

# Nonstationary Flood Frequency Analysis for Annual Flood Peak and Volume Series in Both Univariate and Bivariate Domain

Jianzhu Li<sup>1</sup>  · Yuming Lei<sup>1</sup> · Senming Tan<sup>2</sup> ·  
Colin D. Bell<sup>3</sup> · Bernard A. Engel<sup>3</sup> · Yixuan Wang<sup>1</sup>

Received: 18 October 2016 / Accepted: 9 July 2018 /  
Published online: 9 August 2018  
© Springer Nature B.V. 2018

**Abstract** Flood frequency analysis for practical application is traditionally based on the assumption of stationarity, but this assumption has been open to doubt in recent years. A number of studies have focused on the nonstationary flood frequency analysis, and the associated causes of nonstationarity. In this study, the annual maximum flood peak and flood volume of Wangkuai reservoir watershed were used, and several univariate and bivariate models were established to investigate the nonstationary flood frequency, with the distribution parameters changing over the climate indices (NPO, Niño3) and the check dam indices (CDIp, CDIv). In the univariate models, the Weibull distribution performed best and exhibited an undulate behavior for both flood peak and volume, which tended to describe the nonstationarity reasonably well. The bivariate models were constructed using copulas, of which the optimal Weibull distribution in the univariate flood frequency analysis was considered as marginal distributions within the joint distribution. The results showed that the Gumbel-Hougaard copula offered the best joint distribution, and most of the probability isolines crossed each other, which demonstrated the possibility that the occurrence of combinations of the flood peak and volume may be the same under multiple effects of phase changes in the climate patterns and certain human activities (i.e. soil and water conservation). The most likely events were elaborated in diagrams, and the associated combinations of the flood peak and volume were smaller than that estimated by the fixed parameters (i.e. stationary condition) during most of the study period, while it was the opposite in 1956, 1959 and 1963. The results

---

✉ Jianzhu Li  
lijianzhu@tju.edu.cn

<sup>1</sup> State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, Tianjin 300072, China

<sup>2</sup> Pearl River Comprehensive Technology Center (Information Center) of Pearl River Water Resources Commission of the Ministry of Water Resources, Guangzhou 510611, China

<sup>3</sup> Department of Agricultural and Biological Engineering, Purdue University, 225 S. University St, West Lafayette, IN 47907, USA

highlight the necessity of nonstationary flood frequency analysis under various conditions in both univariate and multivariate domains.

**Keywords** Nonstationarity · Univariate model · GAMLSS · Bivariate model · Copulas

## 1 Introduction

Flood frequency analysis is a necessity of hydraulic structure design, flood risk analysis and management. This analysis involves in fitting a probability distribution to the observations of high flows to derive a relationship between the flood magnitude and the exceedance probability, which can then be used to estimate the likelihood of a given flood event in the form of a return period. Traditional methods are based on the assumption of the stationarity of flood time series. However, due to the environmental change (e.g., climate change and human activities), hydrological time series are no longer stationary, and traditional hydrological frequency analysis is rendered invalid (Milly et al. 2008; Um et al. 2017; Cancelliere 2017).

Several studies of nonstationary hydrological frequency analysis have been presented in recent decades. In the nonstationary flood frequency analysis models, the distribution parameters are expressed as a function of different covariates (Ouarda and El-Adlouni 2011; Vasiliades et al. 2015). Due to a long-term trend in the time series, time has been used as an explanatory covariate (Strupczewski et al. 2001). Gül et al. (2014) compared the results of stationary and nonstationary flood frequency analysis using the GEV and GEV-CDN models, respectively, and suggested the estimation of design flood magnitudes for hydraulic structures in the future should account for trends in the time series. However, the parameter's time dependency will alter with the recorded data length. Therefore, the drivers of detected change should be investigated (Merz et al. 2012). Montanari and Koutsoyiannis (2014) also recommended providing scientific evidence of the changes in extreme events before switching to a fully nonstationary modeling paradigm.

Hydrological nonstationarity can be attributed to low-frequency climate variability (Wilson et al. 2010) and human activities such as changes in land use (Hejazi and Markus 2009; Vogel et al. 2011), deforestation, and dam construction. Villarini et al. (2010) used the Generalized Additive Models in Location, Scale and Shape (GAMLSS) in nonstationary modeling, and found the Atlantic Multidecadal Oscillation, the North Atlantic Oscillation and the Mediterranean Index were the most significant predictors of rainfall and temperature in Rome. López and Francés (2013) demonstrated that the dams contribute to nonstationarity by incorporating a reservoir index into nonstationary flood frequency models combined with the climate indices. Prosdocimi et al. (2015) identified the anthropogenic contributors to change high flows by selecting a set of covariates that included the 99th percentile of the daily rainfall, time, and urban extent of each year.

