Accepted version

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

Nonstationary Regional Flood Frequency Analysis Based on the Bayesian Method

Shuhui Guo[a], Lihua Xiong[a,*], Jie Chen[a], Shenglian Guo[a], Jun Xia[a], Ling Zeng[b], Chong-Yu Xu[c]

[a] *State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, P. R. China; Pearl River Water Resources Commission Technology Center, Guangzhou 510630, P. R. China.* [b] *Bureau of Hydrology, Changjiang Water Resources Commission, Wuhan 430010, P. R. China.* [c]*Department of Geosciences, University of Oslo, P.O. Box 1022 Blindern, N-0315 Oslo, Norway.*

E-mail addresses:

S. Guo (shguo1@whu.edu.cn)

L. Xiong (xionglh@whu.edu.cn)

J. Chen (jiechen@whu.edu.cn)

S. Guo (slguo@whu.edu.cn)

J. Xia (xiajun666@whu.edu.cn)

L. Zeng (zengling128@hotmail.com)

C-Y. Xu (c.y.xu@geo.uio.no)

** Corresponding author:*

Lihua Xiong

State Key Laboratory of Water Resources and Hydropower Engineering Science

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

Wuhan University, Wuhan 430072, P. R. China

E-mail: xionglh@whu.edu.cn

Abstract

Most researches on regional flood frequency analysis (RFFA) have proved that the incorporation of hydrologic information (e.g., catchment attributes and flood records) from different sites in a region can provide more accurate flood estimation than using only the observed flood series at the site of concern. One kind of RFFA is based on the Bayesian method with prior information inferred from regional regression by using the generalized least squares (GLS) model, which is more flexible than other RFFA methods. However, the GLS model for regional regression is a stationary method and not suitable for coping with nonstationary prior information. In this study, in nonstationary condition, the Bayesian RFFA with the prior information inferred from regional regression by using the linear mixed effect (LME) model (i.e. a model that adds random effects to the GLS model) is investigated. Both the GLS-based and LME-based Bayesian RFFA methods have been applied to four hydrological stations within the Dongting Lake basin for comparison, and the results show that the performance of nonstationary LME-based Bayesian RFFA method is better than that of stationary GLS-based Bayesian RFFA method according to the deviance information criterion (DIC). Compared with the stationary GLS-based Bayesian RFFA method, changes in uncertainty of regression coefficients estimation of at-site flood distribution parameters are different from site to site by using the nonstationary LME-based Bayesian RFFA method. The use of nonstationary LME-based Bayesian RFFA method reduces design flood uncertainty, especially for the very small exceedance probability at the tail. This

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

study extends the application of the Bayesian RFFA method to the nonstationary condition, which is helpful for nonstationary flood frequency analysis of ungauged sites.

Keywords: Catchment attributes; regional regression; GLS model; LME model; prior probability distribution; posterior probability distribution

1 Introduction

Flood frequency analysis is very important for hydrological design (e.g., the capacities of reservoir, levees and spillways) and risk management (Reinders and Munoz, 2021; Razmi et al., 2022).

Improving the accuracy of flood frequency estimation is essential to ensure the safety and economy of hydraulic engineering design (Merz and Blöschl, 2008a, 2008b; Pandey et al., 2020; Razmkhah et al., 2022; Viglione et al., 2013). Numerous studies have proven that the combination of hydrological information from different sites in a region can provide more accurate hydrology estimation at a specific site and even making inferences at ungauged hydrological sites, which is the regional flood frequency analysis (RFFA) (Allahbakhshian-Farsani et al., 2020; Kuczera, 1982; Merz and Blöschl, 2008a; Han et al., 2022).

Existing studies on RFFA aim to combine regional information with at-site flood records in two main ways: one is directly use regional information as covariant of the statistics of at-site flood samples (e.g. regional regression method and index flood method) (Gregersen et al., 2017; Gao et al., 2021; Reis et al., 2020), and the other is to use Bayesian method to associate regional information as prior information with at-site flood samples information (Kuczera, 1982; Madsen and Rosbjerg, 1997; Vicens et al., 1975). In terms of directly using the regional information, there are many studies that use regional regression models to link regional covariates to flood sample statistics. In terms of using regional information as prior information, existing studies are generally based on the Bayesian approach, where the prior information of distribution parameter of site-specific flood series is expressed as a probability distribution (i.e., the prior distribution), the prior information is combined with site-specific information to obtain an updated distribution (i.e., the posterior distribution). The parameter prior distribution can be obtained by using the regional regression model that relates prior distribution parameter to catchment attributes (Reis et al., 2020; Thomas and Benson, 1970; Griffis and Stedinger, 2007; Jaffres et al., 2022; Merz and Bloschl, 2005). For example, Cunnane and Nash (1971) proposed an empirical Bayesian T-year event estimator based on the Gumbel distribution, where the mean and coefficient of variation were expressed in terms of catchment area, average annual rainfall, and catchment slop. In contrast to the regional regression and index flood methods, the Bayesian method relaxes the constraints on homogeneous regions in the index flood method by allowing the covariates of frequency distribution parameters to be arbitrary under nonstationary conditions, not just time.

Due to the climate conditions changes and intensive human activities alter natural hydrological cycle regimes, the stationary assumption of hydrological series in the traditional flood frequency analysis has been widely questioned, and therefore most scholars have extensively developed research on nonstationary flood frequency analysis (Han et al., 2022; Jiang et al., 2019; Milly et al., 2008; Razmi et al. 2022; Villarini et al., 2009; Wang et al., 2022; Guo et al., 2021). In stationary RFFA, the flood distribution parameters are assumed constant, the prior information of the flood distribution parameters can be inferred from the regional regression by using the GLS model (Reis et al., 2005). In nonstationary RFFA, the flood distribution parameters are assumed to be time-varying, and the GLS model is not applicable to regional regression of time-varying distribution parameters. Therefore, how to utilize the Bayesian method for nonstationary RFFA needs to be carefully investigated.

The goal of this study is to study the nonstationary RFFA based on the Bayesian method. This paper is structured as follows. Section 2 presents the nonstationary RFFA based on the Bayesian method. Section 3 introduces the study area and data. Section 4 presents the results and discussion of the application of the method. Section 5 presents the conclusion.

2 Methods

In nonstationary condition, RFFA based on the Bayesian method using the prior information inferred from regional regression by using the linear mixed effect (LME) model is investigated. First, the stationary GEV distribution model (SG) and the nonstationary GEV distribution model (NG) for floods are introduced. Second, under the stationary condition, the regional regression prior inferred from the generalized least squares (GLS) model of the at-site regression coefficients of flood distribution parameters are obtained. Third, under the nonstationary condition, the regional regression prior inferred from the LME model of the at-site regression coefficients of flood distribution parameters are obtained. Fourth, Bayesian theory is used to combine the two different prior information with the flood sample information, and the posterior probability distribution of the at-site regression coefficients are obtained. Finally, both the stationary GLS-based and nonstationary LME-based Bayesian RFFA methods (i.e.,

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

SG-GLS and NG-LME) are applied to four hydrological stations within the Dongting Lake basin to compare the results of the two methods. The flowchart of nonstationary RFFA based on the Bayesian method is shown in Fig. 1.

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

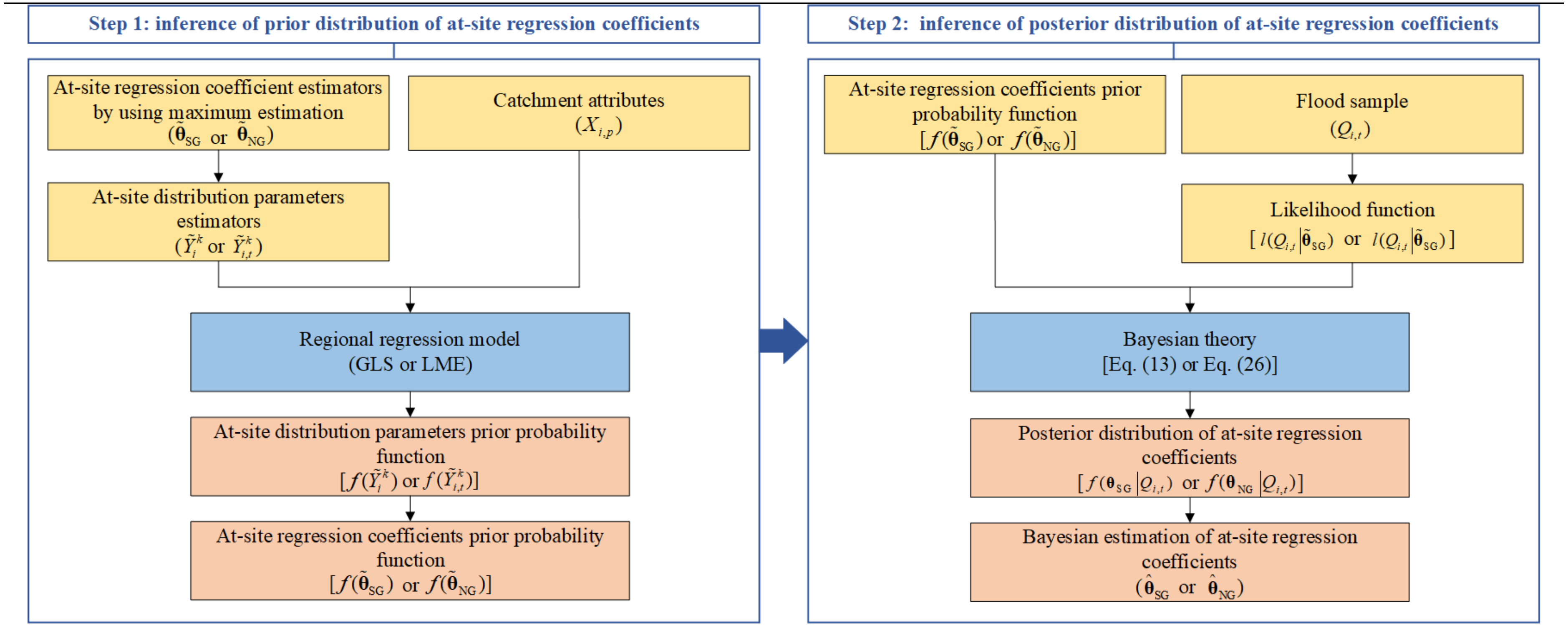

Fig. 1 Flowchart of regional flood frequency analysis based on the Bayesian method

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

2.1 GEV distribution model of flood series

In this study, the annual maximum daily streamflow series is taken as the flood series. Let flood series $Q_{i,t}$ (t=1,…,T) at the i-th site follow the generalized extreme value (GEV) (El Adlouni et al., 2007; Martins and Stedinger, 2000) distribution with a density function $f\left(Q_{i,t}\left|Y_{i,t}^1, Y_{i,t}^2, Y_{i,t}^3\right.\right)$ as follows

$$f\left(Q_{i,t}\left|Y_{i,t}^1, Y_{i,t}^2, Y_{i,t}^3\right.\right)=\frac{1}{Y_{i,t}^2}\left[1+Y_{i,t}^3\left(\frac{Q_{i,t}-Y_{i,t}^1}{Y_{i,t}^2}\right)\right]^{-\frac{1}{Y_{i,t}^3}-1}\exp\left\{-\left[1+Y_{i,t}^3\left(\frac{Q_{i,t}-Y_{i,t}^1}{Y_{i,t}^2}\right)\right]^{-\frac{1}{Y_{i,t}^3}}\right\}, \tag{1}$$
$$-\infty<Y_{i,t}^1<\infty, Y_{i,t}^2>0, -\infty<Y_{i,t}^3<\infty, 1+Y_{i,t}^3\left(\frac{Q_{i,t}-Y_{i,t}^1}{Y_{i,t}^2}\right)>0$$

where $Y_{i,t}^1$ is the first distribution parameter, i.e. the location parameter; $Y_{i,t}^2$ is the second distribution parameter, i.e. the scale parameter; $Y_{i,t}^3$ is the third distribution parameter, i.e. the shape parameter.

