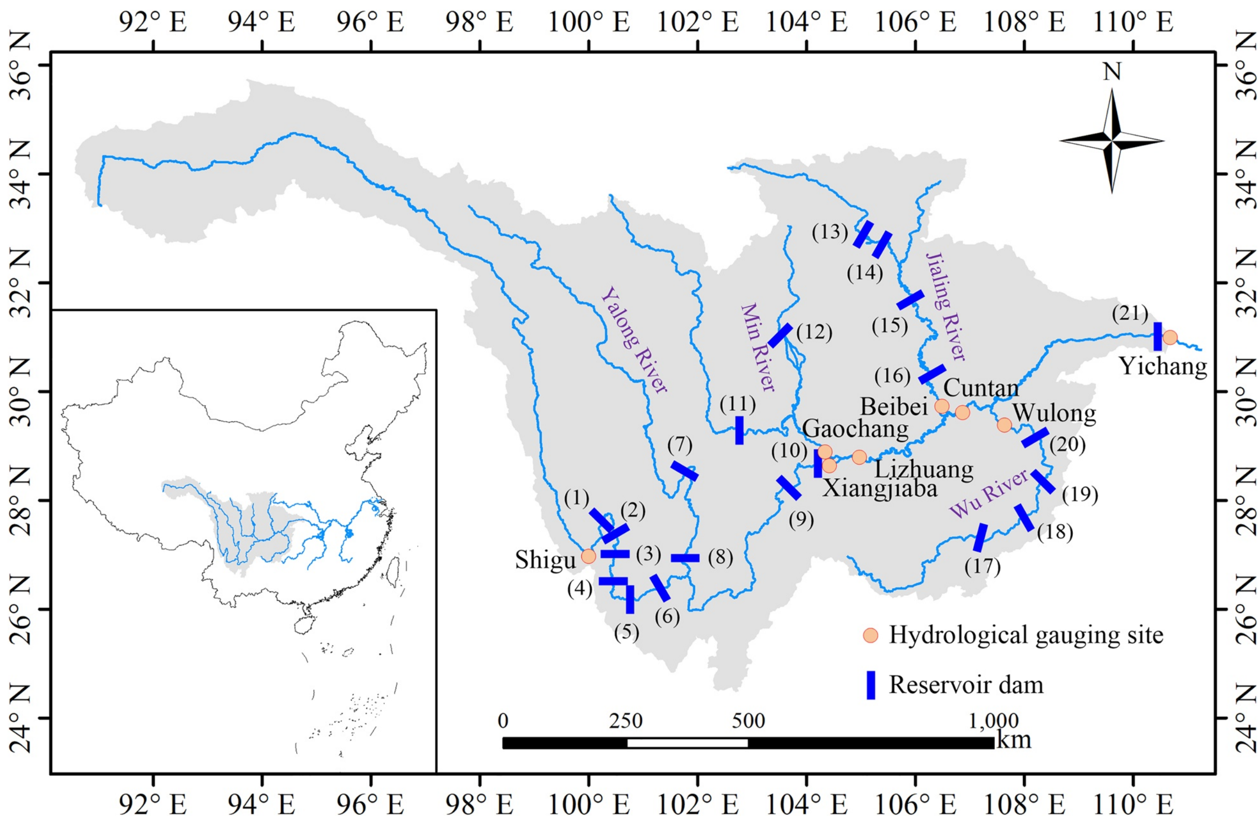

Figure 1. Map of the Upper Yangtze basin.

Section 2 of this paper presents a brief introduction to the study region and data set used in this study, as well as details of the river network-based hierarchical model. Results of the case study for the Upper Yangtze basin are provided in Section 3. Some discussion is given in Section 4, followed by the conclusion in Section 5.

2. Data and Methodology

2.1. Study Region and Data Set

The Yangtze River is the largest river in China, controlling a catchment area of 1.8 million km^2. The Yangtze basin, which contributes more than half of the total GDP of the country and is the home to 459 million people, is the most important economic zone of China. Unfortunately, river flooding has always been a serious natural hazard in the basin because of a huge catchment area as well as a humid subtropical monsoon climate. In this study, we focus on the floods in the Upper Yangtze basin, which has a drainage area of about a million km^2 and contributes the major source of river flooding in the whole basin. Figure 1 presents a map of the Upper Yangtze basin above the gauging site of Yichang, which is exactly the dividing point between the Upper and Middle Yangtze basins.

In the Upper Yangtze basin, the main river channel absorbs the flows of four major tributaries, which are the Yalong River, Min River, Jialing River and Wu River (see Figure 1). In this study, we have gathered the annual maximum daily discharge series at eight key hydrological gauging sites during the observation period from 1951 to 2015. Among these gauging sites, five of them are located at the main river channel, and the rest are located near the outlets of the tributaries of the Min River, Jialing River and Wu River (see Figure 1 and Table S1). Since long-term flood observations of the Yalong River are unavailable, we have to ignore the flood frequency of this tributary.

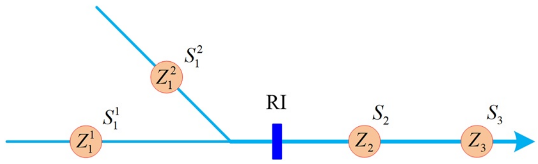

Figure 2. Schematic of a simple river network structure.

For the purpose of both hydropower generation and flood control, dozens of dams have been built in the Upper Yangtze basin, forming the reservoirs with huge storage volume. As a result, these reservoirs are able to regulate the discharge out from the dams, and thereby inevitably disrupt natural flood routing processes in the river network (B. Gao et al., 2013; Jiang, Zhang, & Luo, 2019; B. Xiong et al., 2020). By the end of 2015, a total of 21 reservoirs with considerable flood control capacity have been constructed. Table S2 summarizes the detailed information of these reservoirs, including catchment area, volume of flood control capacity and year of putting into operation. It can be seen that the Three Gorges Reservoir (TGR) has a far larger regulation capacity than other reservoirs. According to the construction timeline of the Three Gorges Dam, the reservoir entered an initial operation period in September 2006, and began to be fully operated in 2009. Given that the maximum flood of 2006 at Yichang occurred prior to September, the TGR should begin to significantly affect the downstream flood processes from 2007.

In this study, the RI, which was first defined by López and Francés (2013), is employed to quantify the effect of reservoir regulation on floods. RI is a dimensionless quantity depending on two reservoir characteristics, i.e. flood control capacity and catchment area, and is calculated by:

$$\mathrm{RI} = \sum_{i=1}^{K} \left(\frac{A_i}{A_T}\right)\left(\frac{V_i}{V_T}\right) \tag{1}$$

where K is the number of the reservoirs upstream the gauging site at a given year, A_i is the catchment area of each reservoir, A_T is the catchment area of the gauging site, V_i is the flood control capacity of each reservoir, and V_T is the total flood control capacity of all reservoirs by the end of 2015. According to the above definition, RI can vary with the construction of new reservoirs, and is therefore able to capture the evolution of the reservoir influence on floods.

2.2. River Network-Based Hierarchical Model for FFD Estimation

2.2.1. General Framework of the Hierarchical Model

To introduce the general framework of the river network-based hierarchical model, a simple river network structure is designed. As shown in Figure 2, a hydrological gauging site S_2 is assumed to be located at the lower reach of two parallel gauging sites S_1^1 and S_1^2, and a reservoir is also assumed to lie between S_2 and its upstream gauging sites. Thus the flood variable Z_2 at S_2 should be dependent on the bivariate flood variables $\left(Z_1^1, Z_1^2\right)$ at S_1^1 and S_1^2, as well as the RI quantifying the reservoir regulation. It is important to note that there is a substantial difference between $\left(Z_1^1, Z_1^2\right)$ and RI, that the former are stochastic variables having a joint probability distribution, whereas the latter is a deterministic variable.

The dependence of Z_2 on $\left(Z_1^1, Z_1^2, \mathrm{RI}\right)$ is denoted by $Z_2 \mid \left(Z_1^1, Z_1^2, \mathrm{RI}\right)$, and is described by a conditional distribution as follows:

$$Z_2 \mid \left(Z_1^1, Z_1^2, \mathrm{RI}\right) \sim f\left[z_2; \boldsymbol{\theta}_{Z_2}\left(z_1^1, z_1^2, \mathrm{RI}; \boldsymbol{\alpha}_{Z_2}\right)\right] \tag{2}$$

where the distribution parameter vector $\boldsymbol{\theta}_{Z_2}$ is stochastic variables conditioned on the variables $\left(z_1^1, z_1^2, \mathrm{RI}\right)$; and $\boldsymbol{\alpha}_{Z_2}$ denotes the hyperparameter vector modeling the relationship between $\boldsymbol{\theta}_{Z_2}$ and $\left(z_1^1, z_1^2, \mathrm{RI}\right)$. The vector $\boldsymbol{\theta}_{Z_2}$ generally contains at most three distribution parameters, i.e., $\boldsymbol{\theta}_{Z_2} = \left(\mu_{Z_2}, \sigma_{Z_2}, \nu_{Z_2}\right)$, which are defined as location parameter (denoted by μ_{Z_2}) with respect to the mean value of the distribution, scale parameter (σ_{Z_2}) with respect to the variance of the distribution, and, if any, shape parameter (ν_{Z_2}) with respect to the skewness of the distribution. Corresponding to μ_{Z_2}, σ_{Z_2} and ν_{Z_2}, the hyperparameter vector $\boldsymbol{\alpha}_{Z_2}$consists of three components, i.e., $\boldsymbol{\alpha}_{Z_2} = \left(\boldsymbol{\alpha}_{\mu}, \boldsymbol{\alpha}_{\sigma}, \boldsymbol{\alpha}_{\nu}\right)$.

GAMLSS is a univariate distributional regression model, where all the statistical parameters of the assumed probability distribution function can be modelled as additive functions of the explanatory variables (Rigby & Stasinopoulos, 2005). In this study, a covariates analysis based on the GAMLSS model is performed to build the relationship between $\boldsymbol{\theta}_{Z_2}$ and the covariates of $\left(z_1^1, z_1^2, \mathrm{RI}\right)$. Taking the location parameter μ_{Z_2} as an example, it is generally expressed as:

$$\mu_{Z_2} = \alpha_{\mu,0} + \alpha_{\mu,1} \cdot g_1\left(z_1^1\right) + \alpha_{\mu,2} \cdot g_2\left(z_1^2\right) + \alpha_{\mu,3} \cdot g_3\left(\mathrm{RI}\right) \tag{3}$$

where $\alpha_{\mu,0}$, $\alpha_{\mu,1}$, $\alpha_{\mu,2}$ and $\alpha_{\mu,3}$ are the parameters of the GAMLSS model, and are also the hyperparameters for μ_{Z_2}, i.e., $\boldsymbol{\alpha}_\mu = \left(\alpha_{\mu,0}, \alpha_{\mu,1}, \alpha_{\mu,2}, \alpha_{\mu,3}\right)$; and $g_i\left(\cdot\right)\left(i = 1,2,3\right)$ are transforming functions and chosen from the candidates of identity, exponential, and logarithmic functions, aiming for capturing linear or nonlinear relationships between μ_{Z_2} and the covariates.