In nonstationary flood frequency analysis, the annual maximum flood peak sampling is widely used. However, in regions where seasonality affects flow, a seasonal maxima approach should be adopted and the nonstationary flood frequency analysis should be made for each seasonal time series (Strupczewski et al. 2009). In the areas where the records are short, the peaks-over-threshold (POT) sampling is used for frequency analysis to extend the data series. Using this sampling method, Tramblay et al. (2013) developed a nonstationary model with climatic covariates for heavy rainfall events. A POT sampling method was also applied by Silva et al. (2012) to obtain the time dependent occurrence rates of the rainfall and flood events, and the authors concluded that traditional flood frequency models should be revised to account for nonstationarities. They also assessed the influence of North Atlantic Oscillation

(NAO) on the occurrence rate of floods, and pointed out the possibility of long-term forecast of flood occurrence rates with NAO indices.

All of the above studies used univariate hydrological frequency analysis. However, because floods can be characterized by flood peak, flood volume and durations, the multivariate flood frequency analysis with copula functions is a good alternative (Zhang and Singh 2006; Zhang and Singh 2007; Karmakar and Simonovic 2009; Sraj et al. 2015; Tsakiris et al. 2015). And Weibull distribution is applicable to flood modeling (Pramanik et al. 2010; Abdollahi et al. 2016), which could be taken as a marginal distribution in Copulas. To date, the studies taking into account nonstationarity in the multivariate frequency analysis are sparse. Bender et al. (2014) fitted the time dependent marginal distribution and the time dependent copulas to 191 years of flood peak and volume data, and found the joint probability varied significantly over time. Considering more than just time as a covariate, Jiang et al. (2015) proposed a reservoir index to perform bivariate frequency analysis to investigate how the reservoirs altered the dependence structure on low flows in the Hanjiang River.

As in the nonstationary frequency analysis, the changes in streamflow series might be related to climate factors. In this paper, for the nonstationary univariate domain, we model the annual maxima of the flood peak and flood volume based on the GAMLSS theory, adopting some climate indices and a reservoir index as covariates. Then we conduct the nonstationary bivariate flood frequency analysis via the Copulas theory. Applying two theories above, the main objective is to explain how flood frequency changes with the climate indices and a reservoir index, which is the innovation as well.