The generalized additive model for location, scale and shape (GAMLSS) (Dixit and JayakumarnAff, 2022; Rigby and Stasinopoulos, 2005; Stasinopoulos and Rigby, 2007) is introduced into the construction of the nonstationary RFFA model in this paper. For the GEV distribution, it is commonly assumed that estimators of the at-site flood distribution parameters $\tilde{Y}_{i,t}^1$ and/or $\tilde{Y}_{i,t}^2$ are dependent on nonstationary covariates whereas $\tilde{Y}_{i,t}^3$ is always constant, because $\tilde{Y}_{i,t}^3$ is quite sensitive and tough to be estimated (Du et al., 2015; Xiong et al., 2020). Therefore, the at-site flood distribution parameters can be expressed as follows

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

$$\tilde{\boldsymbol{\theta}}_{\mathrm{G}} = \left\{ \tilde{a}_{i,0}^{1}, \ldots, \tilde{a}_{i,H}^{1}, \tilde{a}_{i,0}^{2}, \ldots, \tilde{a}_{i,H}^{2}, \tilde{a}_{i,0}^{3} \right\} \begin{cases} g_1\left(\tilde{Y}_{i,t}^{1}\right) = \tilde{a}_{i,0}^{1} + \tilde{a}_{i,1}^{1} Z_{1,t}^{i} + \ldots + \tilde{a}_{i,H}^{1} Z_{H,t}^{i} \\ g_2\left(\tilde{Y}_{i,t}^{2}\right) = \tilde{a}_{i,0}^{2} + \tilde{a}_{i,1}^{2} Z_{1,t}^{i} + \ldots + \tilde{a}_{i,H}^{2} Z_{H,t}^{i} \\ \tilde{Y}_{i,t}^{3} = \tilde{a}_{i,0}^{3} \end{cases} \tag{2}$$

where $\tilde{\boldsymbol{\theta}}_{\mathrm{G}} = \left\{ \tilde{a}_{i,0}^{1}, \ldots, \tilde{a}_{i,H}^{1}, \tilde{a}_{i,0}^{2}, \ldots, \tilde{a}_{i,H}^{2}, \tilde{a}_{i,0}^{3} \right\}$ represents the at-site regression coefficients set of distribution parameters; $\left\{ \tilde{Y}_{i,1}^{1}, \ldots, \tilde{Y}_{i,T}^{1}, \tilde{Y}_{i,1}^{2}, \ldots, \tilde{Y}_{i,T}^{2}, \tilde{Y}_{i}^{3} \right\}$ represents the distribution parameters estimators based on the at-site flood samples only; $\mathbf{Z}_t^i = \left(Z_{1,t}^i, \ldots, Z_{H,t}^i \right)'$ is the vector composed of H nonstationary covariates of the i-th site distribution parameters, when H=0, nonstationary distribution model reduces to stationary distribution model; $g_1(\cdot)$ and $g_2(\cdot)$ represent the link function; $g_1(\cdot)$ is assumed to be the identity or the logarithmic function according to the existing studies (Read and Vogel, 2016; Sarhadi et al., 2016), while $g_2(\cdot)$ is assumed to be the logarithmic function to give the positive scale.

2.2 Catchment attributes selection for regional regression model

We select the catchment attributes that affect flood generating process based on the research of Stein et al. (2021). Three catchment attributes that have great influence on the corresponding dominant flood processes are selected for the following multiple regression model, due to the fact that Merz et al. (2000) concluded that the additional explained variance of regressions using more than three variables is small.

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

2.3 Stationary RFFA based on the Bayesian method

In stationary condition, the flood distribution parameters are constant, which means that H=0 in Eq. (2). The estimation of at-site regression coefficients set of distribution parameters is $\tilde{\boldsymbol{\theta}}_{\text{SG}} = \left\{\tilde{a}_{i,0}^{1}, \tilde{a}_{i,0}^{2}, \tilde{a}_{i,0}^{3}\right\}$ and the at-site flood distribution parameters are expressed as

$$\begin{cases} \tilde{Y}_i^1 = g_1^{-1}\left(\tilde{a}_{i,0}^1\right) \\ \tilde{Y}_i^2 = g_2^{-1}\left(\tilde{a}_{i,0}^1\right) \\ \tilde{Y}_i^3 = \tilde{a}_{i,0}^3 \end{cases} \tag{3}$$

where $g_1^{-1}(\cdot)$ and $g_2^{-1}(\cdot)$ represent the inverse functions of $g_1(\cdot)$ and $g_2(\cdot)$.

2.3.1 GLS model

The generalized least squares (GLS) model is used to establish the relationship between distribution parameters of multiple flood series and catchment attributes (Stedinger and Tasker, 1985, 1986a, 1986b), which assumes that the actual values of the distribution parameters of flood series can be described by a linear function of catchment attributes with additive errors

$$Y_i^k = \sum_{p=0}^{P} \beta_p^k X_{i,p} + \delta_i^k \tag{4}$$

$$\begin{aligned} & E(\delta_i^k) = 0 \\ & Cov(\delta_i^k, \delta_j^k) = \begin{cases} \sigma_{\delta^k}^2, i = j \\ 0, i \neq j \end{cases} \end{aligned} \tag{5}$$

where $X_{i,p}$ (i=1,…,N; p=1,…,P) represents the element of the matrix consisting of P

regional covariates for N sites; β_p^k represents the regional regression coefficient of the k-th distribution parameter; δ_i^k represents the normal distribution model error with the statistical properties as described in Eq. (5), where $\sigma_{\delta^k}^2$ represents the error variance of the GLS model.

Actually, the at-site actual value Y_i^k is generally unavailable, only the at-site estimate $\tilde{Y}_i^k$ of Y_i^k is available, thus requiring the sample error η_i^k to be introduced into Eq. (4) as

$$\begin{aligned}\tilde{Y}_i^k &= Y_i^k + \eta_i^k \\ &= \sum_{p=0}^{P} \beta_p^k X_{i,p} + \delta_i^k + \eta_i^k\end{aligned} \tag{6}$$

$$\begin{aligned}&E(\eta_i^k) = 0 \\ &Cov(\eta_i^k, \eta_j^k) = \begin{cases} \sigma_{\eta_i^k}^2, i = j \\ \sigma_{\eta_i^k}\sigma_{\eta_j^k}\rho_{ij}, i \neq j \end{cases}\end{aligned} \tag{7}$$

where $\sigma_{\eta_i^k}^2$ represents the sample error variance of $\tilde{Y}_i^k$ at site i; $\rho_{ij} = [\mathbf{R}^k]_{ij}$ represents the correlation coefficient between the sample error at site i and j.

The GLS model in matrix form can be expressed as

$$\begin{aligned}\tilde{\mathbf{Y}}^k &= \mathbf{X}\boldsymbol{\beta}^k + \boldsymbol{\delta}^k + \boldsymbol{\eta}^k \\ &= \mathbf{X}\boldsymbol{\beta}^k + \boldsymbol{\varepsilon}^k\end{aligned} \tag{8}$$

where $\mathbf{X}$ represents an $N \times (P+1)$ matrix composed of P regional covariates (i.e. catchment attributes) of N sites; $\boldsymbol{\beta}^k = \left(\beta_0^k, ..., \beta_P^k\right)'$ represents the regional regression coefficients set of the k-th distribution parameters; $\boldsymbol{\delta}^k = \left(\delta_1^k, ..., \delta_N^k\right)'$ represents the vector composed of the GLS model error of the k-th distribution parameters of N sites; $\boldsymbol{\eta}^k = \left(\eta_1^k, ..., \eta_N^k\right)'$ represents the vector composed of sample error of N sites; $\boldsymbol{\varepsilon}^k = \left(\varepsilon_1^k, ..., \varepsilon_N^k\right)'$ represents the total error of the combination of $\boldsymbol{\delta}^k$ and $\boldsymbol{\eta}^k$, where $\varepsilon_i^k = \delta_i^k + \eta_i^k$.

The total error $\boldsymbol{\varepsilon}^k$ has zero mean and covariance matrix $\mathbf{H}^k$. $\mathbf{H}^k$ can expressed as follows

$$\mathbf{H}^k = E\left[\boldsymbol{\varepsilon}^k \left(\boldsymbol{\varepsilon}^k\right)'\right] = \begin{bmatrix} \sigma_{\eta_i^k}^2 + \sigma_{\delta^k}^2 & \dots & \sigma_{\eta_1^k}\sigma_{\eta_N^k}\rho_{1N} \\ \vdots & \ddots & \vdots \\ \sigma_{\eta_N^k}\sigma_{\eta_1^k}\rho_{N1} & \cdots & \sigma_{\eta_N^k}^2 + \sigma_{\delta^k}^2 \end{bmatrix} \tag{9}$$

The parameter set of SG-GLS model is denoted by $\boldsymbol{\theta}_{\text{SG-GLS}} = \left\{\boldsymbol{\theta}_{\text{SG}}, \boldsymbol{\theta}_{\text{GLS}}\right\}$, where $\boldsymbol{\theta}_{\text{GLS}} = \left\{\boldsymbol{\beta}^k, \mathbf{R}^k, \sigma_{\eta_1^k}, ..., \sigma_{\eta_N^k}, \sigma_{\delta^k}\right\}$ represents the parameters set of the GLS model; $\boldsymbol{\theta}_{\text{SG}} = \left\{a_{i,0}^1, a_{i,0}^2, a_{i,0}^3\right\}$ represents the at-site regression coefficients set of the GEV distribution parameters in stationary condition.

2.3.2 GLS-derived prior information of at-site regression coefficients of flood distribution parameters

The estimation $\hat{\boldsymbol{\theta}}_{\text{GLS}} = \left\{\hat{\boldsymbol{\beta}}^k, \hat{\rho}_{ij}, \hat{\sigma}_{\eta_1^k}, ..., \hat{\sigma}_{\eta_N^k}, \hat{\sigma}_{\delta^k}\right\}$ of the GLS model is obtained by the method described in section 2.5.1. The at-site flood distribution parameters follow normal distribution with mean $\hat{\mu}_{\tilde{Y}_i^k}$ and variance $\hat{\sigma}_{\tilde{Y}_i^k}^2$

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

$$\begin{cases} \hat{\mu}_{\tilde{Y}_i^k} = [\mathbf{X}]_i \hat{\boldsymbol{\beta}}^k \\ \hat{\sigma}^2_{\tilde{Y}_i^k} = [\mathbf{X}]_i \Sigma\left(\hat{\boldsymbol{\beta}}^k\right)[\mathbf{X}]_i' + \hat{\sigma}^2_{\delta^k} \end{cases} \tag{10}$$

where $[\mathbf{X}]_i$ represents the element of the i-th row in $\mathbf{X}$; $\Sigma\left(\hat{\boldsymbol{\beta}}^k\right) = \left\{[\mathbf{X}]_i \hat{\mathbf{H}}^k [\mathbf{X}]_i'\right\}^{-1}$ represents the covariance matrix of the GLS model.