It is worth reemphasizing that $f\left[z_2; \boldsymbol{\theta}_{Z_2}\left(z_1^1, z_1^2, \mathrm{RI}; \boldsymbol{\alpha}_{Z_2}\right)\right]$ does represent the conditional distribution of Z_2 given $Z_1^1 = z_1^1$ and $Z_1^2 = z_1^2$, rather than the FFD of Z_2. According to the formula of total probability, the probability density function of Z_2 can be theoretically derived by:

$$f_{Z_2}\left(z_2\right) = \iint f\left[z_2; \boldsymbol{\theta}_{Z_2}\left(z_1^1, z_1^2, \mathrm{RI}; \boldsymbol{\alpha}_{Z_2}\right)\right] h\left(z_1^1, z_1^2\right) \mathrm{d}z_1^1 \mathrm{d}z_1^2 \tag{4}$$

where $h\left(z_1^1, z_1^2\right)$ denotes the density function of the joint distribution of $\left(z_1^1, z_1^2\right)$. Thus the FFD of Z_2 can be characterized by a hierarchical model as follows:

$$\begin{aligned} &Z_2 \mid \left(Z_1^1, Z_1^2, \mathrm{RI}\right) \sim f\left[z_2; \boldsymbol{\theta}_{Z_2}\left(z_1^1, z_1^2, \mathrm{RI}; \boldsymbol{\alpha}_{Z_2}\right)\right] \\ &\left(Z_1^1, Z_1^2\right) \sim h\left(z_1^1, z_1^2\right) \end{aligned} \tag{5}$$

It can be seen that the above hierarchical model consists of two stages: the first stage $\left(Z_1^1, Z_1^2\right) \sim h\left(z_1^1, z_1^2\right)$ modeling the joint probability distribution of the upstream flood variables $\left(Z_1^1, Z_1^2\right)$, and the second stage $Z_2 \middle| \left(Z_1^1, Z_1^2, \mathrm{RI}\right) \sim f\left[z_2; \boldsymbol{\theta}_{Z_2}\left(z_1^1, z_1^2, \mathrm{RI}; \boldsymbol{\alpha}_{Z_2}\right)\right]$ modeling the dependence of Z_2 on the conditioning variables of $\left(Z_1^1, Z_1^2, \mathrm{RI}\right)$. The hierarchical model provides an alternative solution to FFD estimation by thinking of FFD in a hierarchy in addition to a theoretical probability distribution with explicit expression and parameters (Casella & Berger, 2002; Steinschneider & Lall, 2015).

If the bivariate flood variables $\left(Z_1^1, Z_1^2\right)$ are independent, their joint probability distribution $h\left(z_1^1, z_1^2\right)$ can be further expressed as:

$$h\left(z_1^1, z_1^2\right) = f_{Z_1^1}\left(z_1^1\right) \cdot f_{Z_1^2}\left(z_1^2\right) \tag{6}$$

where $f_{Z_1^1}\left(z_1^1\right)$ and $f_{Z_1^2}\left(z_1^2\right)$ denote the probability density functions of Z_1^1 and Z_1^2, respectively. If Z_1^1 and Z_1^2 are correlated, the joint probability distribution of $\left(Z_1^1, Z_1^2\right)$ can be constructed by using the copula technique (Jiang, Xiong, et al., 2019; Nelsen, 2006; Salvadori & De Michele, 2010).

Furthermore, if another hydrological gauging site S_3 is located at the downstream of S_2 (Figure 2), the flood variable Z_3 at S_3 can also be derived by a two-stage hierarchy:

$$\begin{aligned} &Z_3 \middle| Z_2 \sim f\left[z_3; \boldsymbol{\theta}_{Z_3}\left(z_2; \boldsymbol{\alpha}_{Z_3}\right)\right] \\ &Z_2 \sim f_{Z_2}\left(z_2\right) \end{aligned} \tag{7}$$

where the first stage $Z_2 \sim f_{Z_2}\left(z_2\right)$ in the above equation is actually the result of combining the two stages in Equation 5. In other words, the two-stage hierarchy defined by Equation 7 is equivalent to a three-stage one as follows:

$$\begin{aligned} &Z_3 \middle| Z_2 \sim f\left[z_3; \boldsymbol{\theta}_{Z_3}\left(z_2; \boldsymbol{\alpha}_{Z_3}\right)\right] \\ &Z_2 \middle| \left(Z_1^1, Z_1^2, \mathrm{RI}\right) \sim f\left[z_2; \boldsymbol{\theta}_{Z_2}\left(z_1^1, z_1^2, \mathrm{RI}; \boldsymbol{\alpha}_{Z_2}\right)\right] \\ &\left(Z_1^1, Z_1^2\right) \sim h\left(z_1^1, z_1^2\right) \end{aligned} \tag{8}$$

The FFD of Z_2 can be treated as the transitive node of two nested two-stage hierarchies. In other words, the flood variable Z_2 on one hand is dependent on their upstream flood variables, and on the other hand, is also conditioning variable for deriving the FFD of Z_3 at the downstream site. Hence, it is easy to extend a two-stage hierarchical model to the case of more stages when each additional stage is model by another two-stage hierarchy, and conversely, a multi-stage hierarchical model can also be thought of as a combination of several nested two-stage hierarchies.

2.2.2. Theoretical Probability Distributions

In the hierarchical model, the conditional distributions of the flood variables Z_2 and Z_3, as well as the FFDs of Z_1^1 and Z_1^2 are assumed to have some theoretical probability distributions. In this study, the optimal theoretical probability distributions are chosen from gamma, normal, lognormal, Weibull, Pearson type III (PIII) and generalized extreme value distributions, which are all commonly used in flood frequency analysis. The detailed information about these probability distributions is presented in Table S3.

2.2.3. Model Parameter Estimation and Uncertainty Analysis

In this study, the parameters of the river-network hierarchical model are estimated by using the MLE method (Rigby & Stasinopoulos, 2005). For the purpose of a better fitting quality but avoiding overfitting, the proper theoretical distributions, covariates and transforming functions for the GAMLSS models are determined in terms of the Bayesian information criterion (BIC) (Schwarz, 1978).

To assess the uncertainty of the hierarchical model, the Markov chain Monte Carlo-based (MCMC-based) Bayesian approach is also employed to separately estimate the model parameters for each gauging site in the river network. In the application, noniformative prior probability distributions are specified for the unknown parameters, and then the posterior distributions of model parameters are calculated through the MCMC method (Vrugt et al., 2009; B. Xiong et al., 2020).

2.2.4. Computation of FFDs Derived by the Hierarchical Model

To illustrate how to compute the FFDs derived by the hierarchical model, we take the hierarchical model defined by Equation 5 as an example. According to the probability density function of Z_2 in Equation 4, the cumulative distribution function of Z_2 can be theoretically calculated by:

$$F_{Z_2}\left(z_2\right) = \int_0^{z_2} f_{Z_2}\left(\xi\right)\mathrm{d}\xi = \int_0^{z_2}\left[\iint f\left[\xi;\boldsymbol{\theta}_{Z_2}\left(z_1^1, z_1^2, \mathrm{RI}; \boldsymbol{\alpha}_{Z_2}\right)\right] h\left(z_1^1, z_1^2\right)\mathrm{d}z_1^1\mathrm{d}z_1^2\right]\mathrm{d}\xi \tag{9}$$

It is necessary to point out that Equation 9 implicitly define a time-varying probability distribution, which can change with the covariate RI or the probability distributions of the conditioning variables Z_1^1 and Z_1^2. The mean value of Z_2 has the following theoretical expression:

$$E\left(Z_2\right) = E\left[E\left(Z_2 \middle| \; Z_1^1, Z_1^2, \mathrm{RI}\right)\right] \tag{10}$$

In practice, $f_{Z_2}\left(z_2\right)$, $F_{Z_2}\left(z_2\right)$ and $E\left(Z_2\right)$ could have no analytical solutions, but can be calculated using the numerical integration based on the Monte Carlo (MC) sampling technique (Niederreiter, 1978). First, based on the joint probability distribution $h\left(z_1^1, z_1^2\right)$, we can generate random samples of $\left(Z_1^1, Z_1^2\right)$ with the size of N, i.e., $\left(zs_{1,i}^1, zs_{1,i}^2\right)$ $(i = 1,2,\ldots,N)$. Then, the probability density function of Z_2 can be calculated by:

$$f_{Z_2}\left(z_2\right) = \frac{1}{N}\sum_{i=1}^{N} f\left[z_2;\boldsymbol{\theta}_{Z_2,i}\left(zs_{1,i}^1, zs_{1,i}^2, \mathrm{RI}; \boldsymbol{\alpha}_{Z_2}\right)\right] \tag{11}$$

where $\boldsymbol{\theta}_{Z_2,i}$ $(i = 1,2,\ldots,N)$ is calculated from the random sample $\left(zs_{1,i}^1, zs_{1,i}^2\right)$ $(i = 1,2,\ldots,N)$ and RI according to Equation 3. Also with respect to $f\left[z_2;\boldsymbol{\theta}_{Z_2,i}\left(zs_{1,i}^1, zs_{1,i}^2, \mathrm{RI}; \boldsymbol{\alpha}_{Z_2}\right)\right]$ $(i = 1,2,\ldots,N)$, we can generate random samples of Z_2 with the size of N, i.e., $zs_{2,1}, zs_{2,2}, \ldots, zs_{2,N}$. Finally, the cumulative probability function of Z_2 can be computed from the samples $zs_{2,1}, zs_{2,2}, \ldots, zs_{2,N}$ using the empirical distribution function:

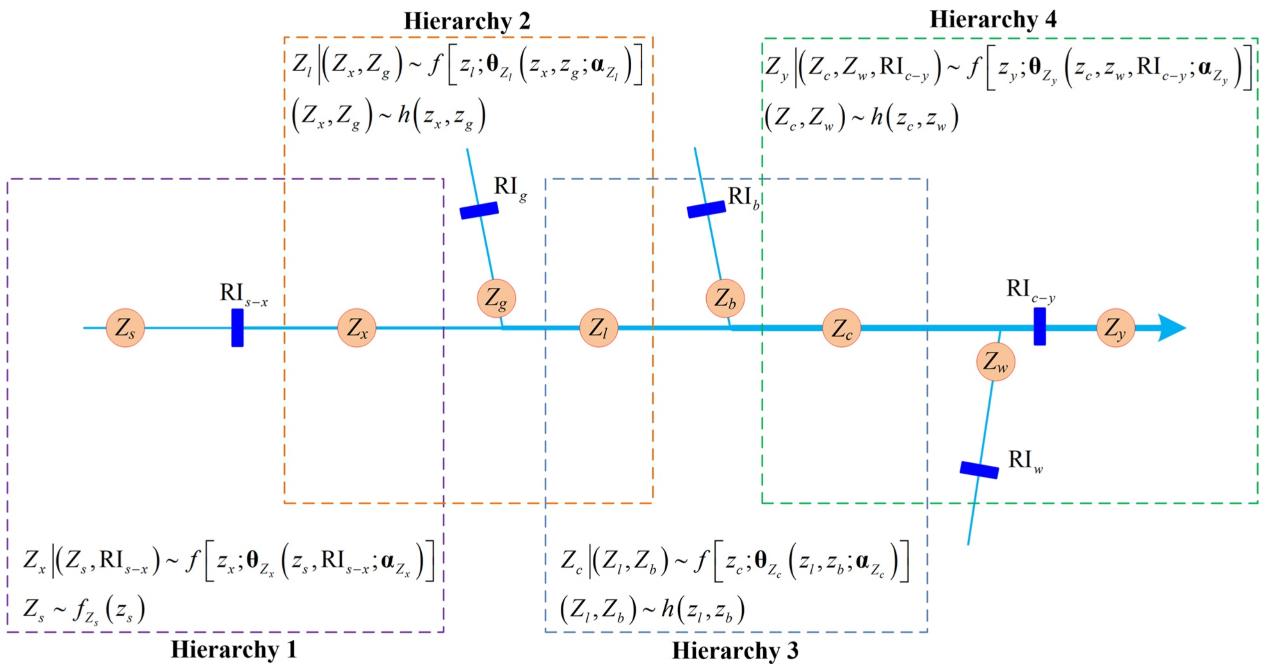

Figure 3. Schematic of the hierarchical model with respect to the river network in the Upper Yangtze basin. Z_s, Z_x, Z_g, Z_l, Z_b, Z_c, Z_w and Z_y represent the flood variables at the hydrological gauging sites of Shigu, Xiangjiaba, Gaochang, Lizhuang, Beibei, Cuntan, Wulong and Yichang, respectively. RI_{s-x} and RI_{c-y} denote the reservoir indices regarding the reservoirs in the catchment between Shigu and Xiangjiaba and the catchment between Cuntan and Yichang, respectively; and RI_g, RI_b and RI_w stand for the reservoir indices regarding the reservoirs in the catchments above Gaochang, Beibei and Wulong, respectively. The flood frequency distributions (FFDs) of Z_s, Z_g, Z_b and Z_w are estimated by directly fitting the assumed theoretical probability distributions to the at-site flood samples, and the FFDs of Z_x, Z_l, Z_c and Z_y are derived by the hierarchical model.

$$F_{Z_2}\left(z_2\right) = \frac{1}{N+1} \sum_{i=1}^{N} \mathbf{1}\left(zs_{2,i} \leq z_2\right) \tag{12}$$

In addition, some important statistical characteristics of Z_2 such as the mean value $E\left(Z_2\right)$, coefficient of variation (Cv) and flood quantiles can also be calculated from $zs_{2,1}, zs_{2,2}, \ldots, zs_{2,N}$.