## 2 Study Region and Data

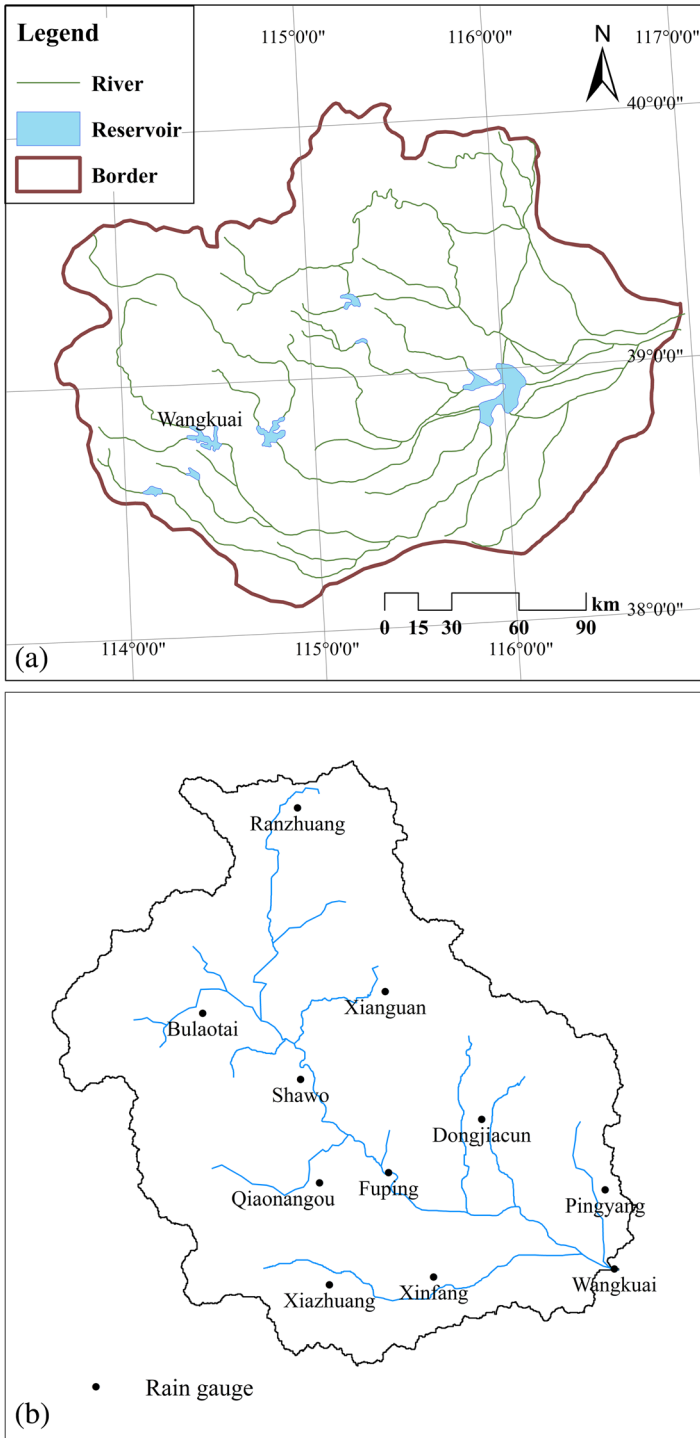
### 2.1 Study Region

Wanguai reservoir, as the study region, is located on the upper reaches of Sha River, which is one of the south branches of Daqinghe River basin, Hebei province, northern China. The reservoir was built between June 1958 and September 1960, controls a drainage area about 3770 km<sup>2</sup>, and the storage capacity is  $13.89 \times 10^8$  m<sup>3</sup>. The reservoir was constructed for comprehensive flood control, irrigation, and electric power generation.

The watershed receives average precipitation of 600 mm, mostly in the summer. (70–80%). The annual mean temperature is 7.4 °C. The watershed is characterized by steep slopes and bedrock outcrops. Forest and grass are dominant on the hill slopes. The urban area is <2%. The rainfall and runoff data have been monitored for a period of 49 years from 1956 to 2004. There are 11 rain gauges available in the study area.

A number of small hydraulic structures were built from 1970 to 1980 for the purpose of flood-control and irrigation in this watershed. For example, in the area controlled by Fuping station (located in the upstream portion of Wanguai reservoir), there are 11 small reservoirs whose total storage capacity is  $591.5 \times 10^4$  m<sup>3</sup>, the largest of which is the Haiyan reservoir with a storage capacity of  $367 \times 10^4$  m<sup>3</sup>.

There was serious soil erosion in this watershed, so measures of soil and water conservation were implemented in the 1980s, most of which were check dams. Small hydraulic structures and the large number of check dams affected the flood processes in this watershed. More than six thousand check dams were built from 1980 to 2000, and the estimated total storage capacity of small hydraulic structures and check dams is about 13.3 million m<sup>3</sup>. The total drainage area of these conservation measures is about 1248 km<sup>2</sup> (Fig. 1).



**Fig. 1** The study area. (a) location of Wangkuai reservoir in the Daqinghe River basin; (b) the Wangkuai reservoir watershed

## 2.2 Data

The flood data of Wangkuai reservoir, provided by the Hydrology and Water Resources Survey Bureau of Hebei Province, were calculated from the reservoir water level and discharge data via water balance equation. The annual maximum flood peak (Q) and flood volume (W) were used herein (Table 1). The Niño3 index with lead time of 3-months and the NPO index with lead time of 6-months were adopted to represent the nonstationarity due to climate change (Li et al. 2015). These Niño3 and NPO indices data were downloaded from Global Climate Observing System (GCOS, [http://www.esrl.noaa.gov/psd/gcos\\_wgsp/](http://www.esrl.noaa.gov/psd/gcos_wgsp/)). Two check dam indices (CDIp and CDIv) were also used as explanatory variables to represent the influences of check dams on the flood peak and volume, respectively. Because the effects of check dams on the flood peak and volume depends on the flood patterns, we analyzed the percentage change of each annual maximum flood peak and volume due to construction of check dams (Fig. 2). All the time series range from 1956 to 2004. For more details of the nonstationary identification of flood data and the calculation of these indices, see Li et al. (2015) and Li and Tan (2015).