The prior probability density function $f\left(\tilde{Y}_i^k\right)$ of at-site flood distribution parameter $\tilde{Y}_i^k$ can be expressed as

$$f\left(\tilde{Y}_i^k\right) = \frac{1}{\sqrt{2\pi}\hat{\sigma}_{\tilde{Y}_i^k}} \exp\left\{-\frac{\left[g_k^{-1}\left(\tilde{a}_{i,0}^k\right) - \hat{\mu}_{\tilde{Y}_i^k}\right]^2}{2\hat{\sigma}^2_{\tilde{Y}_i^k}}\right\} \tag{11}$$

According to the functional relationship between distribution parameters and $\tilde{\boldsymbol{\theta}}_{\mathrm{SG}}$ described in Eq. (3), the prior probability density function $f\left(\tilde{\boldsymbol{\theta}}_{\mathrm{SG}}\right)$ of at-site regression coefficients $\tilde{\boldsymbol{\theta}}_{\mathrm{SG}}$ can be expressed as

$$\begin{aligned} f\left(\tilde{\boldsymbol{\theta}}_{\mathrm{SG}}\right) &= \prod_{k=1}^{K} \frac{\mathrm{d}f\left(\tilde{Y}_i^k\right)}{\mathrm{d}\tilde{a}_{i,0}^k} \\ &= \prod_{k=1}^{K} \frac{\mathrm{d}f\left(\tilde{Y}_i^k\right)}{\mathrm{d}g_k^{-1}\left(\tilde{a}_{i,0}^k\right)} \cdot \frac{g_k^{-1}\left(\tilde{a}_{i,0}^k\right)}{\mathrm{d}\tilde{a}_{i,0}^k} \end{aligned} \tag{12}$$

2.3.3 Posterior distribution of at-site regression coefficients derived from Bayesian theory

According to the Bayesian theory (Ouarda and El-Adlouni, 2011), the posterior probability density function $f\left(\boldsymbol{\theta}_{\mathrm{SG}} \middle| Q_{i,t}\right)$ of the at-site regression coefficients $\tilde{\boldsymbol{\theta}}_{\mathrm{SG}}$

can be expressed as

$$f\left(\boldsymbol{\theta}_{\mathrm{SG}}\left|Q_{i,t}\right.\right)=\frac{l\left(Q_{i,t}\left|\tilde{\boldsymbol{\theta}}_{\mathrm{SG}}\right.\right)f\left(\tilde{\boldsymbol{\theta}}_{\mathrm{SG}}\right)}{\int_{\boldsymbol{\Phi}_{\mathrm{SG}}}l\left(Q_{i,t}\left|\tilde{\boldsymbol{\theta}}_{\mathrm{SG}}\right.\right)f\left(\tilde{\boldsymbol{\theta}}_{\mathrm{SG}}\right)\mathrm{d}\tilde{\boldsymbol{\theta}}_{\mathrm{SG}}} \tag{13}$$

$$l\left(Q_{i,t}\left|\tilde{\boldsymbol{\theta}}_{\mathrm{SG}}\right.\right)=\prod_{t=1}^{T_i}f\left(Q_{i,t}\left|\tilde{\boldsymbol{\theta}}_{\mathrm{SG}}\right.\right) \tag{14}$$

where $l\left(Q_{i,t}\left|\tilde{\boldsymbol{\theta}}_{\mathrm{SG}}\right.\right)$ represents the likelihood function of at-site flood series in stationary condition; $f\left(\tilde{\boldsymbol{\theta}}_{\mathrm{SG}}\right)$ represents the prior probability density function.

2.4 Nonstationary RFFA based on the Bayesian method

In nonstationary condition, the distribution parameters of flood series vary with nonstationary covariates, which means that $H\neq0$ in Eq. (2) and the at-site regression coefficients set of distribution parameters is $\tilde{\boldsymbol{\theta}}_{\mathrm{NG}}=\left\{\tilde{a}_{i,0}^{1},...,\tilde{a}_{i,H}^{1},\tilde{a}_{i,0}^{2},...,\tilde{a}_{i,H}^{2},\tilde{a}_{i,0}^{3}\right\}$. The at-site flood distribution parameters are expressed as

$$\begin{cases}\tilde{Y}_{i,t}^{1}=g_1^{-1}\left(\tilde{a}_{i,0}^{1}+\tilde{a}_{i,1}^{1}Z_{1,t}^{i}+...+\tilde{a}_{i,H}^{1}Z_{H,t}^{i}\right)\\ \tilde{Y}_{i,t}^{2}=g_2^{-1}\left(\tilde{a}_{i,0}^{2}+\tilde{a}_{i,1}^{2}Z_{1,t}^{i}+...+\tilde{a}_{i,H}^{2}Z_{H,t}^{i}\right)\\ \tilde{Y}_{i}^{3}=\tilde{a}_{i,0}^{3}\end{cases} \tag{15}$$

2.4.1 LME model

The linear mixed effects (LME) model is used to establish the relationship between time-varying distribution parameters of multiple flood series and catchment attributes (Laird and Ware, 1982; Pinheiro and Bates, 2000). The actual value

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

$\mathbf{Y}_t^k = \left(Y_{1,t}^k, \ldots, Y_{N,t}^k\right)'$ of k-th (k=1,2,3) distribution parameter of N flood series at time t can be expressed as

$$Y_{i,t}^k = \sum_{p=0}^{P} \beta_p^k X_{i,p} + \sum_{p=0}^{P} b_{p,t}^k X_{i,p} + \delta_{i,t}^k \tag{16}$$

$$\begin{aligned} & E(b_{p,t}^k) = 0 \\ & Cov(b_{p,t}^k, b_{q,t}^k) = \begin{cases} \sigma_{b_p^k}^2, p = q \\ 0, p \neq q \end{cases} \end{aligned} \tag{17}$$

$$\begin{aligned} & E(\delta_{i,t}^k) = 0 \\ & Cov(\delta_{i,t}^k, \delta_{j,t}^k) = \begin{cases} \sigma_{\delta^k}^2, i = j \\ 0, i \neq j \end{cases} \end{aligned} \tag{18}$$

where $b_{p,t}^k$ represents the change of β_p^k at time t, also known as the random effect, which follows the normal distribution with the statistical properties as described in Eq. (17); $\delta_{i,t}^k$ represents the normal distribution model errors with the statistical properties as described in Eq. (18).

The at-site estimate $\tilde{Y}_{i,t}^k$ of $Y_{i,t}^k$ is available, and the sample error η_i^k is introduced into the Eq. (16) as

$$\begin{aligned} \tilde{Y}_{i,t}^k &= Y_{i,t}^k + \eta_{i,t}^k \\ &= \sum_{p=0}^{P} \beta_p^k X_{i,p} + \sum_{p=0}^{P} b_{p,t}^k X_{i,p} + \delta_{i,t}^k + \eta_{i,t}^k \end{aligned} \tag{19}$$

$$\begin{aligned} & E(\eta_{i,t}^k) = 0 \\ & Cov(\eta_{i,t}^k, \eta_{j,t}^k) = \begin{cases} \sigma_{\eta_i^k}^2, i = j \\ \sigma_{\eta_i^k} \sigma_{\eta_j^k} \rho_{ij}, i \neq j \end{cases} \end{aligned} \tag{20}$$

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

where $\sigma^2_{\eta^k_{i,t}}$ represents the sample error variance of $\tilde{Y}_i^k$ of site i at time t; $\rho_{ij} = [\mathbf{R}^k]_{ij}$ represents the correlation coefficient between the sample error at site i and j.

The LME model in matrix form can be expressed as

$$\begin{aligned}\tilde{\mathbf{Y}}_t^k &= \mathbf{X}\boldsymbol{\beta}^k + \mathbf{X}\mathbf{b}_t^k + \boldsymbol{\delta}_t^k + \boldsymbol{\eta}_t^k \\ &= \mathbf{X}\boldsymbol{\beta}^k + \mathbf{X}\mathbf{b}_t^k + \boldsymbol{\varepsilon}_t^k\end{aligned} \tag{21}$$

where $\mathbf{X}$ represents an $N\times(P+1)$ matrix composed of P regional covariates of N sites; $\boldsymbol{\beta}^k = \left(\beta_0^k, ..., \beta_P^k\right)'$ represents the regional regression coefficients set of the k-th distribution parameter; $\mathbf{b}_t^k = \left(b_{0,t}^k, ..., b_{P,t}^k\right)'$ represents the vector composed of random effects of P regional regression coefficients at time t, which has zero mean and covariance matrix $\mathbf{G}^k = diag(\sigma^2_{b_1^k}, ..., \sigma^2_{b_P^k})$; $\boldsymbol{\delta}_t^k = \left(\delta_{1,t}^k, ..., \delta_{N,t}^k\right)'$ represents the vector composed of LME model error of the k-th distribution parameter of N sites at time t; $\boldsymbol{\eta}_t^k = \left(\eta_{1,t}^k, ..., \eta_{N,t}^k\right)'$ represents the vector composed of sample error of N sites at time t; $\boldsymbol{\varepsilon}_t^k = \left(\varepsilon_{1,t}^k, ..., \varepsilon_{N,t}^k\right)'$ represents the total error of the combination of $\boldsymbol{\delta}_t^k$ and $\boldsymbol{\eta}_t^k$, where $\varepsilon_{i,t}^k = \delta_{i,t}^k + \eta_{i,t}^k$.

The total error $\boldsymbol{\varepsilon}_t^k$ at time t has zero mean and covariance matrix $\mathbf{H}_t^k$. $\mathbf{H}_t^k$ can expressed as follows

$$\mathbf{H}_t^k = E\left[\boldsymbol{\varepsilon}_t^k \left(\boldsymbol{\varepsilon}_t^k\right)'\right] = \begin{bmatrix} \sigma^2_{\eta^k_{1,t}} + \sigma^2_{\delta^k} & \dots & \sigma_{\eta^k_{1,t}}\sigma_{\eta^k_{N,t}}\rho_{1N} \\ \vdots & \ddots & \vdots \\ \sigma_{\eta^k_{N,t}}\sigma_{\eta^k_{1,t}}\rho_{N1} & \cdots & \sigma^2_{\eta^k_{N,t}} + \sigma^2_{\delta^k} \end{bmatrix} \tag{22}$$

The parameters set of NG-LME model is denoted by $\boldsymbol{\theta}_{\text{NG-LME}}=\left\{\boldsymbol{\theta}_{\text{NG}},\boldsymbol{\theta}_{\text{LME}}\right\}$, where $\boldsymbol{\theta}_{\text{LME}}=\left\{\boldsymbol{\beta}^k,\mathbf{R}^k,\sigma_{\eta_1^k},...,\sigma_{\eta_N^k},\sigma_{\delta^k}\right\}$ represents the parameters set of the LME model; $\boldsymbol{\theta}_{\text{NG}}=\left\{a_{i,0}^1,...,a_{i,H}^1,a_{i,0}^2,...,a_{i,H}^2,a_{i,0}^3\right\}$ is the at-site regression coefficients set of the GEV distribution parameters in nonstationary condition.

2.4.2 LME-derived prior information of at-site regression coefficients of flood distribution parameters

The estimation $\hat{\boldsymbol{\theta}}_{\text{LME}}=\left\{\hat{\boldsymbol{\beta}}^k,\hat{\rho}_{ij},\hat{\sigma}_{b_1^k},...,\hat{\sigma}_{b_P^k},\hat{\sigma}_{\eta_{1,t}^k},...,\hat{\sigma}_{\eta_{N,T}^k},\hat{\sigma}_{\delta^k}\right\}$ of the LME model is obtained by the method described in section 2.5.2. The at-site flood distribution parameters follow normal distribution with mean $\hat{\mu}_{\tilde{Y}_{i,t}^k}$ and variance $\hat{\sigma}^2_{\tilde{Y}_{i,t}^k}$

$$\begin{cases}\hat{\mu}_{\tilde{Y}_{i,t}^k}=\left[\mathbf{X}\right]_i\hat{\boldsymbol{\beta}}^k\\ \hat{\sigma}^2_{\tilde{Y}_{i,t}^k}=\left[\mathbf{X}\right]_i\Sigma\left(\hat{\boldsymbol{\beta}}^k\right)\left[\mathbf{X}\right]_i'+\left[\mathbf{X}\right]_i\mathbf{G}^k\left[\mathbf{X}\right]_i'+\hat{\sigma}^2_{\delta^k}\end{cases} \tag{23}$$

where $[\mathbf{X}]_i$ represents the element of the i-th row in $\mathbf{X}$; $\Sigma\left(\hat{\boldsymbol{\beta}}^k\right)=\left\{\left[\mathbf{X}\right]_i\hat{\mathbf{H}}_t^k\left[\mathbf{X}\right]_i'\right\}^{-1}$ represents the covariance matrix of the LME model.