2.2.5. Goodness-of-Fit Test for FFDs

The goodness of fit (GoF) of FFDs is evaluated by the graphical Probability-Probability (PP) plot and the Kolmogorov-Smirnov (KS) test (Frank & Massey, 1951). The PP plot tests the fitting quality of a FFD by checking the agreement between the theoretical probabilities calculated by the FFD and empirical probabilities. The KS test is used to decide if the flood variables follow a given FFD based on a hypothesis testing. If the p-value of the KS test is larger than a critical significance level, the null hypotheses that the flood variables follow the FFD should be accepted. In this study, the critical significance level for the KS test is set to be 0.05.

2.3. River Network-Based Hierarchical Model for the Upper Yangtze Basin

According to the structure of the river network as well as the spatial distribution of both hydrological gauging sites and reservoirs in the Upper Yangtze basin (Figure 1), the river-network hierarchical model for this basin is outlined in Figure 3.

As shown in Figure 1, Shigu is located at the upstream from Xiangjiaba, and 10 reservoirs lie in the catchment between these two gauging sites, thus the flood variable Z_x at Xiangjiaba should be associated with

10.1029/2020WR029374

both the flood variable Z_s at Shigu and the reservoir index RI_{s-x}. Thus the distribution of Z_x can be modeled by a two-stage hierarchy as follows:

$$\begin{aligned} &Z_x \mid \left(Z_s, \mathrm{RI}_{s-x}\right) \sim f\left[z_x; \boldsymbol{\theta}_{Z_x}\left(z_s, \mathrm{RI}_{s-x}; \boldsymbol{\alpha}_{Z_x}\right)\right] \\ &Z_s \sim f_{Z_s}\left(z_s\right) \end{aligned} \tag{13}$$

where the distribution $f_{Z_s}\left(z_s\right)$ of Z_s is modeled by an assumed theoretical probability distribution.

The flood variable Z_l at Lizhuang depends on those at Xiangjiaba and Gaochang, and thus the second two-stage hierarchy is given by:

$$\begin{aligned} &Z_l \mid \left(Z_x, Z_g\right) \sim f\left[z_l; \boldsymbol{\theta}_{Z_l}\left(z_x, z_g; \boldsymbol{\alpha}_{Z_l}\right)\right] \\ &\left(Z_x, Z_g\right) \sim h\left(z_x, z_g\right) \end{aligned} \tag{14}$$

where Z_g is the flood variable at Gaochang, and $h\left(z_x, z_g\right)$ denotes the density function of the joint probability distribution of Z_x and Z_g. The FFD of Z_g is estimated by using the GAMLSS model with the covariate of the reservoir index RI_g.

After estimating the FFD of Z_l from Equation 14, the probability distribution of the flood variable Z_c at Cuntan is given by:

$$\begin{aligned} &Z_c \mid \left(Z_l, Z_b\right) \sim f\left[z_c; \boldsymbol{\theta}_{Z_c}\left(z_l, z_b; \boldsymbol{\alpha}_{Z_c}\right)\right] \\ &\left(Z_l, Z_b\right) \sim h\left(z_l, z_b\right) \end{aligned} \tag{15}$$

where Z_b denotes the flood variable at Beibei and its probability distribution is estimated by using the GAMLSS model with the covariate of the reservoir index RI_b.

Finally, the fourth two-stage hierarchy for deriving the probability distribution of the flood variable Z_y at Yichang is defined by:

$$\begin{aligned} &Z_y \mid \left(Z_c, Z_w, \mathrm{RI}_{c-y}\right) \sim f\left[z_y; \boldsymbol{\theta}_{Z_y}\left(z_c, z_w, \mathrm{RI}_{c-y}; \boldsymbol{\alpha}_{Z_y}\right)\right] \\ &\left(Z_c, Z_w\right) \sim h\left(z_c, z_w\right) \end{aligned} \tag{16}$$

where Z_w represents the flood variable at Wulong; $h\left(z_c, z_w\right)$ represents the joint probability distribution of the flood variables at Cuntan and Wulong; and the reservoir index RI_{c-y} actually indicates the regulation effect of the TGR, which lies between Cuntan and Yichang.

Figure 4 displays a flow chart to depict the organization of the entire case study for the Upper Yangtze basin. First, the flood dependence within the river network is preliminarily examined by a spatial correlation analysis of flood variables. Then, according to the hierarchical model for the Upper Yangtze basin presented above, the FFD at each gauging site in the river network will be estimated. In particular, the FFDs at the upstream gauging sites of Shigu, Gaochang, Beibei and Wulong are estimated by directly fitting the assumed theoretical probability distributions to the at-site flood samples using the GAMLSS model, and on this basis the FFDs at the downstream gauging sites of Xiangjiaba, Lizhuang, Cuntan and Yichang are derived by the hierarchical model in sequence. Based on the modeling results, the effect of reservoir regulation on floods is evaluated. Finally, the uncertainty of the hierarchical model is assessed by using the MCMC-based Bayesian approach.

3. Results and Analysis

3.1. Spatial Correlations of Floods in the Upper Yangtze Basin

Before employing the hierarchical model to derive the FFDs in the Upper Yangtze basin, a preliminary analysis for the spatial correlations of flood variables is performed. Figure S1 displays the correlations of the flood variables at Xiangjiaba, Lizhuang, Cuntan and Yichang to those at their upstream gauging sites. It can be seen that the linear correlation coefficients r^2 for almost all flood pairs are larger than 0.2, indicating a correlation at the 0.01 significance level. This finding reveals a strong flood dependence within the network in the Upper Yangtze River.

10.1029/2020WR029374

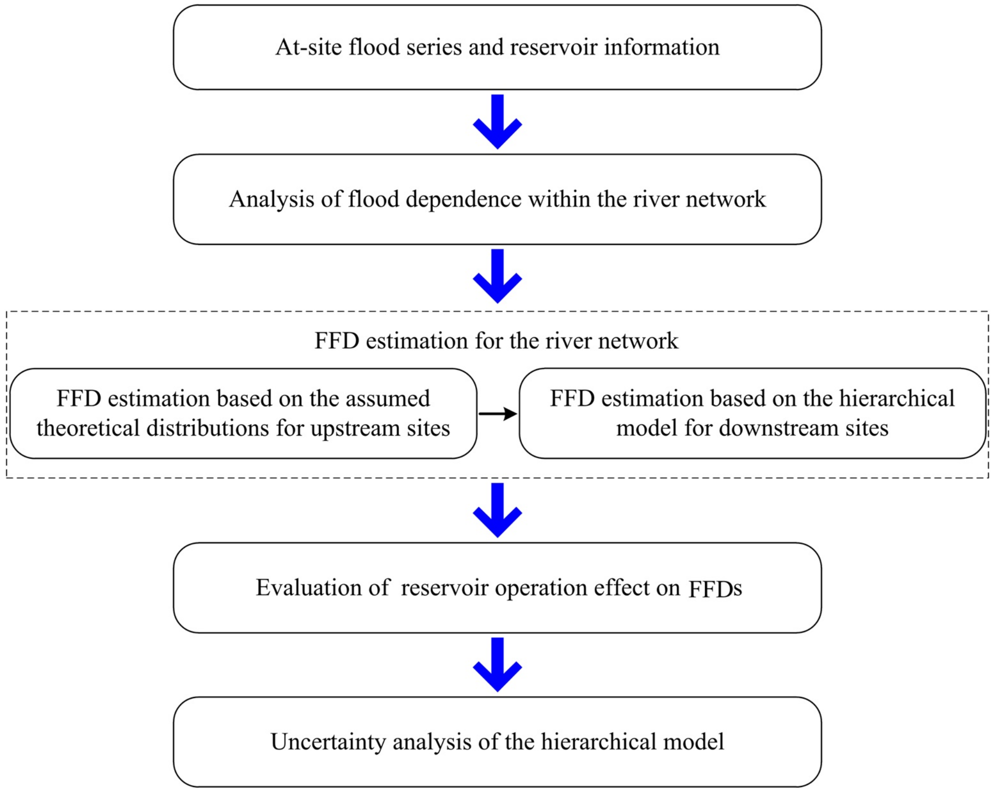

Figure 4. Flow chart of the organization of the entire study.

For the hierarchical model outlined in Figure 3, we need to construct the joint probability distributions of the flood variables at three pairs of parallel gauging sites, i.e., Xiangjiaba-Gaochang, Lizhuang-Beibei, and Cuntan-Wulong. Figure S2 illustrates that the linear correlation coefficients r^2 of all bivariate flood variables are near zero, which indicates a very weak dependence. Therefore, we prefer to assume that these bivariate flood variables are all independent, and thus the joint probability distributions are calculated by the method shown in Equation 6, i.e., multiplying the probability densities of the univariate flood variables.

3.2. FFD Estimation for the River Network

3.2.1. FFD Estimation Based on the Assumed Theoretical Probability Distributions for Upstream Sites

The FFDs at the upstream gauging sites of Shigu, Gaochang, Beibei and Wulong are estimated by directly fitting the assumed theoretical probability distributions to the at-site flood samples. The optimally fitted theoretical probability distribution for each gauging site is determined by using the GAMLSS model with the covariate of RI. The PP plots in Figure S3 suggest that the probabilities calculated by the optimally fitted theoretical probability distributions are generally consistent with the empirical probabilities. The KS test indicates that these FFDs present a satisfactory fitting quantity with passing the GoF examination at the 0.05 significance level.

Table 1 displays the results of the covariate analysis for the flood variables at Shigu, Gaochang, Beibei and Wulong. For Gaochang and Wulong, the fitted theoretical probability distributions with the covariate of RI outperform those with constant statistical parameters. The location parameters (which refer to mean values of flood variables) of the FFDs at these two gauging sites have negative relationships with RI. This finding verifies the reservoir regulation effect of declining the flood mean values at Gaochang and Wulong.