C.V.: coefficient of variation, is dimensionless. P is the exceedance probability.

## 3 Methods

### 3.1 Univariate Flood Frequency Analysis Based on GAMLSS Theory

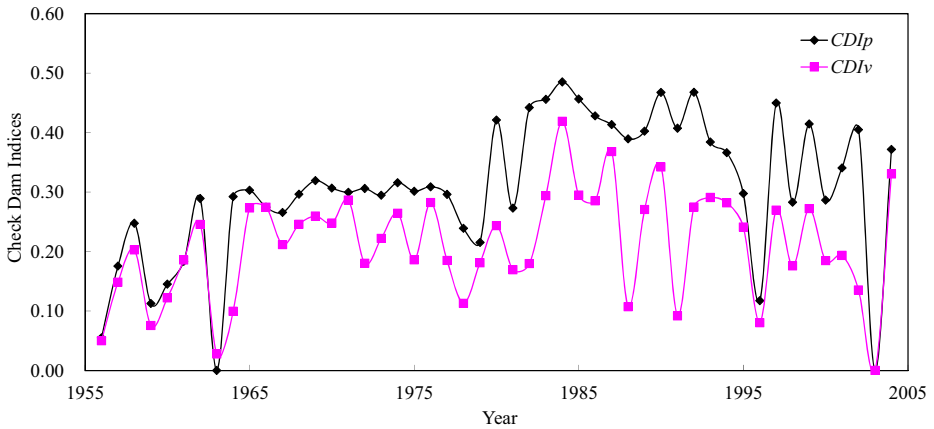
In this paper, a nonstationary model was built based on the GAMLSS theory. The GAMLSS theory, whose full name is “Generalized Additive Models in Location, Scale and Shape”, can be defined as parametric models, assuming the response variable  $y$  given the relationships between the parameters and the explanatory variables (Rigby and Stasinopoulos 2005; Stasinopoulos and Rigby 2007). Within this new framework, four two-parameter distributions (Table 1) were chosen as potential candidates to model the observed flood time series at the Wangkuai reservoir, and the parametric model is given as:

$$\begin{aligned} g_1(\mu) &= X_1\beta_1 \\ g_2(\sigma) &= X_2\beta_2 \end{aligned} \tag{1}$$

Where  $g_k(\cdot)$ ,  $k = 1, 2$  is a known monotonic link function which demonstrates the explanatory variables,  $X_k$  is an explanatory matrix of order  $n \times I_k$ , and  $\beta_k$  is the coefficient of parameter vector of length  $I_k$ , which can be estimated by the RS algorithm. The RS algorithm is a general, modular algorithm. It has an outer cycle which maximizes the penalized likelihood with respect to  $\beta_k$  in the model successively for  $\mu$  and  $\sigma$ . First initialize fitted values  $\mu$  and  $\sigma$ , evaluate the initial linear predictors  $g_k(\cdot)$ . Then start the outer cycle until convergence and update the value of  $k$ . If the change in the likelihood is sufficiently small, end the outer

**Table 1** Main characteristics of the analysed variables Q and W

Variables	period of record	Units	Design Value under the Assumption of Stationarity			
			Mean Value	C.V.	$P = 0.01$	$P = 0.02$
Q	1956–2004	m <sup>3</sup> /s	2000	1.45	14,050	11,300
W	1956–2004	10 <sup>4</sup> m <sup>3</sup>	8700	1.65	55,800	71,800



**Fig. 2** Check dam indices in Wangkuai reservoir watershed

cycle; otherwise, continue the outer cycle. At each calculation the current updated values of all the quantities are used. Readers can refer to Rigby and Stasinopoulos (2005) for details.

For a given data set, models were built with different combinations of explanatory variables, and compared using a goodness of fit criteria. The Akaike information criterion (Akaike 1974) was used to measure the goodness of fit because it penalizes for over-fitting the models. In addition, with no validated models, the quality of model fitting was verified by analyzing the normality and independence of the residuals of each model. The model was deemed adequate if the residuals mean is nearly zero, variance is nearly one, coefficient of skewness and kurtosis are near zero and three, respectively, and the Filliben coefficient (Filliben 1975) is greater than the critical value given a certain sample size.