The probability density function $f\left(\tilde{Y}_{i,t}^k\right)$ of at-site flood distribution parameter $\tilde{Y}_{i,t}^k$ can be expressed as

$$f\left(\tilde{Y}_{i,t}^k\right)=\frac{1}{\sqrt{2\pi}\hat{\sigma}_{\tilde{Y}_{i,t}^k}}\exp\left\{-\frac{\left[g_k^{-1}\left(\tilde{a}_{i,0}^k,...,\tilde{a}_{i,H}^k\right)-\hat{\mu}_{\tilde{Y}_{i,t}^k}\right]^2}{2\hat{\sigma}^2_{\tilde{Y}_{i,t}^k}}\right\} \tag{24}$$

According to the functional relationship between distribution parameters and

$\tilde{\boldsymbol{\theta}}_{\mathrm{NG}}$ described in Eq. (15), the prior probability density function $f\left(\tilde{\boldsymbol{\theta}}_{\mathrm{NG}}\right)$ of at-site regression coefficients $\tilde{\boldsymbol{\theta}}_{\mathrm{NG}}$ can be expressed as

$$\begin{aligned} f\left(\tilde{\boldsymbol{\theta}}_{\mathrm{NG}}\right) &= \prod_{k=1}^{K}\prod_{h=0}^{H}\frac{\mathrm{d}f\left(\tilde{Y}_{i,t}^{k}\right)}{\mathrm{d}\tilde{a}_{i,h}^{k}} \\ &= \prod_{k=1}^{K}\prod_{h=0}^{H}\frac{\mathrm{d}f\left(\tilde{Y}_{i,t}^{k}\right)}{\mathrm{d}g_{k}^{-1}\left(\tilde{a}_{i,0}^{k},\ldots,\tilde{a}_{i,H}^{k}\right)}\cdot\frac{\mathrm{d}g_{k}^{-1}\left(\tilde{a}_{i,0}^{k},\ldots,\tilde{a}_{i,H}^{k}\right)}{\mathrm{d}\tilde{a}_{i,h}^{k}} \end{aligned} \tag{25}$$

2.4.3 Posterior distribution of at-site regression coefficients derived from Bayesian theory

According to the Bayesian theory, the posterior probability density function $f\left(\boldsymbol{\theta}_{\mathrm{NG}}\left|Q_{i,t}\right.\right)$ of the at-site regression coefficients $\tilde{\boldsymbol{\theta}}_{\mathrm{NG}}$ can be expressed as

$$f\left(\boldsymbol{\theta}_{\mathrm{NG}}\left|Q_{i,t}\right.\right) = \frac{l\left(Q_{i,t}\left|\tilde{\boldsymbol{\theta}}_{\mathrm{NG}}\right.\right) f\left(\tilde{\boldsymbol{\theta}}_{\mathrm{NG}}\right)}{\int_{\boldsymbol{\Phi}_{\mathrm{NG}}} l\left(Q_{i,t}\left|\tilde{\boldsymbol{\theta}}_{\mathrm{NG}}\right.\right) f\left(\tilde{\boldsymbol{\theta}}_{\mathrm{NG}}\right)\mathrm{d}\tilde{\boldsymbol{\theta}}_{\mathrm{NG}}} \tag{26}$$

$$l\left(Q_{i,t}\left|\tilde{\boldsymbol{\theta}}_{\mathrm{NG}}\right.\right) = \prod_{t=1}^{T} f\left(Q_{i,t}\left|\tilde{\boldsymbol{\theta}}_{\mathrm{NG}}\right.\right) \tag{27}$$

where $l\left(Q_{i,t}\left|\tilde{\boldsymbol{\theta}}_{\mathrm{NG}}\right.\right)$ represents the likelihood function of at-site flood series in nonstationary condition; $f\left(\tilde{\boldsymbol{\theta}}_{\mathrm{NG}}\right)$ represents the prior probability density function.

2.5 Parameter Estimation

2.5.1 Estimation of the GLS model parameters

The parameters to be estimated in the GLS model described in Eq. (8) are

$\boldsymbol{\theta}_{\text{GLS}} = \left\{ \boldsymbol{\beta}^k, \mathbf{R}^k, \sigma_{\eta_1^k}, \ldots, \sigma_{\eta_N^k}, \sigma_{\delta^k} \right\}$.

The estimations $\tilde{Y}_i^k$ of at-site distribution parameters used in GLS model are obtained by the maximum likelihood method using only the at-site flood samples. The correlation coefficient $\rho_{ij} = [\mathbf{R}^k]_{ij}$ between the sample error at site i and site j can be calculated using the function of the distance between two sites (Reis et al., 2020).

The sample error variance $\sigma_{\eta_i^k}^2$ uses the estimate suggested by Gregersen et al. (2017). The maximum likelihood method is also used to estimate $\left\{ \boldsymbol{\beta}^k, \sigma_{\delta^k} \right\}$. We assume that the total error $\boldsymbol{\varepsilon}^k$ follows the normal distribution with mean zero and covariance matrix $\mathbf{H}^k$, the log-likelihood functions of $\boldsymbol{\varepsilon}^k$ can be expressed as

$$\begin{aligned} \ln f\left(\boldsymbol{\varepsilon}^k \middle| \hat{\mathbf{R}}^k, \boldsymbol{\beta}^k, \hat{\sigma}_{\eta_1^k}, \ldots, \hat{\sigma}_{\eta_N^k}, \sigma_{\delta^k} \right) &= -\frac{1}{2}\left[\ln\left|\mathbf{H}^k\right| + \left(\boldsymbol{\varepsilon}^k\right)' \left(\mathbf{H}^k\right)^{-1} \boldsymbol{\varepsilon}^k \right] \\ &= -\frac{1}{2}\left[\ln\left|\mathbf{H}^k\right| + \left(\tilde{\mathbf{Y}}^k - \mathbf{X}\boldsymbol{\beta}^k\right)' \left(\mathbf{H}^k\right)^{-1} \left(\tilde{\mathbf{Y}}^k - \mathbf{X}\boldsymbol{\beta}^k\right) \right] \end{aligned} \tag{28}$$

The estimation $\left\{ \hat{\boldsymbol{\beta}}^k, \hat{\sigma}_{\delta^k} \right\}$ is obtained by maximizing the log-likelihood function $\ln f\left(\boldsymbol{\varepsilon}^k \middle| \hat{\mathbf{R}}^k, \boldsymbol{\beta}^k, \hat{\sigma}_{\eta_1^k}, \ldots, \hat{\sigma}_{\eta_N^k}, \sigma_{\delta^k} \right)$.

2.5.2 Estimation of the LME model parameters

The parameters need to be estimated in the LME model described in Eq. (21) are $\boldsymbol{\theta}_{\text{LME}} = \left\{ \boldsymbol{\beta}^k, \mathbf{R}^k, \mathbf{G}^k, \sigma_{\eta_{1,t}^k}, \ldots, \sigma_{\eta_{N,T}^k}, \sigma_{\delta^k} \right\}$.

The at-site time-varying distribution parameter $\tilde{Y}_{i,t}^k$ used in the LME model is

Accepted version

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

estimated by the maximum likelihood method using only the at-site flood samples. The correlation coefficient matrix $\mathbf{R}^k$ is estimated in the same way as in the GLS model. The sample error variance $\sigma^2_{\eta^k_{i,t}}$ is estimated as

$$\hat{\sigma}^2_{\eta^k_{i,t}} = \frac{\sum_{i=1}^{N}\left(\tilde{Y}^k_{i,t} - \bar{Y}^k_{i,t}\right)^2}{N-1} \tag{29}$$

$$\bar{Y}^k_{i,t} = \frac{1}{N}\sum_{i=1}^{N}\tilde{Y}^k_{i,t} \tag{30}$$

where $\bar{Y}^k_{i,t}$ represents the average of $\tilde{Y}^k_{i,t}$.

The maximum likelihood estimation method has also been used to estimate $\left\{\boldsymbol{\beta}^k, \mathbf{G}^k, \sigma_{\delta^k}\right\}$. The total error $\boldsymbol{\varepsilon}^k_t$ follows the normal distribution with mean zero and covariance matrix $\mathbf{H}^k_t$, the log-likelihood functions of $\boldsymbol{\varepsilon}^k_t$ is

$$\begin{aligned} &\ln f\left(\boldsymbol{\varepsilon}^k_1,...,\boldsymbol{\varepsilon}^k_T \left| \hat{\mathbf{R}}^k, \hat{\sigma}_{\eta^k_{1,t}},...,\hat{\sigma}_{\eta^k_{N,T}}, \mathbf{G}^k, \boldsymbol{\beta}^k, \sigma_{\delta^k}\right.\right) \\ &= -\frac{1}{2}\sum_{t=1}^{T}\left[\ln\left|\mathbf{H}^k_t\right| + \left(\boldsymbol{\varepsilon}^k_t\right)'\left(\mathbf{H}^k_t\right)^{-1}\boldsymbol{\varepsilon}^k_t\right] \\ &= -\frac{1}{2}\sum_{t=1}^{T}\left[\ln\left|\mathbf{H}^k_t\right| + \left(\tilde{\mathbf{Y}}^k_t - \mathbf{X}\boldsymbol{\beta}^k - \mathbf{X}\mathbf{b}^k_t\right)'\left(\mathbf{H}^k_t\right)^{-1}\left(\tilde{\mathbf{Y}}^k_t - \mathbf{X}\boldsymbol{\beta}^k - \mathbf{X}\mathbf{b}^k_t\right)\right] \end{aligned} \tag{31}$$

The estimation $\left\{\hat{\boldsymbol{\beta}}^k, \hat{\mathbf{G}}^k, \hat{\sigma}_{\delta^k}\right\}$ is obtained by maximizing the log-likelihood function $\ln f\left(\boldsymbol{\varepsilon}^k_1,...,\boldsymbol{\varepsilon}^k_T \left| \hat{\mathbf{R}}^k, \hat{\sigma}_{\eta^k_{1,t}},...,\hat{\sigma}_{\eta^k_{N,T}}, \mathbf{G}^k, \boldsymbol{\beta}^k, \sigma_{\delta^k}\right.\right)$.