Figure 5 displays the evolutions of the FFDs during the observation period from 1951 to 2015. It can be found that the covariate of RI is able to reasonably capture the changes of the observed flood samples,

Table 1
Results of the FFD Estimation Based on the Assumed Theoretical Distributions for the Flood Variables at Shigu, Gaochang, Beibei and Wulong

Hydrological gauging site	Distribution	μ	σ	BIC
Shigu	Gamma	4949.7	0.262	1121.4
Gaochang	Gamma	15756.4	0.296	1287.0
	Lognormal	$9.662 - 0.898\ln\left(\mathrm{RI}_g + 1\right)$	0.282	1283.0 (optimal)
Beibei	Weibull	25,280	3.400	1349.0
Wulong	Gamma	11418.6	0.362	1269.5
	Lognormal	$9.331 - 0.700\mathrm{RI}_w$	0.338	1261.9 (optimal)

Abbreviations: BIC, Bayesian information criterion; FFD, flood frequency distribution.

which generally occurred in the period after 2005. In particular, the flood mean value at Gaochang declines from 16,349 to 11,762 m^3/s, indicating an absolute reduction of 4,587 m^3/s and a relative reduction of 28%. For the flood at Wulong, the flood mean value decreases from 11,945 to 6,512 m^3/s, indicating an absolute reduction of 5,433 m^3/s as well as a relative reduction more than 45%. For the FFD at Beibei, none of the distribution parameters is related to the RI despite there are four reservoirs in the catchment above this gauging site (Figure 1). In a nonstationarity situation, the flood quantile corresponding to a given exceedance probability will change over time, since the FFD no longer holds constant (Obeysekera & Salas, 2016; Yan et al., 2017). To assess the effect of reservoir regulation on the hydrological designs at Gaochang and Lizhuang, Figure 5 also presents the evolutions of the flood quantiles of 75%, 90%, 95%, 98% and 99%. It is

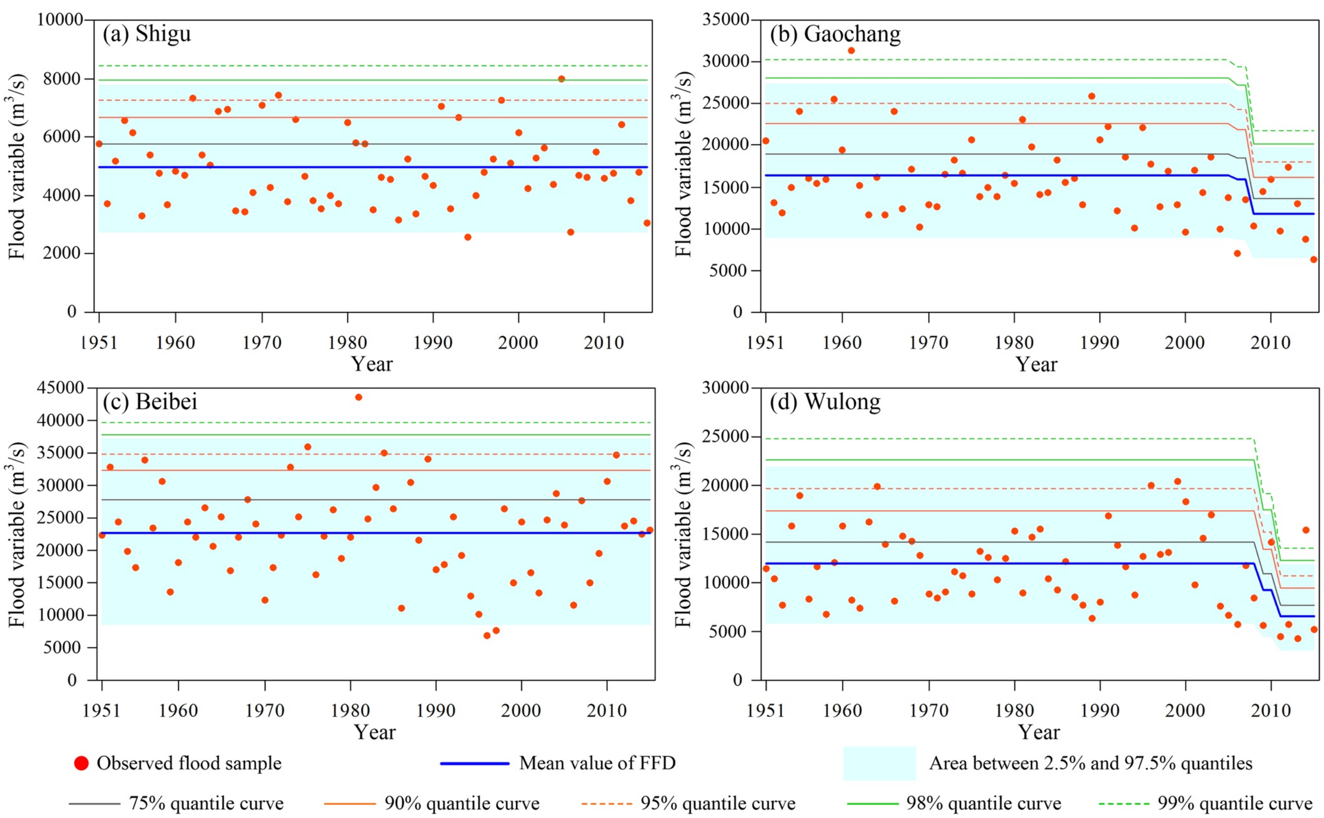

Figure 5. Evolutions of the optimally fitted theoretical probability distributions for the flood variables at the gauging sites of (a) Shigu, (b) Gaochang, (c) Beibei, and (d) Wulong, respectively.

Table 2
Results of Covariate Analysis for the Conditional Distributions of the Floods at Xiangjiaba, Lizhuang, Cuntan and Yichang in the Hierarchical Model

Hydrological gauging site	Distribution	μ	σ	BIC
Xiangjiaba	Lognormal	9.71	0.240	1269.4
	Lognormal	$4.562 + 0.608\ln\left(z_s\right)$	0.177	1234.7 (optimal)
Lizhuang	Gamma	25848.3	0.195	1299.5
	Gamma	$5832 + 0.802z_g + 0.438z_x$	$0.003 + 0.034e^{z_x/16971.7}$	1216.9 (optimal)
Cuntan	Gamma	46630.0	0.217	1392.3
	Gamma	$-261600 + 28840\ln\left(z_l\right) + 0.744z_b$	$0.048 + 0.027e^{z_b/22720.8}$	1330.9 (optimal)
Yichang	Normal	48348	8857	1374.4
	Normal	$-301784 + 30288\ln\left(z_c\right) + 3371\ln\left(z_w\right) - 5689e^{\mathrm{RI}_{c-y}}$	4494	1269.4 (optimal)

Abbreviation: BIC, Bayesian information criterion.

found that the reduction magnitude of flood quantiles generally gets to be enlarged with increasing cumulative probability. This finding indicates that the reservoir regulation tends to have a greater impact on the large floods at the upper tail.

3.2.2. FFD Estimation Based on the Hierarchical Model for Downstream Sites

Under the river network-based hierarchical model, the FFDs at the downstream sites of Xiangjiaba, Lizhuang, Cuntan and Yichang are derived by coupling the distributions conditioned on the upstream flood variables and RI with the probability distributions of the upstream floods. By using the GAMLSS model based on MLE, the conditional distributions of the flood variables at these four gauging sites are estimated, and the modeling results are summarized in Table 2. In agreement with the spatial correlation analysis for the flood variables (see Figure S1), there are positive relations between the location parameters of the conditional distributions and the upstream floods variables. Moreover, the scale parameters for Lizhuang and Cuntan are also found to be positively related to the upstream flood variables.

As shown in Figure 1, by the end of 2015, 10 reservoirs had been put into operation in the catchment between Shigu and Xiangjiaba, nevertheless, none of the parameters of the conditional distribution at Xiangjiaba is related to the RI. For the flood at Yichang, the location parameter of the conditional distribution presents a negative relation with the RI regarding the TGR, verifying the ability of this reservoir to reduce the downstream flood magnitude. The conditional distributions against with the observed flood samples are illustrated in Figure 6. It can be seen that the conditional distributions present noisy evaluations, which are able to capture the annual variability exhibited by the observed flood samples.

Figure S4 indicates that the flood probabilities calculated by the FFDs derived by the hierarchical model are generally consistent with the empirical probabilities. The *p*-values of the KS test suggest that all the derived FFDs pass the GoF test at the 0.05 significance level. Hence, the hierarchical model exhibits an acceptable performance in the Upper Yangtze basin. Figure 7 displays the evolutions of the FFDs derived by the hierarchical model in contrast to the observed flood samples. It is found that the mean values and quantile regions of the derived FFDs are able to reasonably capture the scatters of the observed flood samples. According to the modeling results, the FFD at Xiangjiaba presents a stable process, while the FFDs at Lizhuang, Cuntan and Yichang exhibit declines in mean values after 2005. The flood mean values at Lizhuang and Cuntan present similar reductions of about 4,000 m^3/s. In particular, the flood mean value at Lizhuang decreases from 26,392 to 22,710 m^3/s, and the flood mean value at Cuntan decreases from 48,402 to 44,123 m^3/s. During the period from 2005 to 2015, the flood mean value at Yichang falls from 50,070 to 35,378 m^3/s, resulting

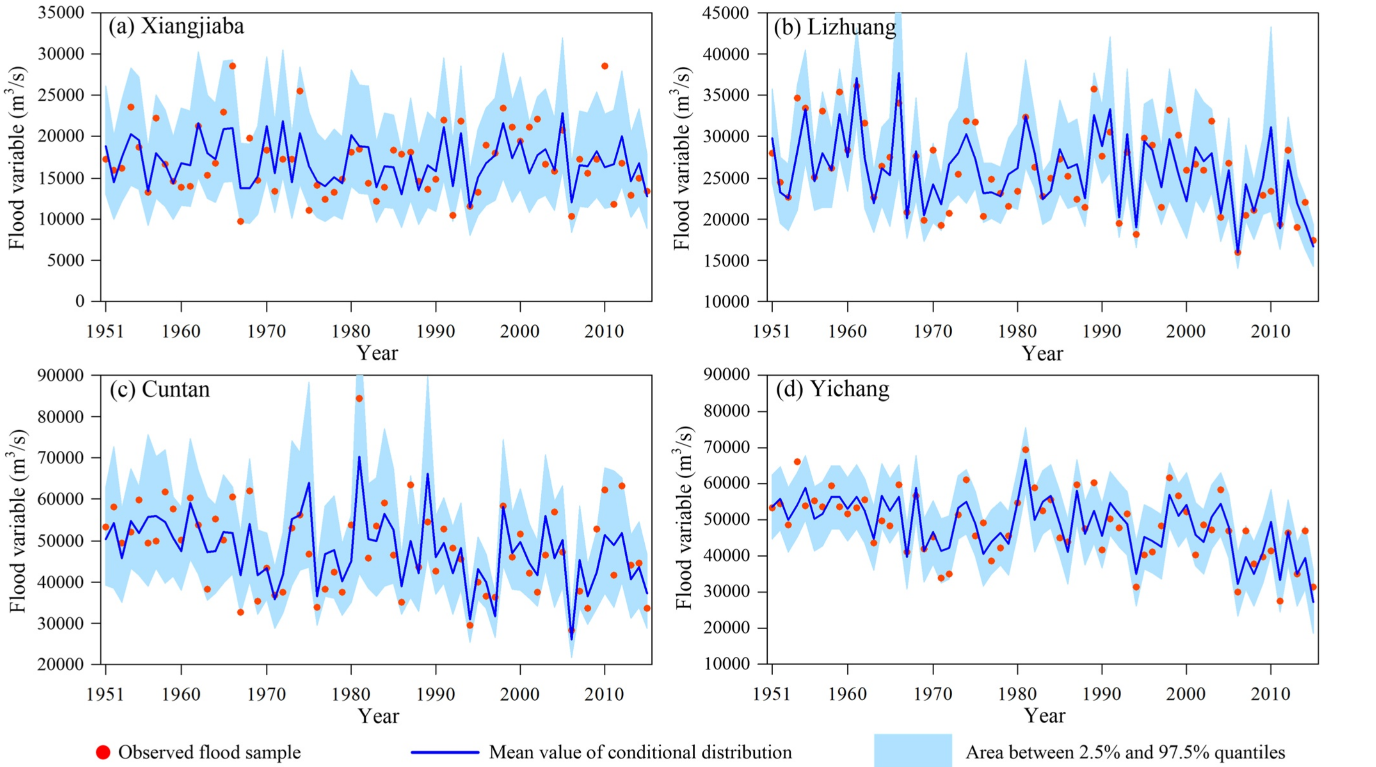

Figure 6. Conditional distributions of the flood variables at (a) Xiangjiaba, (b) Lizhuang, (c) Cuntan and (d) Yichang given the observed floods at the upstream gauging sites and reservoir index (RI).

in an absolute reduction of about 15,000 m^3/s, which is the largest absolute reduction among all the flood variables considered in this study. The flood quantiles with given typical probabilities also exhibit substantial declines, suggesting the benefit of reservoir regulation to mitigating the flooding risk in the basin.