### 3.2 Bivariate-Joint Flood Frequency Analysis Based on Copulas

Copula functions are efficient mathematical tools for modeling the dependence structure of two or more random variables. Copulas were first introduced by Sklar (1959), and have been widely used in the decades since. For a pair of random variables  $X$  and  $Y$ , with marginal distribution functions  $u = F_X(x) = P(X \leq x)$  and  $v = F_Y(y) = P(Y \leq y)$ , there is a copula function  $C$  to describe the associated relationship:

$$F_{X,Y}(x,y) = C[F_X(x), F_Y(y)] = C(u, v) \tag{2}$$

Where  $F_{X,Y}(x,y)$  is a joint cumulative distribution function (cdf) with margins  $u$  and  $v$ , all  $(u, v) \in (0, 1)^2$ . If  $u$  and  $v$  are continuous, then  $C$  is unique; otherwise,  $C$  is uniquely determined on  $RanF_X \times RanF_Y$  (Nelsen 2006).

One of the frequently-used group of copulas is Archimedean Copulas, with three types: the Clayton, the Frank and the Gumbel copulas (Table 2). Because they are easily constructed and solved, and can capably model the full range of tail dependencies, they are widely used in the multivariate domain of hydrology and also herein.

Table 3 shows three types of Archimedean Copula functions, using  $\theta$  as the copula parameter. The nonstationary models in this study are constructed through copulas composed of two marginal distributions and a copula parameter  $\theta$ . The marginal distributions are determined by the best nonstationary univariate models, outlined in Section 3.1, and only the copula parameter  $\theta$  needs to be solved herein. Several methods have been proposed in the literature for

**Table 2** Implemented GAMLSS distributions with two parameter family (Stasinopoulos et al. 2004)

Distributions	Probability density function	Comments
Gumbel	$f_Y(y \mu, \sigma) = \frac{1}{\sigma} \exp\left[\left(\frac{y}{\sigma}\right) - \exp\left(\frac{y}{\sigma}\right)\right]$	$-\infty < y < \infty$ $-\infty < \mu < \infty$ $\sigma > 0$
Weibull	$f_Y(y \mu, \sigma) = \frac{\sigma y^{\sigma-1}}{\mu^\sigma} \exp\left[-\left(\frac{\sigma y^{\sigma-1}}{\mu^\sigma}\right)^\sigma\right]$	$y > 0, \mu > 0, \sigma > 0$
Gamma	$f_Y(y \mu, \sigma) = \frac{1}{(\sigma^2 \mu)^{1/\sigma^2}} \frac{1}{(\sigma^2 \mu)^{1/\sigma^2}}$	$y > 0, \mu > 0, \sigma > 0$
Log-Normal	$f_Y(y \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{\sqrt{2\pi\sigma^2}}\right\}$	$y > 0, \mu > 0, \sigma > 0$

estimating the copula parameter  $\theta$ , but when the data has already been transformed to the unit hypersquare domain by a parametric estimation of the marginal cdf's, the inference function of margins method (IFM) should be adopted. The IFM is given as:

$$L(u, v; \theta) = \sum \ln C[F_X(x), F_Y(y); \theta] \tag{3}$$

Letting  $\partial L / \partial \theta = 0$ , then  $\theta$  can be calculated.

The procedure for model selection follows two steps. First, the test of fit for the copula functions is checked by the Kolmogorov- Smirnov (K-S) test. Second, from all models that pass the K-S test, the optimum model is then selected using the goodness-of-fit (GoF) statistics of ordinary least squares (OLS) and Akaike information criterion (AIC). The K-S test statistic  $D$ , and the estimates of OLS and AIC are given as:

$$D = \max \left\{ \left| C_k - \frac{i}{n} \right|, \left| C_k - \frac{i-1}{n} \right| \right\} \tag{4}$$

Where  $C_k$  is the copula value of measured sample series,  $i$  is the number of samples which meet the requirements that  $x \leq x_i, y \leq y_i$ , and  $n$  is the length of the sample series.

$$OLS = \sqrt{\frac{1}{n} \sum_{i=1}^n (Pe_i - P_i)^2} \tag{5}$$

$$Pe_i = \frac{i}{n + 1} \tag{6}$$

Where  $Pe_i$  and  $P_i$  are the empirical frequency and the theoretic frequency of measured sample series, respectively.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Pe_i - P_i)^2 \tag{7}$$

$$AIC = n \ln(MSE) + 2m \tag{8}$$

**Table 3** Three types of Archimedean Copula function (Shiau et al. 2007)

Copulas	Copula function	Comments
Gumbel-Hougaard Copula	$C_\theta(u, v) = \exp\{-[(-\ln u)^\theta + (-\ln v)^\theta]^{\frac{1}{\theta}}\}$	$\theta \in [-1, \infty)$
Clayton Copula	$C_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{\frac{1}{\theta}}$	$\theta \in (0, \infty)$
Frank Copula	$C_\theta(u, v) = -\frac{1}{\theta} \ln\left\{1 + \frac{1}{\theta}\right\}$	$\theta \in R$

where *MSE* is the sum of squared residuals of model fitting, *m* is the number of the model parameters. The lower the value of AIC is, the better the model will fit.