2.5.3 Bayesian estimation of at-site regression coefficients of flood distribution parameters

The Bayesian estimation of the at-site regression coefficients of flood distribution parameters are the expectation of its posterior probability density distribution, which can be expressed as

$$\hat{\boldsymbol{\theta}}_{\mathrm{SG}} = \int_{\boldsymbol{\Phi}_{\mathrm{SG}}} \boldsymbol{\theta}_{\mathrm{SG}} \cdot f\left(\boldsymbol{\theta}_{\mathrm{SG}} \left| Q_{i,t}\right.\right) \mathrm{d}\boldsymbol{\theta}_{\mathrm{SG}} \tag{32}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{NG}} = \int_{\boldsymbol{\Phi}_{\mathrm{NG}}} \boldsymbol{\theta}_{\mathrm{NG}} \cdot f\left(\boldsymbol{\theta}_{\mathrm{NG}} \left| Q_{i,t}\right.\right) \mathrm{d}\boldsymbol{\theta}_{\mathrm{NG}} \tag{33}$$

where $\hat{\boldsymbol{\theta}}_{\mathrm{SG}}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{NG}}$ represent the Bayesian estimators set of at-site regression coefficients of flood distribution parameters in stationary and nonstationary conditions, respectively. The posterior distribution of at-site regression coefficients of flood distribution parameters can be calculated by the Markov chain Monte Carlo (MCMC) algorithm (El Adlouni et al., 2007; Laloy and Vrugt, 2012; Martins and Stedinger, 2000; Vrugt et al., 2009).

2.6 Model selection and diagnosis

In the Bayesian method, the selection of nonstationary covariate $Z_{h,t}^{i}$ for *i* site is based on the Deviance Information Criterion (DIC; Spiegelhalter et al., 2002, 2014)

In testing the goodness-of-fit of the model, the quantile-quantile plot based on the diagnosis method is used (Coles, 2001), which assumes that a good model should have plotted points close to the 1:1 line. The fitted models are further evaluated by testing the goodness of fit and the uncertainty in quantile estimation.

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

3 Study area and data

3.1 Study area

Dongting Lake, the second-largest freshwater lake in China, is located in the northeastern of Hunan Province, on the southern bank of the Yangtze River mainstream. The Yangtze River discharges water and sediment to the East Dongting Lake and West Dongting Lake through the Songzi River, the Hudu River and the Dahei River, i.e. the Three Inlets. The Southern Dongting Lake and Western Dongting Lake are fed by four main tributaries: the Xiangjiang River, the Zishui River, the Yuanjiang River, and the Lishui River, which is referred to as the Four Waters. Dongting Lake discharges water and sediment into the Yangtze River through the Chenglingji station, which is the only outlet of the Dongting Lake to the Yangtze river.

3.2 Data

This study covers four hydrological gauges in the Dongting Lake basin (i.e. Shimen, Taoyuan, Taojiang and Xiangtan). The observed daily streamflow records for the four gauges were provided by the Hydrology Bureau of the Changjiang Water Resources Commission, China (http://www.cjh.com.cn/en/index. html). There are a total of 48 meteorological stations located in and around drainage areas of the four hydrological gauges, with meteorological data obtained from the National Climate Center of the China Meteorological Administration (http://www.cma.gov.cn/).

The spatial distribution of the hydrological and meteorological gauges is presented in Fig. 2, and information on streamflow series is presented in Table 1.

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

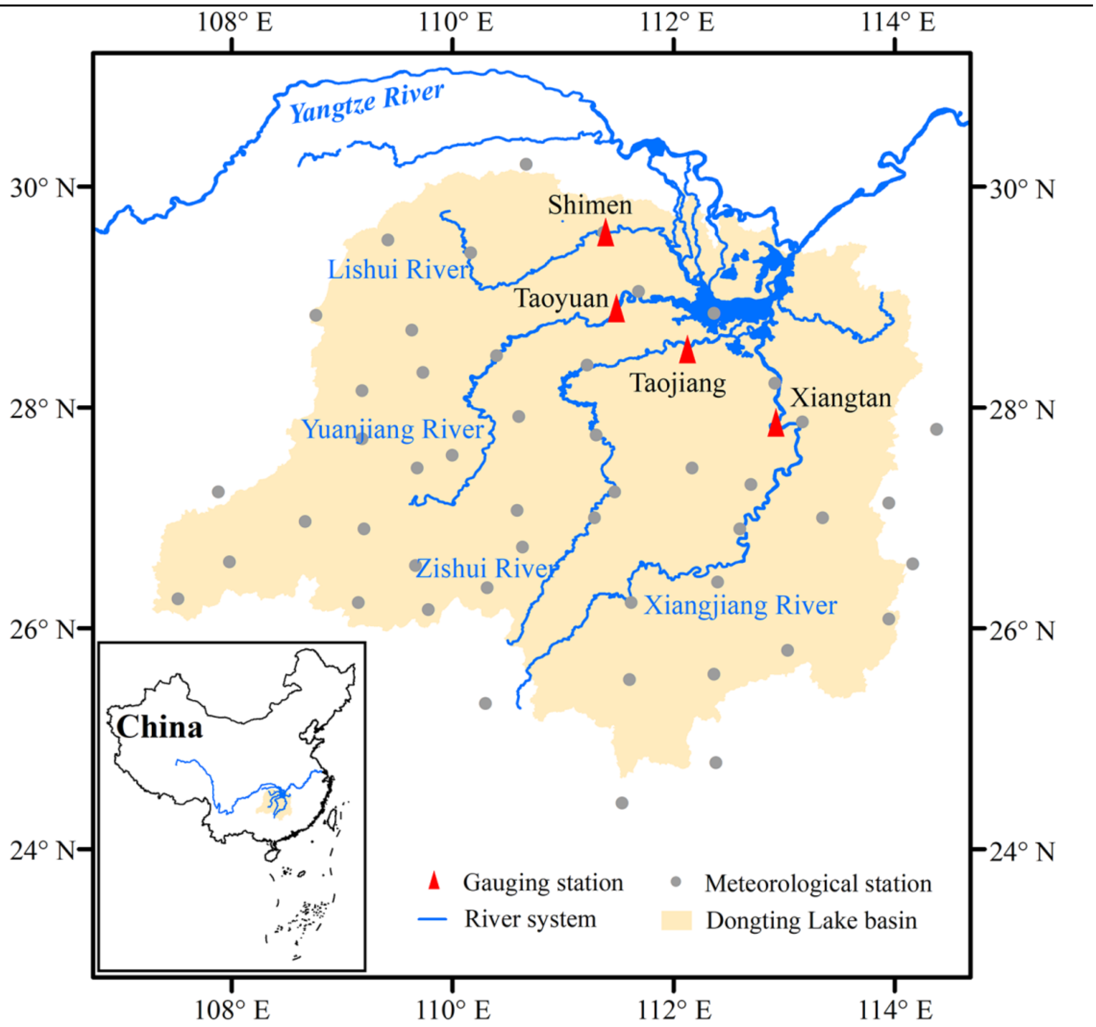

Fig. 2 Map of the Dongting Lake basin

Table 1. List of streamflow gauging stations in the Dongting Lake basin used in this study.

Station	Location	Station code	Station number	Catchment area (km^2)
Shimen	Lishui River	SM	1	15139
Taoyuan	Yuanjiang River	TY	2	87571
Taojiang	Zishui River	TJ	3	27033

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

Xiangtan	Xiangjiang River	XT	4	81638

4 Results and discussion

4.1 Preliminary analysis

We extracted the annual maximum daily streamflow and the corresponding date from the 52-year daily streamflow series as the flood information for this study. The Mann-Kendall test (Mann, 1945; Kendall, 1975) is used to test the nonstationary of flood series SM, TY, TJ and XT, and the results show a significant downward trend in SM. We take the annual precipitation P and potential evapotranspiration Ep as candidate covariates of the NG model. The nonstationary models are optimal for SM, TY and XT, while the stationary model is optimal for TJ as shown in Table 6. The DIC values of nonstationary model are lower than those of stationary model, so it is concluded that the performance of nonstationary model based on covariates is better than that of stationary model.

4.2 Selection of catchment attributes

Table 2 shows the classification results of the annual flood generating process for SM, TY, TJ and XT. According to the research of Stein et al. (2021), we select the three indicators of mean slope (*MS*), mean precipitation (*MP*) and forest fraction (*FF*) (Table 2) as catchment attributes for the following multiple regression.

Table 2. Catchment attributes selection list for the regional regression model.

Station	Climate type (E_P/P)	Dominant flood type	Catchment attributes		
			MS (°)	*MP* (mm)	*FF* (km^2)

SM	0.642	LR	19.65	1384	10151
TY	0.676	LR	17.56	1328	59533
TJ	0.658	LR	14.91	1430	17492
XT	0.663	LR	12.38	1506	51902

4.3 Simulation results

4.3.1 Regional regression results

The GLS model is used for the regional regression of flood distribution parameters of SM, TY, TJ and XT in stationary condition. Table 3 shows that the R^2 is above 0.85 for all four sites, which indicates that the GLS model has good fitting performance and can be used for subsequent analysis. The variances $\sigma^2_{\eta_i^k}$ of sample errors of the three distribution parameters are 5124.41, 2487.59 and 1.24, respectively. The variances $\sigma^2_{\delta^k}$ of regression errors of the three distribution parameters are 4013.42, 1134.01 and 2.79, respectively. It can be seen that the variance of the sample residuals and the variance of the regression residuals are reduced for the location, scale and shape parameters.

Accepted version

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

Table 3. The regional regression results of flood distribution parameters by using the GLS model and the LME model for SM, TY, TJ and XT in stationary condition.

Distribution parameter	GLS model ($\tilde{Y}_i^k = \beta_0^k + \beta_1^k \cdot MS + \beta_2^k \cdot MP + \beta_3^k \cdot FF + \varepsilon_i^k$) parameter estimation							R^2
	β_0^k	β_1^k	β_2^k	β_3^k	$\mathbf{R}^k$	$\sigma_{\eta_i^k}^2$	$\sigma_{\delta^k}^2$	
$\tilde{Y}_i^1$	-25147.47	534.78	12.46	0.16	$\begin{bmatrix} 1 & 0.42 & 0.32 & 0.25 \\ 0.42 & 1 & 0.43 & 0.28 \\ 0.32 & 0.43 & 1 & 0.36 \\ 0.25 & 0.28 & 0.36 & 1 \end{bmatrix}$	5124.41	4013.42	0.85
$\tilde{Y}_i^2$	-1078.18	117.46	-0.15	0.47		2487.59	1134.01	0.94
$\tilde{Y}_i^3$	0.04	0.03	-0.02	0.02		1.24	2.79	0.97

Distribution parameter	LME model ($\tilde{Y}_{i,t}^k = (\beta_0^k + b_{0,t}^k) + (\beta_1^k + b_{1,t}^k) \cdot MS + (\beta_2^k + b_{2,t}^k) \cdot MP + (\beta_3^k + b_{3,t}^k) \cdot FF + \varepsilon_{i,t}^k$) parameter estimation							R^2
	β_0^k	β_1^k	β_2^k	β_3^k	$\mathbf{R}^k$	$\mathbf{G}^k$	$\sigma_{\delta^k}^2$	

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

$\tilde{Y}_{i,t}^{1}$	-24910.05	623.57	11.29	0.22	$\begin{bmatrix} 1 & 0.42 & 0.32 & 0.25 \\ 0.42 & 1 & 0.43 & 0.28 \\ 0.32 & 0.43 & 1 & 0.36 \\ 0.25 & 0.28 & 0.36 & 1 \end{bmatrix}$	$\begin{bmatrix} 2.82^2 & & & \\ & 0.01^2 & & \\ & & 0.02^2 & \\ & & & 0.02^2 \end{bmatrix}$	2471.43	0.87
$\tilde{Y}_{i,t}^{2}$	-1013.85	129.01	-0.03	0.06		$\begin{bmatrix} 0.03^2 & & & \\ & 0.01^2 & & \\ & & 0.01^2 & \\ & & & 0.01^2 \end{bmatrix}$	814.04	0.93
$\tilde{Y}_{i}^{3}$	0.06	0.02	0.01	0.01		-	1.27	0.91

The LME model is used to regional regression the time-varying flood distribution parameters of SM, TY, TJ and XT in nonstationary condition. Table 4 shows that the R^2 is above 0.87 for all four sites, indicating that the LME model has good fitting performance and can be used for the subsequent nonstationary conditional analysis. Consistent with the findings in stationary condition, The variance of the sample residuals and the variance of the regression residuals are reduced for the location, scale and shape parameters.