3.3. Attribution of Flood Decline in the Upper Yangtze Basin

The cause-effect relationship between reservoir regulation and flood magnitude decline has been revealed in numerous basins around the world (Jiang, Xiong, et al., 2019; López & Francés, 2013; Merz & Blöschl, 2008b; Volpi et al., 2018; B. Xiong et al., 2020). With the consideration of the flood dependence within the river network, the hierarchical model has the advantage of being able to capture the transmission of flood characteristics along the river network. Thus the changes of the floods in the Upper Yangtze basin can easily be attributed to the relevant reservoirs. Figure 8 compares the probability density curves of the FFDs with and without the effect of reservoir regulation, which are represented by the FFDs of 1951 and 2015, respectively. Since the flood variable at Lizhuang is dependent on that at Gaochang, the regulation effect of the reservoirs above Gaochang can transfer to the FFD at Lizhuang. In the similar spirit, the regulation effect of the reservoirs above Gaochang can also transfer to the FFD at Cuntan via Lizhuang. The flood variable at Yichang is conditioned on the flood variables at Cuntan and Wulong, and therefore the flood declines at Cuntan and Wulong can both spread to Yichang. In natural condition before the regulation of the TGR, the flood magnitude at Yichang should generally be larger than that at the upstream gauging site of Cuntan (see panel (d) in Figure 8), since the former gauging site has a larger catchment area. But the probability density curves of the FFDs of 2015 suggest that the flood magnitude at Cuntan is generally larger than that at Yichang. This finding indicates that the TGR located between these two gauges has remarkably declined the flood magnitude at Yichang.

Some previous study revealed that the decline of the flood mean value at Yichang, which controls the discharge of the whole Upper Yangtze basin, is attributed to the joint regulation of multiple reservoirs, including the TGR and these reservoirs above the gauging sites of Wulong and Cuntan (Jiang, Zhang, & Luo, 2019;

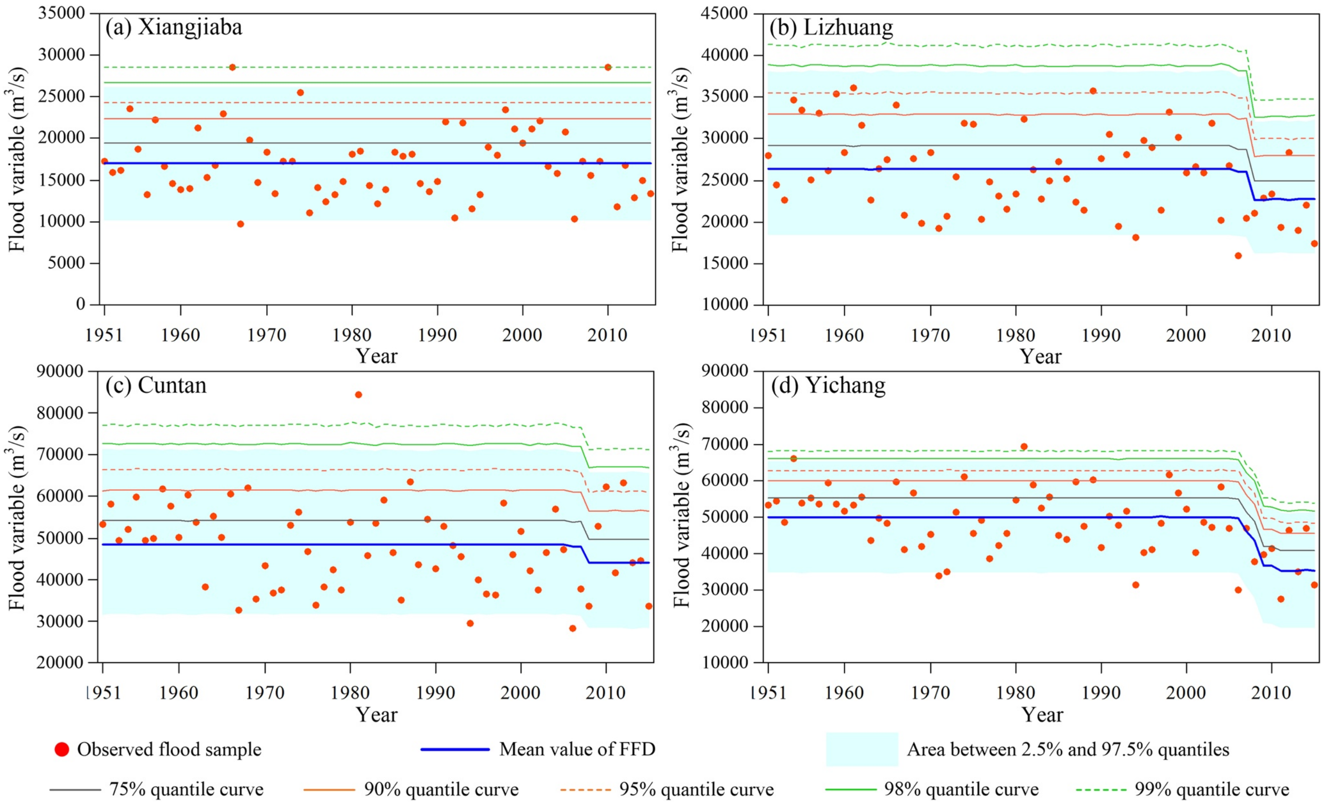

Figure 7. Evolutions of the flood frequency distributions (FFDs) derived by the hierarchical model. Panels (a), (b), (c) and (d) present the results for the gauging sites of Xiangjiaba, Lizhuang, Cuntan and Yichang, respectively.

Li et al., 2020; Wang et al., 2017). According to Equation 10, the mean value (or expectation) of the FFD at Yichang is given by:

$$E\left(Z_y\right) = E\left\{E\left[Z_y \mid \left(Z_c, Z_w, \mathrm{RI}_{c-y}\right)\right]\right\} \tag{17}$$

Table 2 illustrates that $Z_y \big| \left(Z_c, Z_w, \mathrm{RI}_{c-y}\right)$ follows a normal distribution, in which the location parameter is actually the expectation of $Z_y \big| \left(Z_c, Z_w, \mathrm{RI}_{c-y}\right)$. Thus, combining Equation 17 and the expression of the location parameter of the conditional distribution in Table 2, the mean value of the FFD at Yichang can be expressed by:

$$\begin{aligned} E\left(Z_y\right) &= E\left[-301784 + 30288\ln\left(Z_c\right) + 3371\ln\left(Z_w\right) - 5689e^{\mathrm{RI}_{c-y}}\right] \\ &= -301784 + 30288E\left[\ln\left(Z_c\right)\right] + 3371E\left[\ln\left(Z_w\right)\right] - 5689e^{\mathrm{RI}_{c-y}} \end{aligned} \tag{18}$$

Based on the above equation, it is easy to separate the contributions of different reservoirs to declining the flood mean value at Yichang. As shown in Figure 9, the flood mean value at Yichang presents a stepwise decline during the period from 2006 to 2015, when the reservoirs in the Upper Yangtze basin were put into operation in succession. The TGR was initially operated in 2006 and then put into full operation in 2009, resulting in apparent reductions in the flood magnitude at Yichang. The flood declines in 2008 and 2011 are mainly attributed to the reservoirs upstream Cuntan and Wulong, respectively. In general, the TGR contributes the majority of the reduction in the flood mean value at Yichang, followed by the reservoirs above Cuntan and those above Wulong. In 2015, the TGR leads to a flood reduction of 9,775 m^3/s, accounting for about 67% of the total reduction as well as about 20% of the flood mean value of 1951. The reservoirs above Cuntan induces a flood reduction of 2,853 m^3/s, i.e., a relative contribution of about 19% to the total reduction. The rest reduction induced by the reservoirs above Wulong is 2,045 m^3/s, making a relative contribution of 14% to the total reduction. This finding generally agrees with the previous studies, in which the TGR was also

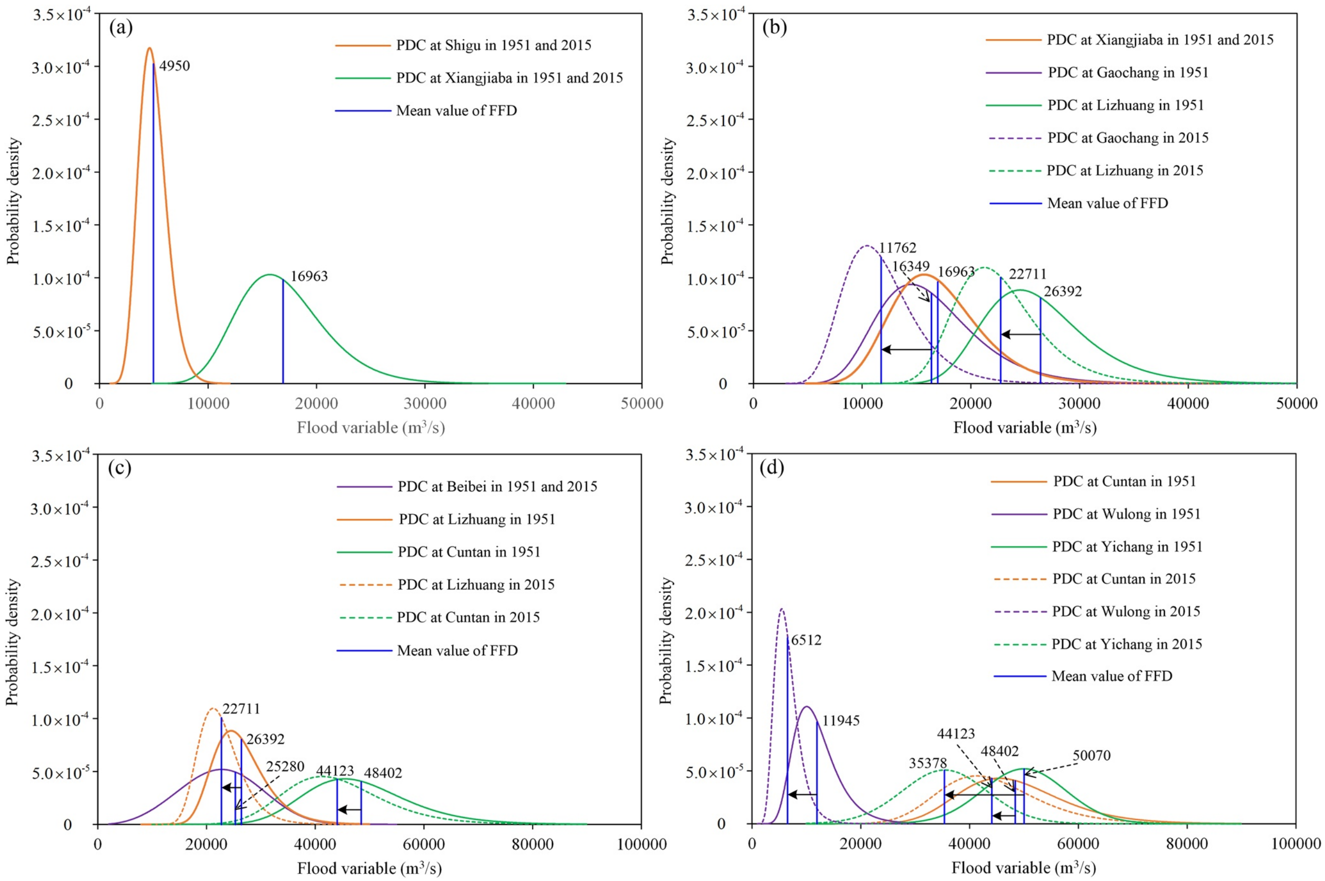

Figure 8. Changes of flood probability density curves (PDC) in the Upper Yangtze basin. The probability density curves in panels (a), (b), (c) and (d) correspond to the flood variables in hierarchies 1, 2, 3 and 4 in Figure 3, respectively.

found to have the largest contribution in declining the flood magnitude of the whole Upper Yangtze basin (Jiang, Zhang, & Luo, 2019; B. Xiong et al., 2020).

The reservoirs in the Upper Yangtze River basin normally follow a general operation rule: if the flood magnitude downstream a reservoir is forecasted to exceed a security threshold, the reservoir will be operated

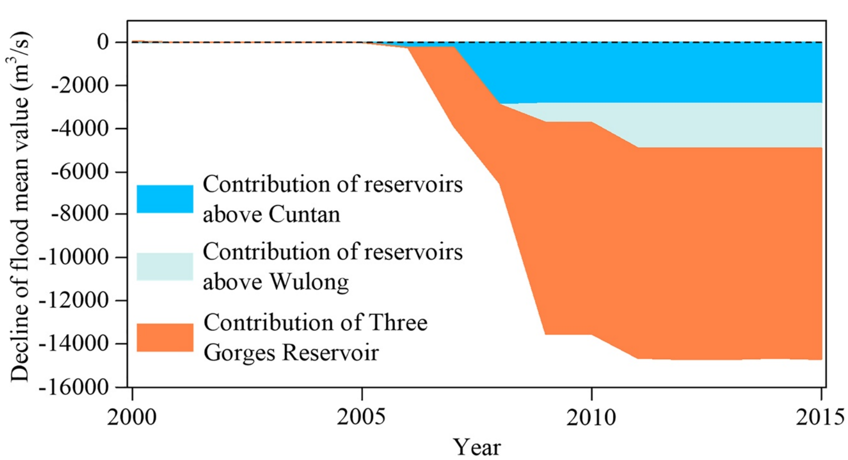

Figure 9. Contributions of reservoirs to reducing the flood mean value at Yichang.

to restrain the discharge out of the reservoir; and if the downstream flood magnitude is below the security threshold, the reservoir shall not be used to impound flood flow, guaranteeing the water level of the reservoir below a limited stage. In other words, to a certain extent, the specific operation strategy of a reservoir for flood control shall be determined by the flood risk of the downstream region. The TGR is designed to protect the flooding plain in the Middle Yangtze basin, which is more susceptible to river flooding than the Upper Yangtze basin. Because of the mountainous topography, the flood risk in the region upstream the TGR is relatively small. The TGR has a far more significant influence on downstream flood than other reservoirs do, not only due to its larger flood control capacity but a greater flood risk of the downstream region.

3.4. Uncertainty Analysis of the Rive-Network Based Hierarchical Model

3.4.1. Uncertainty Intervals of Flood Quantiles

In addition to MLE, the MCMC-based Bayesian approach is used to individually estimate the empirical posterior distributions of the model parameters at each node in the rive-network based hierarchical model. Figures S5–S12 present the histograms of the posterior distributions, each of which is actually characterized by a set of parameter samples. It is found that the model parameters estimated by MLE are generally around the medians of the posterior distributions, indicating the reliability of the MLE method. As displayed in panel (b) of Figure S7, panel (b) of Figure S11 and panel (d) of Figure S12, the posterior distributions of the model parameters related to reservoir indices are generally below zero. This finding further verifies that the reservoir regulation presents a significant effect of reducing the flood magnitude in the Upper Yangtze basin.

Given the empirical posterior distributions of model parameters, we can calculate the uncertainty intervals of flood quantiles at each hydrological gauging site. It is important to note that the uncertainty of the FFDs modeled by the theoretical distributions is only associated with model parameter estimation, and the uncertainty of the FFDs derived by the hierarchical model is not only due to the uncertainty in parameter estimation but also comes from numerical integration. For the hierarchical model, the uncertainty associated with numerical integration is calculated through repeating the numerical integration with the model parameters estimated by MLE 10,000 times. Figure 10 displays the 95% uncertainty intervals of flood quantiles, which are estimated based on the FFDs of 2015. It is found that the width of the uncertainty intervals due to numerical integration is very narrow, indicating that the uncertainty of the FFDs derived by the hierarchical model is mainly from model parameter estimation.

3.4.2. Propagation of Model Parameter Uncertainty Along the Network

According to the structure of the hierarchical model outlined in Figure 3, the uncertainty associated with model parameters can propagate along the river network downwards. For example, the uncertainty associated with the model parameters of the FFD at Shigu will spread to the FFDs at the downstream gauging sites of Xiangjiaba, Lizhuang, Cuntan and Yichang. Given both the model parameter samples at Shigu which are randomly drawn from the empirical posterior distributions and the model parameters estimated by MLE at the related gauging sites, the corresponding FFDs at Xiangjiaba, Lizhuang, Cuntan and Yichang can be estimated, respectively. By repeating this procedure 10,000 times, the uncertainty of the FFDs at these four gauging sites can be calculated. The results for the propagation of the uncertainty associated with the model parameters at Shigu are presented in Figure S13. The model parameter uncertainty of the FFD at Shigu presents pronounced effect on the FFD at Xiangjiaba, however, the effect on the FFDs at Cuntan and Yichang is almost invisible. The results for the propagation of the uncertainty associated with the model parameters of the other seven gauging sites are displayed in Figures S14–S20, all of which suggest that the influence of the model parameter uncertainty tends to substantially diminish along the river network downwards.

10.1029/2020WR029374

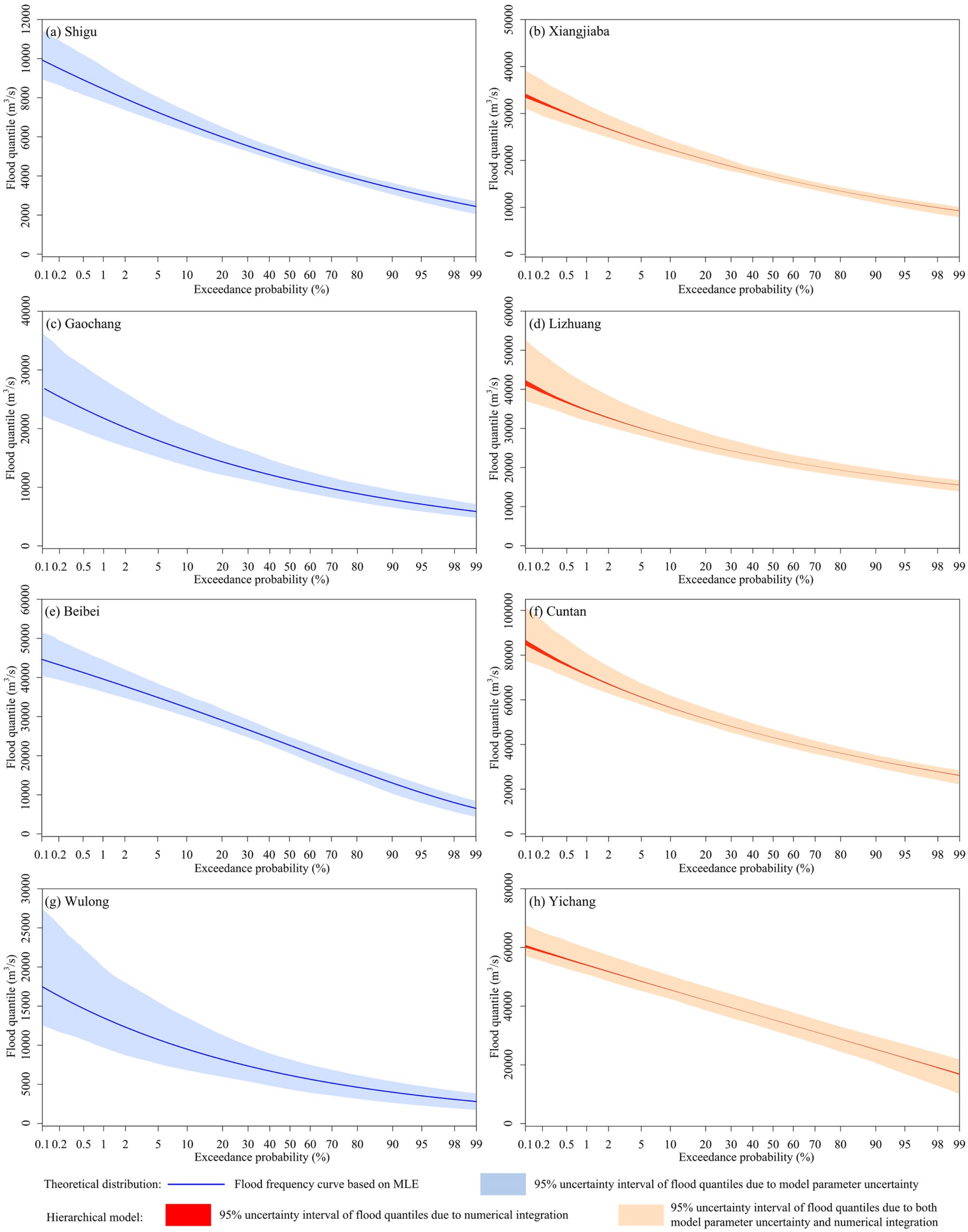

Figure 10. Uncertainty intervals of flood quantiles based on the flood frequency distributions (FFDs) of 2015.

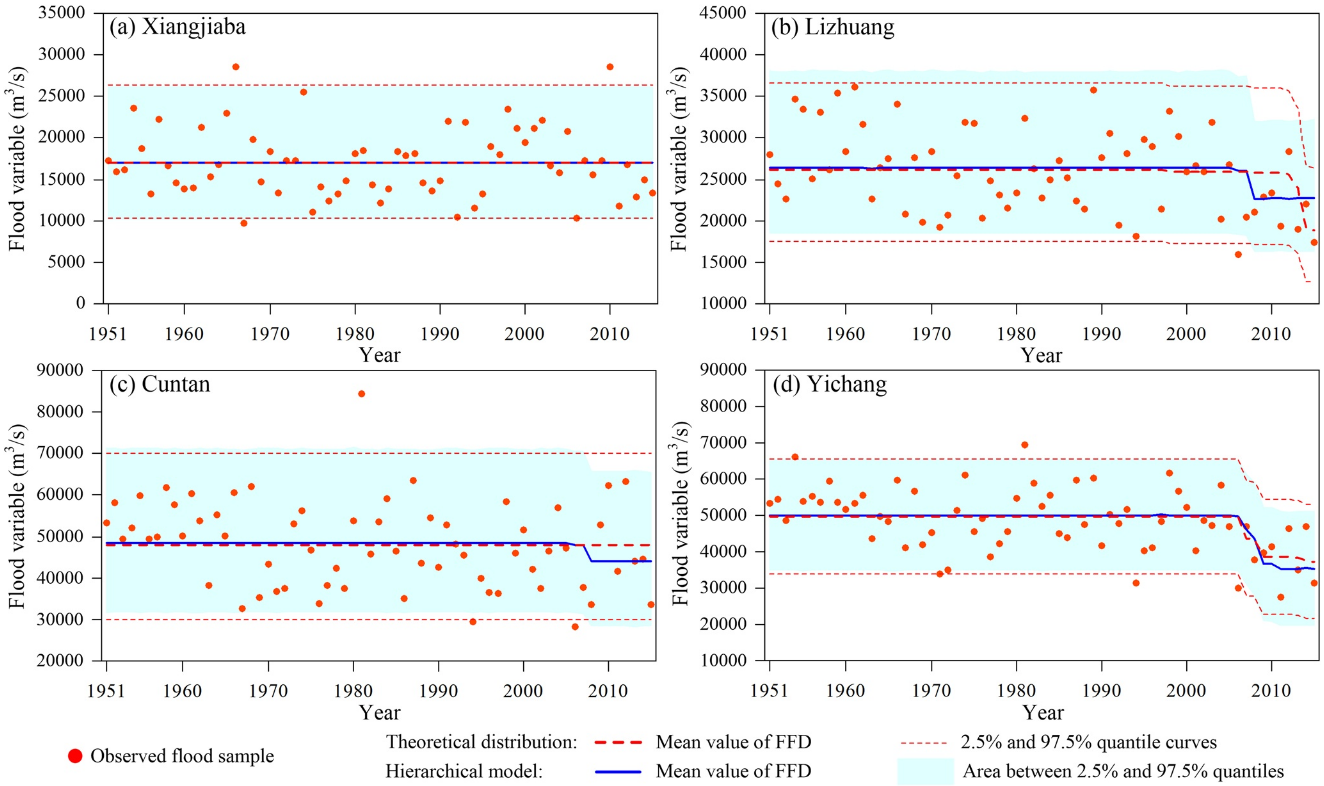

Figure 11. Evolutions of the flood frequency distributions (FFDs) both derived by the hierarchical model and by fitting the theoretical probability distributions to the at-site flood samples.