The performance of the regional regression model in the nonstationary condition is superior compared to the stationary condition. Tables 3 shows that the variances of regression errors in the nonstationary condition are much smaller than that in the stationary condition. This is due to the use of random effect term in the nonstationary condition. Compared with the stationary condition, the random effect term is used to consider the errors of regression coefficients in the nonstationary condition, thus significantly reducing the variance of regression errors.

4.3.2 Inference results of prior probability distribution function of at-site regression coefficients

The multivariate normal distributions of distribution parameters are obtained from the regional regression of three distribution parameters of SM, TY, TJ and XT, and the prior probability distribution functions of parameters of flood distribution parameters are then further derived.

The formulas of at-site flood distribution parameters in stationary condition are

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

shown in Table 4. The prior probability distribution functions of at-site regression coefficients inferred from the probability distribution of distribution parameters at the four stations in stationary condition are shown below

SM:

$$
\begin{aligned}
f\left(a_{1,0}^{1}\right) &= \frac{1}{\sqrt{2\pi}\cdot 26.01}\exp\left[-\frac{\left(a_{1,0}^{1}-4229.75\right)^{2}}{2\cdot 676.43}\right] \\
f\left(a_{1,0}^{2}\right) &= -\frac{\left[\exp\left(a_{1,0}^{2}\right)-5793.27\right]\cdot\exp\left(a_{1,0}^{2}\right)}{2\pi\cdot 16.63^{3}}\exp\left\{-\frac{\left[\exp\left(a_{1,0}^{2}\right)-5793.27\right]^{2}}{276.48}\right\} \\
f\left(a_{1,0}^{3}\right) &= \frac{1}{\sqrt{2\pi}\cdot 2.39}\exp\left[-\frac{\left(a_{1,0}^{3}-1.30\right)^{2}}{2\cdot 5.73}\right]
\end{aligned}
\tag{34}
$$

TY:

$$
\begin{aligned}
f\left(a_{2,0}^{1}\right) &= -\frac{\left[\exp\left(a_{2,0}^{1}\right)-10315.42\right]\cdot\exp\left(a_{2,0}^{1}\right)}{2\pi\cdot 57.84^{3}}\exp\left\{-\frac{\left[\exp\left(a_{2,0}^{1}\right)-10315.42\right]^{2}}{3346.15}\right\} \\
f\left(a_{2,0}^{2}\right) &= -\frac{\left[\exp\left(a_{2,0}^{2}\right)-28765.72\right]\cdot\exp\left(a_{2,0}^{2}\right)}{2\pi\cdot 61.43^{3}}\exp\left\{-\frac{\left[\exp\left(a_{2,0}^{2}\right)-28765.72\right]^{2}}{3773.14}\right\} \\
f\left(a_{2,0}^{3}\right) &= \frac{1}{\sqrt{2\pi}\cdot 1.78}\exp\left[-\frac{\left(a_{2,0}^{3}-0.45\right)^{2}}{2\cdot 3.18}\right]
\end{aligned}
\tag{35}
$$

TJ:

$$f\left(a_{3,0}^{1}\right)=\frac{1}{\sqrt{2\pi}\cdot 23.41}\exp\left[-\frac{\left(a_{3,0}^{1}-3442.62\right)^{2}}{2\cdot 548.24}\right]$$

$$f\left(a_{3,0}^{2}\right)=-\frac{\left[\exp\left(a_{3,0}^{2}\right)-8679.88\right]\cdot\exp\left(a_{3,0}^{2}\right)}{2\pi\cdot 31.70^{3}}\exp\left\{-\frac{\left[\exp\left(a_{3,0}^{2}\right)-8679.88\right]^{2}}{1004.76}\right\}$$

$$f\left(a_{3,0}^{3}\right)=\frac{1}{\sqrt{2\pi}\cdot 1.26}\exp\left[-\frac{\left(a_{3,0}^{3}-1.76\right)^{2}}{2\cdot 1.59}\right]$$

(36)

XT:

$$f\left(a_{4,0}^{1}\right)=-\frac{\left[\exp\left(a_{4,0}^{1}\right)-8542.18\right]\cdot\exp\left(a_{4,0}^{1}\right)}{2\pi\cdot 14.80^{3}}\exp\left\{-\frac{\left[\exp\left(a_{4,0}^{1}\right)-8542.18\right]^{2}}{218.43}\right\}$$

$$f\left(a_{4,0}^{2}\right)=-\frac{\left[\exp\left(a_{4,0}^{2}\right)-24544.01\right]\cdot\exp\left(a_{4,0}^{2}\right)}{2\pi\cdot 49.81^{3}}\exp\left\{-\frac{\left[\exp\left(a_{4,0}^{2}\right)-24544.01\right]^{2}}{2481.16}\right\} \quad (37)$$

$$f\left(a_{4,0}^{3}\right)=\frac{1}{\sqrt{2\pi}\cdot 1.65}\exp\left[-\frac{\left(a_{4,0}^{3}-0.76\right)^{2}}{2\cdot 2.73}\right]$$

Eqs. (41) to (44) are the prior probability distribution functions for at-site regression coefficients of SM, TY, TJ and XT in stationary condition.

The formulas of at-site flood distribution parameters in nonstationary condition are shown in Table 4. Compared with the stationary condition, the mean and variance of distribution parameters are slightly reduced in the nonstationary condition. The probability distributions of at-site regression coefficients for the four sites in nonstationary condition are shown below

SM:

$$
\begin{aligned}
f\left(a_{1,0}^{1}\right) &= \frac{\left(a_{1,0}^{1}+a_{1,1}^{1}P\right)-4473.18}{\sqrt{2\pi}\cdot 23.31}\exp\left\{-\frac{\left[\left(a_{1,0}^{1}+a_{1,1}^{1}P\right)-4473.18\right]^{2}}{2\cdot 543.15}\right\} \\
f\left(a_{1,1}^{1}\right) &= \frac{P\cdot\left[\left(a_{1,0}^{1}+a_{1,1}^{1}P\right)-4473.18\right]}{\sqrt{2\pi}\cdot 23.31}\exp\left\{-\frac{\left[\left(a_{1,0}^{1}+a_{1,1}^{1}P\right)-4473.18\right]^{2}}{2\cdot 543.15}\right\} \\
f\left(a_{1,0}^{2}\right) &= -\frac{\left[\exp\left(a_{1,0}^{2}\right)-3483.43\right]\cdot\exp\left(a_{1,0}^{2}\right)}{2\pi\cdot 14.04^{3}}\exp\left\{-\frac{\left[\exp\left(a_{1,0}^{2}\right)-3483.43\right]^{2}}{197.25}\right\} \\
f\left(a_{1,0}^{3}\right) &= \frac{1}{\sqrt{2\pi}\cdot 1.87}\exp\left[-\frac{\left(a_{1,0}^{3}-1.46\right)^{2}}{2\cdot 3.48}\right]
\end{aligned}
\tag{38}
$$

TY:

$$
\begin{aligned}
f\left(a_{2,0}^{1}\right) &= \frac{\left(a_{2,0}^{1}+a_{2,1}^{1}\cdot P+a_{2,2}^{1}Ep\right)-9316.46}{\sqrt{2\pi}\cdot 54.93}\exp\left\{-\frac{\left[\left(a_{2,0}^{1}+a_{2,1}^{1}\cdot P+a_{2,2}^{1}Ep\right)-9316.46\right]^{2}}{2\cdot 3017.43}\right\} \\
f\left(a_{2,1}^{1}\right) &= \frac{P\cdot\left[\left(a_{2,0}^{1}+a_{2,1}^{1}P+a_{2,2}^{1}Ep\right)-9316.46\right]}{\sqrt{2\pi}\cdot 54.93}\exp\left\{-\frac{\left[\left(a_{2,0}^{1}+a_{2,1}^{1}P+a_{2,2}^{1}Ep\right)-9316.46\right]^{2}}{2\cdot 3017.43}\right\} \\
f\left(a_{2,2}^{1}\right) &= \frac{Ep\cdot\left[\left(a_{2,0}^{1}+a_{2,1}^{1}P+a_{2,2}^{1}Ep\right)-9316.46\right]}{\sqrt{2\pi}\cdot 54.93}\exp\left\{-\frac{\left[\left(a_{2,0}^{1}+a_{2,1}^{1}P+a_{2,2}^{1}Ep\right)-9316.46\right]^{2}}{2\cdot 3017.43}\right\} \\
f\left(a_{2,0}^{2}\right) &= -\frac{\left[\exp\left(a_{2,0}^{2}\right)-27428.43\right]\cdot\exp\left(a_{2,0}^{2}\right)}{2\pi\cdot 64.14^{3}}\exp\left\{-\frac{\left[\exp\left(a_{2,0}^{2}\right)-27428.43\right]^{2}}{4113.42}\right\} \\
f\left(a_{2,0}^{3}\right) &= \frac{1}{\sqrt{2\pi}\cdot 1.87}\exp\left[-\frac{\left(a_{2,0}^{3}-1.46\right)^{2}}{2\cdot 3.48}\right]
\end{aligned}
$$

(39)

TJ:

$$
\begin{aligned}
&f\left(a_{3,0}^{1}\right)=-\frac{\left[\exp\left(a_{3,0}^{1}+a_{3,1}^{1}Ep\right)-3648.73\right]\cdot\exp\left(a_{3,0}^{1}+a_{3,1}^{1}Ep\right)}{2\pi\cdot 25.93^{3}}\exp\left\{-\frac{\left[\exp\left(a_{3,0}^{1}+a_{3,1}^{1}Ep\right)-3648.73\right]^{2}}{672.15}\right\}\\
&f\left(a_{3,1}^{1}\right)=-\frac{Ep\cdot\left[\exp\left(a_{3,0}^{1}+a_{3,1}^{1}Ep\right)-3648.73\right]\cdot\exp\left(a_{3,0}^{1}+a_{3,1}^{1}Ep\right)}{2\pi\cdot 25.93^{3}}\exp\left\{-\frac{\left[\exp\left(a_{3,0}^{1}+a_{3,1}^{1}Ep\right)-3648.73\right]^{2}}{672.15}\right\}\\
&f\left(a_{3,0}^{2}\right)=-\frac{\left[\exp\left(a_{3,0}^{2}\right)-8472.43\right]\cdot\exp\left(a_{3,0}^{2}\right)}{2\pi\cdot 28.90^{3}}\exp\left\{-\frac{\left[\exp\left(a_{3,0}^{2}\right)-8472.43\right]^{2}}{835.47}\right\}\\
&f\left(a_{3,0}^{3}\right)=\frac{1}{\sqrt{2\pi}\cdot 1.65}\exp\left[-\frac{\left(a_{3,0}^{3}-2.01\right)^{2}}{2\cdot 2.73}\right]
\end{aligned}
\tag{40}
$$