4. Discussion

4.1. Comparison Between the Hierarchical Model and the Assumed Theoretical Probability Distributions

In addition to the proposed river network-based hierarchical model, the FFDs at Xiangjiaba, Lizhuang, Cuntan and Yichang can also be estimated by fitting the assumed theoretical probability distributions to the observed flood samples. The GAMLSS model with the covariate of RI is also employed to obtain the optimal theoretical probabilities to fit the at-site observed flood samples. Figure 11 compares the FFDs derived by the hierarchical model and the optimally fitted theoretical distributions in describing the evolutions of the observed flood samples. It can be seen that these two methods exhibit similar modeling results for the floods at Xiangjiaba and Yichang (see panels (a) and (d) in Figure 11), but show apparent discrepancies between the modeling results for the floods at Lizhuang and Cuntan (see panels (b) and (c) in Figure 11). Intuitively, the river network-based hierarchical model has a more reasonable performance in capturing the evolution of the observed flood samples at Lizhuang (panel (b) in Figure 11). Since the flood variable at Cuntan is significantly related to that at Lizhuang, the FFD at Cuntan should present an evolution similar to the FFD at Lizhuang due to reservoir regulation. However the fitted theoretical distribution for the flood at Cuntan suggests a stationary process. This indicates that the hierarchical model considering the flood dependence within the river network should be more effective in identifying the effect of reservoir regulation on the flood at Cuntan.

The results of uncertainty analysis for the FFDs both derived by the hierarchical model and by directly fitting theoretical distributions to the at-site flood samples are displayed in Figure 12. In this figure, the uncertainty intervals of flood quantiles are calculated by the FFDs of 2015. Compared to the fitted theoretical distributions, the FFDs derived by the hierarchical model generally present narrower uncertainty intervals of flood quantiles. This finding indicates that the hierarchical model is able to reduce the uncertainty in flood quantile estimation by incorporating the flood dependence within the river network, even though it

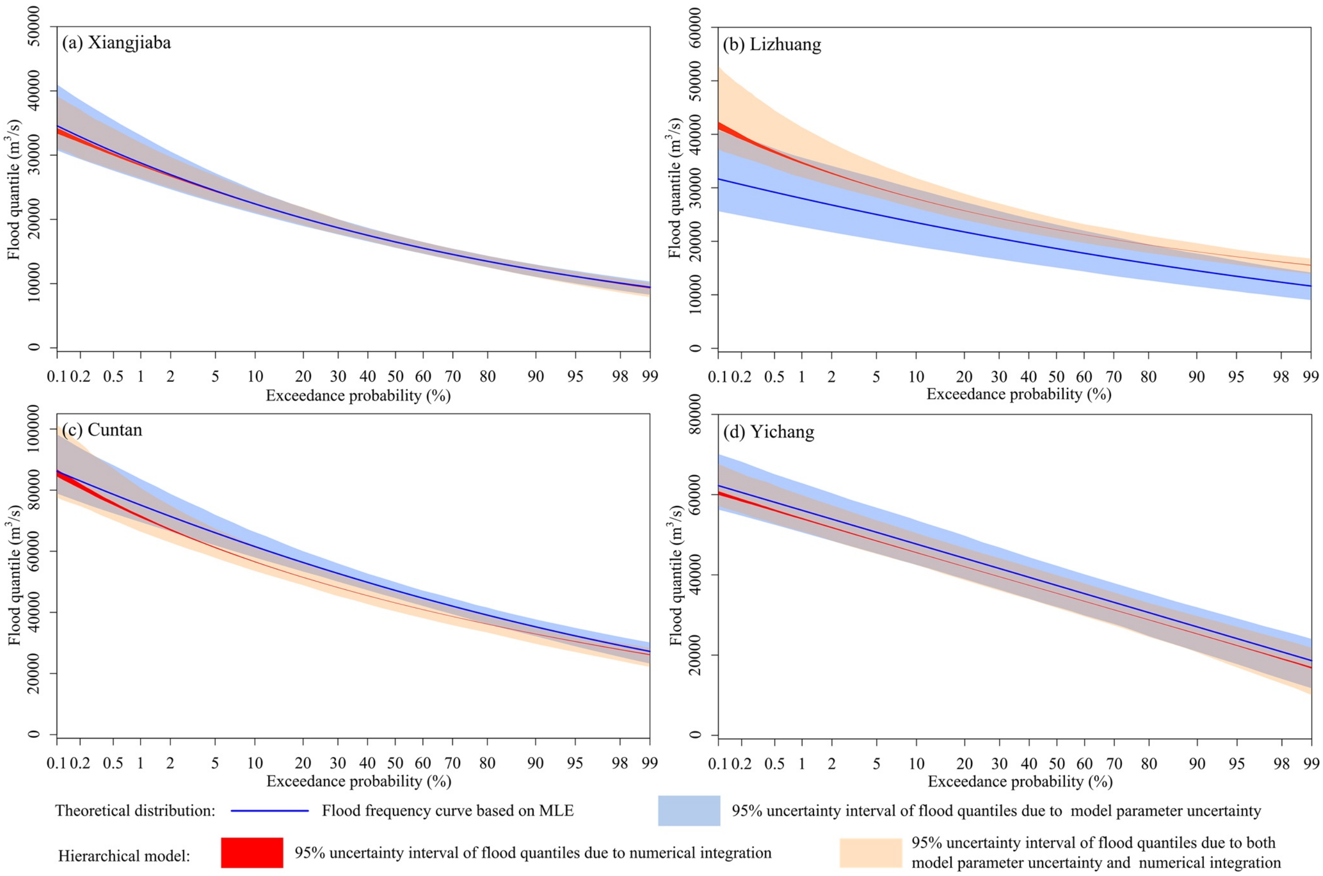

Figure 12. Uncertainty intervals of flood quantiles derived by both the hierarchical model and the fitted theoretical probability distributions. The flood quantiles are calculated from the flood frequency distributions (FFDs) of 2015.

has a complex structure with more parameters. Ravindranath et al. (2019) also found that a spatial river network structure produced a substantial reduction in the uncertainty associated with paleo-streamflows while reconstructing the streamflow in the Upper Missouri River. Merz and Blöschl (2008a) grouped the meaningful additional information in flood frequency analysis into three types: temporal, spatial and causal expansions. According to this concept, the flood dependence within the river network should be the additional information regarding spatial and causal expansions.

In this study, the proposed hierarchical model presents an advantage over the conventional FFD estimation method based on the assumed theoretical distributions in estimating the FFDs in the Upper Yangtze basin, due to the capturing of the flood dependence within the river network. However, the hierarchical model is still a statistical approach, which is unable to consider the physical factors dominating flood dependence, such as the distance between gauging sites, the timing of the floods, the travel time between floods, and the specific rule of reservoir operation. The above limitation could be a challenge for the hierarchical model while being employed to a more complex and larger river network with numerous gauging sites and reservoirs.

4.2. Effect of Reservoir Regulation on Floods

Although all the reservoirs considered in this study have considerable capacities of flood control, the modeling results in Tables 1 and 2 indicate that the flood variables at Xiangjiaba and Beibei are independent of the covariates of reservoir indices. This finding does not necessarily mean that these reservoirs make no difference to the downstream flood processes. As displayed in Table S2, most reservoirs in the Upper Yangtze basin were put into operation after 2010, whereas the flood samples used in this study are end in 2015.

The changes of the flood variables caused by reservoir regulation may be not statistically significant and are unlikely to be detected by a statistical analysis, since the flood samples under reservoir regulation are too short. For example, if the flood samples at Shigu and Xiangjiaba are extended from 2015 to 2019, we will obtain a different modeling result of the covariate analysis for the flood variable at Xiangjiaba. As displayed in Table S4, the flood samples with different lengths lead to different modeling results of covariate selection for the RI. The modeling results for the samples from 1951 to 2015 indicate that the RI is not selected as a significant covariate for the flood variable at Xiangjiaba in terms of the optimal value of BIC. For the flood samples from 1951 to 2019, the GAMLSS model with the covariate of RI outperforms the model without this covariate in terms of BIC. The results in Table S4 suggest that, using the extended flood samples, the RI is chosen as a significant covariate for the location parameter of the conditional distribution of the flood at Xiangjiaba, and therefore reservoir regulation shall have a significant impact on the flood at Xiangjiaba. In order to examine the impacts of perturbation from calibrated parameter values, or the impacts of parameter calibration errors, an uncertainty analysis for the regression coefficient for the RI is done by using the MCMC-based Bayesian approach. The empirical posterior distribution of the regression coefficient for the RI is displayed in Figure S21. For the flood samples from 1951 to 2019, the posterior distribution of the regression coefficient is generally below zero, suggesting a significantly negative relationship between the flood at Xiangjiaba and the RI. For the flood samples from 1951 to 2015, the median of the posterior distribution of the regression coefficient for the RI is near to zero, suggesting a negligible relationship between the flood variable at Xiangjiaba and the RI. This finding verifies the results of covariate selection for the RI and the length of flood samples shall be an important factor affecting the modeling results of covariate analysis.

The above findings also indicate that the RI based on storage capacity and catchment area might fail to reflect the complete effect of reservoir regulation on floods. In practice, the flood control ability of a reservoir could be dependent on multiple factors, including its storage capacity, catchment area, reservoir position, operation rule and even the flood regime in the basin (Ayalew et al., 2013; S. Gao et al., 2019; Volpi et al., 2018; B. Xiong et al., 2019). The effect of reservoir regulation on floods requires a deeper investigation with involving more reasoning information about reservoir characteristics.

5. Conclusion

In this study, we present a hierarchical model to estimate the FFDs by considering both the flood dependence within the river network and the effect of reservoir regulation on flood characteristics. Under this hierarchical model, the FFD is not directly modeled by an assumed theoretical distribution relying only on local at-site flood samples, but arises from a hierarchy, which combines the conditional distribution of the interested flood variable given both its upstream flood variables and the RI with the probability distribution of the upstream flood variables.