XT:

$$
\begin{aligned}
&f\left(a_{4,0}^{1}\right)=\frac{1}{\sqrt{2\pi}\cdot 23.32}\exp\left[-\frac{\left(a_{4,0}^{1}-8673.46\right)^{2}}{2\cdot 543.76}\right]\\
&f\left(a_{4,0}^{2}\right)=-\frac{\left[\exp\left(a_{4,0}^{2}+a_{4,1}^{2}Ep\right)-3483.43\right]\cdot\exp\left(a_{4,0}^{2}+a_{4,1}^{2}Ep\right)}{2\pi\cdot 51.71^{3}}\exp\left\{-\frac{\left[\exp\left(a_{4,0}^{2}+a_{4,1}^{2}Ep\right)-23446.23\right]^{2}}{2673.53}\right\}\\
&f\left(a_{4,0}^{2}\right)=-\frac{Ep\cdot\left[\exp\left(a_{4,0}^{2}+a_{4,1}^{2}Ep\right)-3483.43\right]\cdot\exp\left(a_{4,0}^{2}+a_{4,1}^{2}Ep\right)}{2\pi\cdot 51.71^{3}}\exp\left\{-\frac{\left[\exp\left(a_{4,0}^{2}+a_{4,1}^{2}Ep\right)-23446.23\right]^{2}}{2673.53}\right\}\\
&f\left(a_{4,0}^{3}\right)=\frac{1}{\sqrt{2\pi}\cdot 1.65}\exp\left[-\frac{\left(a_{4,0}^{3}-1.43\right)^{2}}{2\cdot 2.73}\right]
\end{aligned}
\tag{41}
$$

4.3.3 Calculation results of at-site regression coefficients

Table 4 shows the results of the at-site regression coefficients estimation for SM, TY, TJ, and XT when using the SG-GLS and NG-LME models. Fig. 3 shows the posterior probability distributions of at-site regression coefficients when using SG-GLS and NG-LME models.The results show that the uncertainty in the at-site regression

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

coefficients estimation is reduced when using the NG-LME model compared to the SG-GLS model, such as for SM, the uncertainty of at-site regression coefficients $a^1_{1,0}$ of location parameter by using the SG-GLS model is 157^2, which is divided into the uncertainty of $a^1_{1,0}$ and $a^1_{1,1}$ are 72^2 and 0.16^2 by using the NG-LME model. The uncertainty of at-site regression coefficients $a^2_{1,0}$ and $a^3_{1,0}$ when using the SG-GLS model are 0.15^2 and 0.09^2, which is reduced to 0.01^2 and 0.04^2 when using the NG-LME model.

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

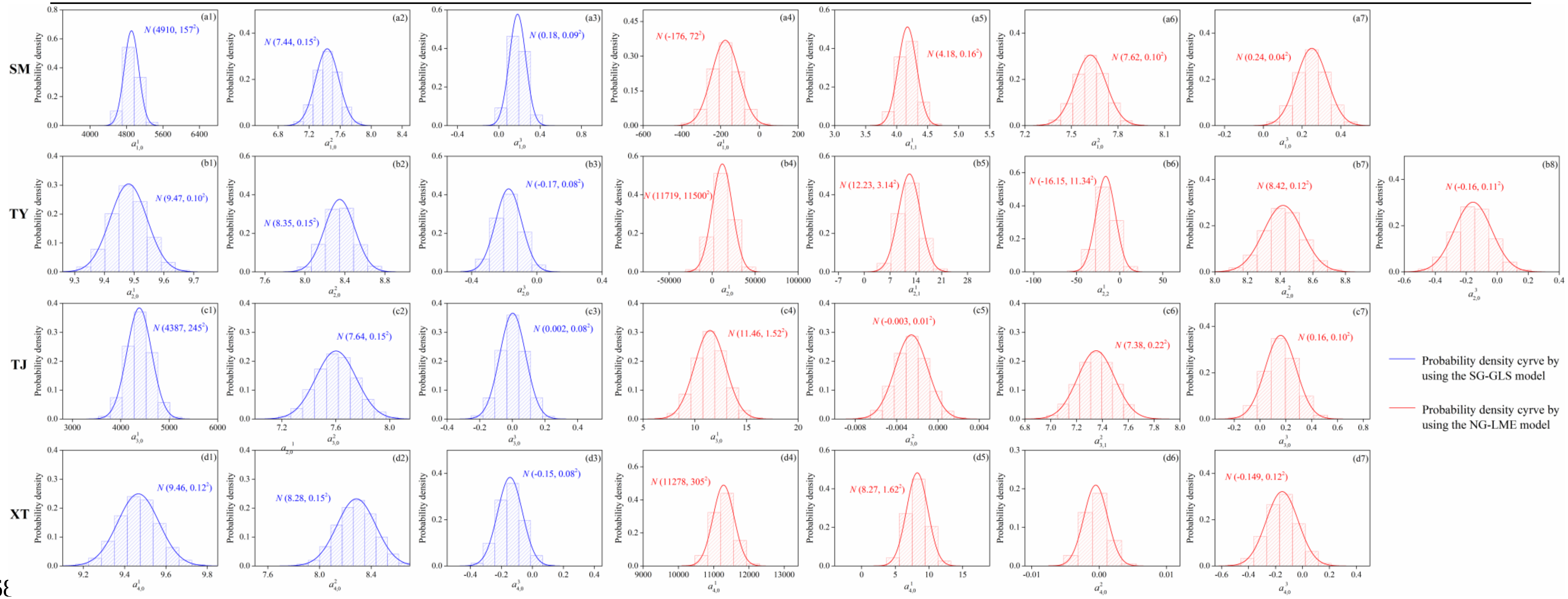

Fig. 3 Posterior probability density curve of at-site regression coefficients of SM, TY, TJ and XT by using the SG-GLS and NG-LME models

Table 4. The Bayesian estimation results of SG-GLS and NG-LME models of SM, TY, TJ, and XT.

Station	Model	Formulas for calculating distribution parameters	Estimation of at-site regression coefficients						DIC
			$a_{i,0}^{1}$	$a_{i,1}^{1}$	$a_{i,2}^{1}$	$a_{i,0}^{2}$	$a_{i,1}^{2}$	$a_{i,0}^{3}$	
SM	SG-GLS	$Y_1^1 = a_{1,0}^1$ $Y_1^2 = \exp(a_{1,0}^2)$ $Y_1^3 = a_{1,0}^3$	4910.274	-	-	7.435	-	0.181	1011.829
	NG-LME	$Y_{1,t}^1 = a_{1,0}^1 + a_{1,1}^1 P$ $Y_{1,t}^2 = \exp(a_{1,0}^2)$ $Y_1^3 = a_{1,0}^3$	-176.345	4.175	-	7.622	-	0.250	1010.758
TY	SG-GLS	$Y_2^1 = \exp(a_{2,0}^1)$ $Y_2^2 = \exp(a_{2,0}^2)$ $Y_2^3 = a_{2,0}^3$	9.617	-	-	8.347	-	-0.174	1078.189
	NG-LME	$Y_{2,t}^1 = a_{2,0}^1 + a_{2,1}^1 P + a_{2,2}^1 Ep$ $Y_{2,t}^2 = \exp(a_{2,0}^2)$ $Y_2^3 = a_{2,0}^3$	11719.045	12.226	-16.154	8.417	-	-0.156	1035.728

TJ	SG-GLS	$Y_3^1 = a_{3,0}^1$ $Y_3^2 = \exp(a_{3,0}^2)$ $Y_3^3 = a_{3,0}^3$	4387.487	-	-	7.671	-	0.002	1047.746
	NG-LME	$Y_{3,t}^1 = \exp(a_{3,0}^1 + a_{3,1}^1 Ep)$ $Y_{3,t}^2 = \exp(a_{3,0}^2)$ $Y_3^3 = a_{3,0}^3$	11.575	-0.003		7.586		0.157	987.586
XT	SG-GLS	$Y_4^1 = \exp(a_{4,0}^1)$ $Y_4^2 = \exp(a_{4,0}^2)$ $Y_4^3 = a_{4,0}^3$	9.476	-	-	8.284	-	-0.145	1074.864
	NG-LME	$Y_{4,t}^1 = a_{4,0}^1$ $Y_{4,t}^2 = \exp(a_{4,0}^2 + a_{4,1}^2 Ep)$ $Y_4^3 = a_{4,0}^3$	11278.578	-	-	8.276	-0.002	-0.149	976.761

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

4.4 Comparison of the SG-GLS and NG-LME Models

The uncertainty of flood quantiles given by the SG-GLS and NG-LME models can be calculated based on the empirical posterior distribution of model parameters given the specific values of covariates. Fig. 4 shows the 95% uncertainty interval of the tails of cumulative probability distributions given by the SG-GLS and NG-LME models with the *P* of the 50^{th}, 90^{th}, and 99^{th} percentiles and the *Ep* of the 50^{th}, 90^{th}, and 99^{th} percentiles. The width of the uncertainty intervals in Fis. 4 indicates that the results of the SG-GLS model have a larger uncertainty than the NG-LME model, especially SM and TJ. Therefore, the NG-LME model can reduce the uncertainty interval of flood quantile compared with the SG-GLS model, especially for the case where the exceedance probability in the tail is very small.

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

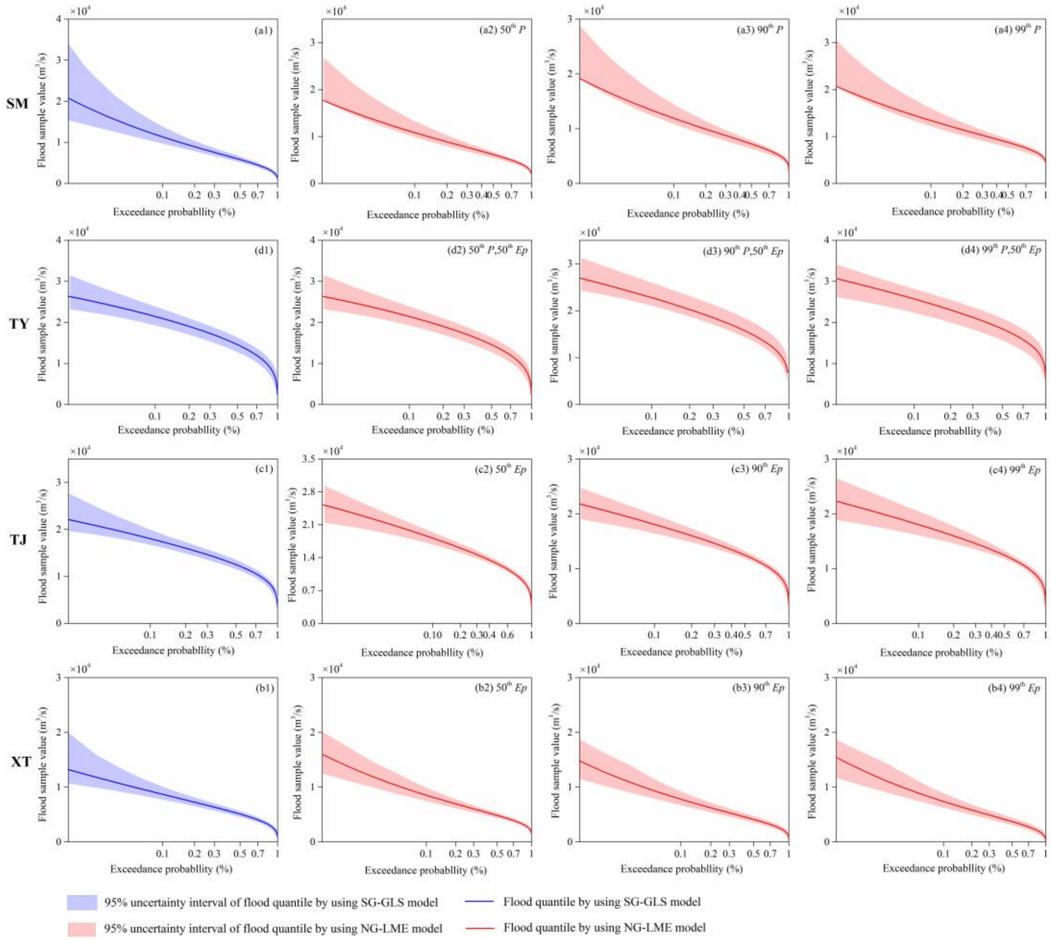

Fig. 4 Uncertainty intervals of the tails of cumulative probability distribution of SM, TY, TJ and XT given by the SG-GLS model and NG-LME model

5 Conclusions

The main purpose of this paper is to study nonstationary RFFA based on the Bayesian method. The proposed method has been applied to four hydrological stations within the Dongting Lake basin, and the major conclusions are as follows:

(1) The performances of nonstationary models outperforms the stationary models by the deviance information criterion (DIC). The DIC values of nonstationary models (i.e. NG-LME) are lower than those of the stationary models (i.e. SG-GLS).

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

(2) The LME model outperforms the GLS model in the regional regression of the flood distribution parameters. This may be due to the random effect term of the LME model being used in the nonstationary condition to consider the errors of regression coefficients in the nonstationary condition, thus greatly reducing the variance of regression model residuals.

(3) Compare with the stationary model, the increase or decrease in the uncertainty of regression coefficients estimation of at-site flood distribution parameters is different from site to site by using the nonstationary model, while the use of nonstationary model reduces the uncertainty in the estimation of the flood quantile.

This study extends the application of the RFFA based on the Bayesian method to the nonstationary condition by using the LME model to infer regional prior information, which can be developed to obtain regional estimates of the statistics of flood at ungauged sites and is helpful for ungauged sites nonstationary flood frequency estimation.

Water Resources Management, https://doi.org/10.1007/s11269-022-03394-9

Author Contributions SH Guo: Conceptualization, Methodology, Software, Formal analysis, Writing - Original Draft. L Xiong: Conceptualization, Resources, Writing - Review & Editing, Project administration, Funding acquisition. J Chen: Writing - Review & Editing. SL Guo: Writing - Review & Editing, Project administration, Funding acquisition. J Xia: Writing - Review & Editing, Project administration, Funding acquisition. L Zeng: Writing - Review & Editing. C-Y Xu: Writing - Review & Editing, Project administration, Funding acquisition. All authors read and approved the final manuscript.

Funding This research was financially supported by National Natural Science Foundation of China (NSFC Grants 41890822 and 51525902), the Ministry of Education "Plan 111" Fund of China (B18037), China Three Gorges Corporation (0799254), and the Research Council of Norway (FRINATEK Project 274310), all of which are greatly appreciated.

Availability of Data All data are available from the corresponding author.

Declarations

Ethical Approval The author promises to compliance with Ethical Standards.

Consent to Participate Informed consent was obtained from all individual participants included in the study.

Consent to Publish The participants have consented to the submission of the research article to the journal.

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

References

Allahbakhshian-Farsani P, Vafakhah M, Khosravi-Farsani H, Hertig E (2020) Regional flood frequency analysis through some machine learning models in semi-arid regions. *Water Resources Management*, 34(9): 2887-2909. https://doi.org/10.1007/s11269-020-02589-2

Cunnane C, Nash J E (1971) Bayesian estimation of frequency of hydrological events. *Mathematical Models in Hydrology*, IAHS Publish, 100: 47-55.

Coles S (2001) An introduction to statistical modeling of extreme values. London, England: Springer. https://doi.org/10.1007/978-1-4471-3675-0

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

Dixit S, JayakumarnAff K V (2022) A non-stationary and probabilistic approach for drought characterization using trivariate and pairwise copula construction (PCC) model. *Water Resources Management*, 36(4):1217-1236. https://doi.org/10.1007/s11269-022-03069-5

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

Griffis V W, Stedinger J R (2007) The use of GLS regression in regional hydrologic analyses. *Journal of Hydrology*, 344(1-2): 82-95.

https://doi.org/10.1016/j.jhydrol.2007.06.023

Guo S, Xiong L, Zha X, Zeng L, Cheng L (2021) Impacts of the Three Gorges Dam on the streamflow fluctuations in the downstream region. *Journal of Hydrology*, 598(3): 126480. https://doi.org/10.1016/j.jhydrol.2021.126480

Gregersen I B, Madsen H, Rosbjerg D, Arnbjerg-Nielsen K (2017). A regional and nonstationary model for partial duration series of extreme rainfall. *Water Resources Research*, 53, 2659-2678. https://doi.org/10.1002/2016WR019554

Gao Q, Li G, Bao J, Wang J (2021) Regional frequency analysis based on precipitation regionalization accounting for temporal variability and a nonstationary index flood model. *Water Resources management*, 35: 4435-4456. https://doi.org/10.1007/s11269-021-02959-4

Han X, Mehrotra R, Sharma A, Rahman A (2022) Incorporating nonstationarity in regional flood frequency analysis procedures to account for climate change impact. *Journal of Hydrology*, 2022, 612: 128235. https://doi.org/10.1016/j.jhydrol.2022.128235

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

Jaffres J, Cuff B, Cuff C, Knott M, Rasmussen C (2022) Hydrological characteristics of Australia: national catchment classification and regional relationships. *Journal of hydrology*, 612: 127969. https://doi.org/10.1016/j.jhydrol.2022.127969

Kuczera G (1982) Combining site-specific and regional information: An empirical

Bayes approach. *Water Resources Research*, 18(2): 306-314. https://doi.org/10.1029/WR018i002p00306

Kendall M G (1975) Rank Correlation Methods. London: Charles Griffin.

Laird N M, Ware J H (1982) Random-effects models for longitudinal data, Biometrics, 38: 963-974.

Laloy E, Vrugt J A (2012) High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high-performance computing. *Water Resources Research*, 48: W01526. https://doi.org/10.1029/2011WR010608

Madsen H, Rosbjerg D (1997) Generalized least squares and empirical bayes estimation in regional partial duration series index-flood modeling, *Water Resources Research*, 33(4): 771-781. https://doi.org/10.1029/96WR03850

Martins E S, Stedinger J R (2000) Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36(3): 737-744. https://doi.org/10.1029/1999WR900330

Merz R, Piock-Ellena U, Blöschl G, Kirnbauer R (2000) Skalierungsprobleme bei der Regionalisierung von Hochwässern (Scale problems in regionalising floods). Endbericht 2000, Österreichische Akademie der Wissenschaften (HÖ-18). Institut für Hydraulik, TU-Wien.

Mann H B (1945) Non-parametric test against trend. Econometic, 13(3): 245-259.

Merz R, Bloschl G (2005) Flood frequency regionalisation—Spatial proximity vs. catchment attributes. *Journal of Hydrology*, 302(1-4): 283-306. https://doi.org/10.1016/j.jhydrol.2004.07.018

Milly P C, Betancourt J, Falkenmark M, Hirsch R M, Kundzewicz Z W, Lettenmaier D

P, Stouffer R J (2008) Stationarity is dead: Whither water management? *Science*, 319(5863): 573-574. https://doi.org/10.1126/science.1151915

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

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

Ouarda T B M J, El‐Adlouni S (2011) Bayesian nonstationary frequency analysis of hydrological variables 1. *Journal of the American Water Resources Association*, 47(3): 496-505. https://doi.org/10.1111/j.1752-1688.2011.00544.x

Pandey K K, Abhash A, Tripathi R P (2020) Revised Dicken's method for flood frequency estimation of Upper Ganga basin. *Journal of Hydrology*, 586: 124904. https://doi.org/10.1016/j.jhydrol.2020.124904

Pinheiro J C, Bates D M (2000) Mixed-Effects Models in S and S-PLUS, *Springer*.

Reis D S Jr, Stedinger J R, Martins E S (2005) Bayesian generalized least squares regression with application to log Pearson type 3 regional skew estimation. *Water Resources Research*, 41: W10419. https://doi.org/10.1029/2004WR003445

Reis D S, Veilleux A G, Lamontagne J R, Stedinger J R, Martins E S (2020) Operational Bayesian GLS regression for regional hydrologic analyses. *Water Resources Research*, 56: e2019WR026940. https://doi.org/10.1029/2019WR026940

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

Read L K, Vogel R M (2016) Hazard function analysis for flood planning under nonstationarity. *Water Resources Research*, 52: 4116-4131. https://doi.org/10.1002/2015WR018370

Reinders J B and Munoz S E (2021) Improvements to flood frequency analysis on Alluvial Rivers using paleoflood data. *Water Resources Research*, 57(4): e2020WR028631. https://doi.org/10.1029/2020WR028631

Razmi A, Mardani-Fard H A, Golian S, Zahmatkesh Z (2022) Time-varying univariate and bivariate frequency analysis of nonstationary extreme sea level for New York City. *Environmental Processes*, 9: 8. https://doi.org/10.1007/s40710-021-00553-9

Razmkhah H, Fararouie A, Ravari A R (2022) Multivariate flood frequency analysis using bivariate copula functions. *Water Resources Management*, 36(2): 729-743. https://doi.org/10.1007/s11269-021-03055-3

Stasinopoulos D M, Rigby R A (2007) Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software*, 23(7): 1-46. Retrieved from. https://www.jstatsoft.org/index.php/jss/article/view/v023i07/v23i07.pdf

Stedinger J R, Tasker G D (1985) Regional hydrologic analysis: 1. Ordinary, weighted, and generalized least squares compared. *Water Resources Research*, 21(9): 1421-1432. https://doi.org/10.1029/WR021i009p01421

Stedinger J R, Tasker G (1986a) Correction to "Regional hydrologic analysis, 1, Ordinary, weighted and generalized least squares compared". *Water Resources Research*, 22(5): 844. https://doi.org/10.1029/WR022i005p00844

Stedinger J R, Tasker G (1986b) Regional hydrologic analysis, 2: Model-error estimators, estimation of sigma and log-Pearson type 3 distributions. *Water Resources Research*, 22(10): 1487-1499. https://doi.org/10.1029/WR022i010p01487

Sarhadi A, Burn D H, Ausín M C, Wiper M P (2016) Time-varying nonstationary multivariate risk analysis using a dynamic Bayesian copula. *Water Resources Research*, 52: 2327-2349. https://doi.org/10.1002/2015WR018525

Spiegelhalter D J, Best N G, Carlin B P, Van Der Linde A (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 64(4): 583-639. https://doi.org/10.1111/1467-9868.00353

Spiegelhalter D J, Best N G, Carlin B P, Van Der Linde A (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 76(3): 485-493. https://doi.org/10.1111/rssb.12062

Stein L, Clark M P, Knoben W J M, Pianosi F, Woods R A (2021) How do climate and catchment attributes influence flood generating processes? A large-sample study for 671 catchments across the contiguous USA. *Water Resources Research*, 57: e2020WR028300. https://doi.org/10.1029/2020WR028300

Thomas D M, Benson M A (1970) Generalization of streamflow characteristics from drainage-basin characteristics. U.S. Geological Survey Water-Supply Paper 1975.

Vicens G J, Rodríguez-Iturbe I, Schaake J C Jr (1975) A Bayesian framework for the use of regional information in hydrology. *Water Resources Research*, 11(3): 405-414. https://doi.org/10.1029/WR011i003p00405

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

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

Viglione A, Merz R, Salinas J L, Blöschl G (2013) Flood frequency hydrology: 3. A Bayesian analysis. *Water Resources Research*, 49: 675-692. https://doi.org/10.1029/2011WR010782

Wang M, Jiang S, Ren L, Xu C-Y, Shi P, Yuan S, Liu Y, Fang X (2022) Nonstationary flood and low flow frequency analysis in the upper reaches of Huaihe River Basin, China, using climatic variables and reservoir index as covariates. *Journal of Hydrology*, 612: 128266. https://doi.org/10.1016/j.jhydrol.2022.128266

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