The application of the proposed hierarchical model to the Upper Yangtze basin suggests a satisfactory performance in estimating the FFDs, which not only exhibit acceptable fitting quality but also reasonably capture the effect of reservoir regulation on floods. The reservoir regulation especially that of the TGR is found to remarkably decline the flood magnitude in the Upper Yangtze basin. In addition, the uncertainty analysis of the hierarchical model shows that the influence of the uncertainty associated with model parameters tends to substantially diminish along the river network downwards.

Compared to the conventional FFD estimation method that directly fits the assumed theoretical probability distributions to the at-site flood samples, the river network-based hierarchical model incorporating the additional information of the flood dependence within the river network is more capable of modeling the effect of reservoir regulation on the floods in the Upper Yangtze basin. On the other hand, the hierarchical model produces a reduction in the uncertainty in flood quantile estimation, even though it has a complex structure with more parameters. This finding indicates that it is worthy of expanding additional information with hydrological reasoning in FFD estimation.

Data Availability Statement

The flood data used in this study are collected by Bureau of Hydrology, Changjiang Water Resources Commission, China, and available under request. Most of the flood data used in this study are also available at the website of Hydrology and Water Researches Center of Hubei (http://113.57.190.228:8001/#!/web/Report/RiverReport).

Acknowledgments

This research is financially supported jointly by the National Natural Science Foundation of China (NSFC Grants 51809243 and 41890822), the Research Council of Norway (FRINATEK Project 274310), and the Ministry of Education "Plan 111" Fund of China (B18037), all of which are greatly appreciated.

References

Ayalew, T. B., Krajewski, W. F., & Mantilla, R. (2013). Exploring the effect of reservoir storage on peak discharge frequency. *Journal of Hydrologic Engineering*, *18*(12), 1697–1708. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000721

Casella, G., & Berger, R. L. (2002). *Statistical Inference*. Thomson Learning.

Dawdy, D., Griffis, V., & Gupta, V. (2012). Regionalood-frequency analysis: How we got here and where we are going. *Journal of Hydrologic Engineering*, *17*(9), 953–959. https://doi.org/10.1061/(asce)he.1943-5584.0000584

Debele, S. E., Strupczewski, W. G., & Bogdanowicz, E. (2017). A comparison of three approaches to non-stationary flood frequency analysis. *Acta Geophysica*, *65*(4), 863–883. https://doi.org/10.1007/s11600-017-0071-4

Dingman, S. L. (2015). *Physical hydrology*. Long Grove IL. Waveland press.

Eagleson, P. S. (1972). Dynamics of flood frequency. *Water Resources Research*, *8*(4), 878–898. https://doi.org/10.1029/WR008i004p00878

El Adlouni, S., Ouarda, T. B., Zhang, X., Roy, R., & Bobée, B. (2007). Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research*, *43*, W03410. https://doi.org/10.1029/2005WR004545

Engeland, K., Wilson, D., Borsányi, P., Roald, L., & Holmqvist, E. (2018). Use of historical data in flood frequency analysis: A case study for four catchments in Norway. *Hydrology Research*, *49*, 466–486. https://doi.org/10.2166/nh.2017.069

Frank, J., & Massey, J. R. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, *46*(253), 68–78.

Gao, B., Yang, D., & Yang, H. (2013). Impact of the Three Gorges Dam on flow regime in the middle and lower Yangtze River. *Quaternary International*, *304*, 43–50. https://doi.org/10.1016/j.quaint.2012.11.023

Gao, S., Liu, P., Pan, Z., Ming, B., Guo, S., Cheng, L., & Wang, J. (2019). Incorporating reservoir impacts into flood frequency distribution functions. *Journal of Hydrology*, *568*, 234–246. https://doi.org/10.1016/j.jhydrol.2018.10.061

Guo, S. L., & Cunnane, C. (1991). Evaluation of the usefulness of historical and palaeological floods in quantile estimation. *Journal of Hydrology*, *129*(1–4), 245–262. https://doi.org/10.1016/0022-1694(91)90053-K

Hosking, J. R. M. (Ed.), (1990). L-Moments: Analysis and estimation of distributions using linear combinations of order statistics. *American Statistician*, *52*, 105–124. http://www.jstor.org/stable/2345653

Hosking, J. R. M., & Wallis, J. R. (1997). *Regional frequency analysis*. UK. Cambridge University Press.

Jiang, C., Xiong, L., Yan, L., Dong, J., & Xu, C. Y. (2019). Multivariate hydrologic design methods under nonstationary conditions and application to engineering practice. *Hydrology and Earth System Sciences*, *23*(3), 1683–1704. https://doi.org/10.5194/hess-23-1683-2019

Jiang, C., Zhang, Q., & Luo, M. (2019). Assessing the effects of the Three Gorges Dam and upstream inflow change on the downstream flow regime during different operation periods of the dam. *Hydrological Processes*, *33*(22), 2885–2897. https://doi.org/10.1002/hyp.13535

Khaliq, M. N., Ouarda, T. B. M. J., Ondo, J. C., Gachon, P., & Bobée, B. (2006). Frequency analysis of a sequence of dependent and/or nonstationary hydrometeorological observations: A review. *Journal of Hydrology*, *329*, 534–552. https://doi.org/10.1016/j.jhydrol.2006.03.004

Kroll, C. N., & Stedinger, J. R. (1998). Regional hydrologic analysis: Ordinary and generalized least squares revisited. *Water Resources Research*, *34*(1), 121–128. https://doi.org/10.1029/97wr02685

Li, H., Liu, P., Guo, S., Cheng, L., & Yin, J. (2020). Climatic control of upper Yangtze River flood hazard diminished by reservoir groups. *Environmental Research Letters*, *15*(12), 124013. https://doi.org/10.1088/1748-9326/abc4fe

López, J., & Francés, F. (2013). Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates. *Hydrology and Earth System Sciences*, *17*(8), 3189–3203. https://doi.org/10.5194/hess-17-3189-2013

Merz, R., & Blöschl, G. (2008a). Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information. *Water Resources Research*, *44*, W08432. https://doi.org/10.1029/2007WR006744

Merz, R., & Blöschl, G. (2008b). Flood frequency hydrology: 2. Combining data evidence. *Water Resources Research*, *44*, W08433. https://doi.org/10.1029/2007WR006745

Naghettini, M., Potter, K. W., & Illangasekare, T. (1996). Estimating the upper tail of flood-peak frequency distributions using hydrometeorological information. *Water Resources Research*, *32*(6), 1729–1740. https://doi.org/10.1029/96WR00200

Nelsen, R. B. (2006). *An introduction to copulas*. NY: Springer.

Niederreiter, H. (1978). Quasi-Monte Carlo methods and pseudo-random numbers. *Bulletin of the American Mathematical Society*, *197*(84), 957–1041. https://doi.org/10.1090/s0002-9904-1978-14532-7

Obeysekera, J., & Salas, J. (2016). Frequency of recurrent extremes under nonstationarity. *Journal of Hydrologic Engineering*, *21*(5), 04016005. https://doi.org/10.1061/(asce)he.1943-5584.0001339

Ravindranath, A., Devineni, N., Lall, U., Cook, E. R., Pederson, G., Martin, J., & Woodhouse, C. (2019). Streamflow reconstruction in the upper Missouri River basin using a novel Bayesian network model. *Water Resources Research*, *55*, 7694–7716. https://doi.org/10.1029/2019WR024901

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, *54*(3), 507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x

Salvadori, G., & De Michele, C. (2010). Multivariate multiparameter extreme value models and return periods: A copula approach. *Water Resources Research*, *46*, W10501. https://doi.org/10.1029/2009WR009040

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464. https://doi.org/10.1016/s0079-6638(08)70079-9

Sivapalan, M., Blöschl, G., Merz, R., & Gutknecht, D. (2005). Linking flood frequency to long-term water balance: Incorporating effects of seasonality. *Water Resources Research*, *41*, W06012. https://doi.org/10.1029/2004WR003439

Steinschneider, S., & Lall, U. (2015). A hierarchical Bayesian regional model for nonstationary precipitation extremes in Northern California conditioned on tropical moisture exports. *Water Resources Research*, *51*, 1472–1492. https://doi.org/10.1002/2014WR016664

Thorarinsdottir, T. L., Hellton, K. H., Steinbakk, G. H., Schlichting, L., & Engeland, K. (2018). Bayesian regional flood frequency analysis for large catchments. *Water Resources Research*, *54*, 6929–6947. https://doi.org/10.1029/2017WR022460

Toonen, W. H. (2015). Flood frequency analysis and discussion of non-stationarity of the Lower Rhine flooding regime (AD 1350–2011): Using discharge data, water level measurements, and historical records. *Journal of Hydrology*, *528*, 490–502. https://doi.org/10.1016/j.jhydrol.2015.06.014

Villarini, G., Smith, J. A., & Napolitano, F. (2010). Nonstationary modeling of a long record of rainfall and temperature over Rome. *Advances in Water Resources*, *33*, 1256–1267. https://doi.org/10.1016/j.advwatres.2010.03.013

Villarini, G., Smith, J. A., Serinaldi, F., Bales, J., Bates, P. D., & Krajewski, W. F. (2009). Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. *Advances in Water Resources*, *32*(8), 1255–1266. https://doi.org/10.1016/j.advwatres.2009.05.003

Vogel, R. M., Yaindl, C., & Walter, M. (2011). Nonstationarity: Flood magnification and recurrence reduction factors in the United States. *Journal of the American Water Resources Association*, *47*(3), 464–474. https://doi.org/10.1111/j.1752-1688.2011.00541.x

Volpi, E., Di Lazzaro, M., Bertola, M., Viglione, A., & Fiori, A. (2018). Reservoir effects on flood peak discharge at the catchment scale. *Water Resources Research*, *54*, 9623–9636. https://doi.org/10.1029/2018WR023866

Vrugt, J. A., Ter Braak, C., Diks, C., Robinson, B. A., Hyman, J. M., & Higdon, D. (2009). Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, *10*(3), 273–290. https://doi.org/10.1515/ijnsns.2009.10.3.273

Wang, Y., Wang, D., Lewis, Q. W., Wu, J., & Huang, F. (2017). A framework to assess the cumulative impacts of dams on hydrological regime: A case study of the Yangtze River. *Hydrological Processes*, *31*(17), 3045–3055. https://doi.org/10.1002/hyp.11239

Xiong, B., Xiong, L., Guo, S., Xu, C.-Y., Xia, J., Zhong, Y., & Yang, H. (2020). Nonstationary frequency analysis of censored data: A case study of the floods in the Yangtze River from 1470 to 2017. *Water Resources Research*, *56*, e2020WR027112. https://doi.org/10.1029/2020WR027112

Xiong, B., Xiong, L., Xia, J., Xu, C. Y., Jiang, C., & Du, T. (2019). Assessing the impacts of reservoirs on downstream flood frequency by coupling the effect of scheduling-related multivariate rainfall with an indicator of reservoir effects. *Hydrology and Earth System Sciences*, *23*(11), 4453–4470. https://doi.org/10.5194/hess-23-4453-2019

Xiong, L., Jiang, C., Xu, C.-Y., Yu, K.-X., & Guo, S. (2015). A framework of change-point detection for multivariate hydrological series. *Water Resources Research*, *51*, 8198–8217. https://doi.org/10.1002/2015WR017677

Yan, L., Xiong, L., Guo, S., Xu, C.-Y., Xia, J., & Du, T. (2017). Comparison of four nonstationary hydrologic design methods for changing environment. *Journal of Hydrology*, *551*, 132–150. https://doi.org/10.1016/j.jhydrol.2017.06.001

Yu, G., Wright, D. B., Zhu, Z., Smith, C., & Holman, K. D. (2019). Process-based flood frequency analysis in an agricultural watershed exhibiting nonstationary flood seasonality. *Hydrology and Earth System Sciences*, *23*(5), 2225–2243. https://doi.org/10.5194/hess-23-2225-2019