The concept of return period in stationary frequency analysis is prone to misconception and misuse, and attempts to solve this problem have been made with new methods, but with limited success (Olsen et al. 1998; Salas and Obeysekera 2014; Serinaldi 2015). Since the return period is based on probability, we explore the effect of nonstationarity on flood data focusing on the exceedance probability for the present study. In the bivariate-joint domain, the OR-joint exceedance probability is the likelihood that at least one of the hydrologic variables *X* and *Y* exceeds the values *x* and *y*, respectively:  $P^{\cup} = P(X > x \cup Y > y)$ . While the AND-joint exceedance probability is the likelihood that both variables *X* and *Y* exceed the values *x* and *y*, respectively:  $P^{\cap} = P(X > x \cap Y > y)$ . The  $P^{\cup}$  and  $P^{\cap}$  are given as:

$$P^{\cup} = 1 - C[F_X(x), F_Y(y)] \tag{9}$$

$$P^{\cap} = 1 - F_X(x) - F_Y(y) + C[F_X(x), F_Y(y)] \tag{10}$$

All data copula  $C(u, v)$  on the same probability level have the same exceedance probability. However, at least one combination of a given probability is more likely than others, namely the most-likely events. Therefore, the most-likely events can be selected as the point with the largest joint probability on the level curve, which is given by Gräler et al. (2013):

$$(u, v) = \underset{C_{uv}(u,v)=k}{\operatorname{argmax}} f_{XY}(F_X^{-1}(u), F_Y^{-1}(v)) \quad x = F_X^{-1}(u), \quad y = F_Y^{-1}(v) \tag{11}$$

Where *k* is a given value of copula, *x* and *y* can be calculated by the inverse cdf of the marginal distributions.

## 4 Results

### 4.1 Nonstationary Model of Univariate Flood Frequency Analysis

Two nonstationary models of the independent response variables *Q* and *W* were constructed based on GAMLSS, using the Niño3 index, NPO index and CDIP/CDIV as the explanatory variables. First, the functional relationships of the explanatory variables and the distribution parameters were estimated using four distributions (Table 4). The Weibull distribution performed best for both *Q* and *W* as they produced the lowest values of AIC. The functional relationships which were generated using the Weibull distribution were written as Eq. (12) for *Q* and Eq. (13) for *W*. All coefficients of the distribution parameters were proven to satisfy the significance level 0.05 via a T test. Second, the residuals of optimal distributions are shown in Table 5. With an approximate mean of 0, variance of 1, skewness coefficient of 0, kurtosis coefficient of 3, and the Filliben coefficients both greater than 0.975 for a sample size of 49, the residuals were deemed acceptable. Also, visual inspection of the worm plot and the normal QQ plot (not shown herein) exhibited an approximately normal distribution, which further proved the satisfactory performance of the two Weibull distributions.



**Table 4** The functional relationships of the explanatory variables and the distribution parameters

Distributions	Q			W		
	AIC	$\theta_1$	$\theta_2$	AIC	$\theta_1$	$\theta_2$
Gumbel	825.47	CDIp NPO	CDIp	1045.70	CDIv	NINO3
Weibull	<b>771.73</b>	CDIp NPO	–	<b>893.24</b>	CDIv NPO	CDIv
Gamma	772.51	CDIp NPO	–	897.11	CDIv NPO	CDIv
Log-Normal	783.82	NPO	–	918.78	CDIv NINO3	–

$$\theta_1 = \log(\mu) = -135.268 - 5.418CDI_p + 0.142I_{NPO} \quad (12)$$

$$\theta_2 = \sigma = 1.263$$

$$\theta_1 = \log(\mu) = -86.221 - 7.700CDI_v + 0.095I_{NPO} = \sigma = 1.385 - 2.565CDI_v \quad (13)$$

$$\theta_2$$

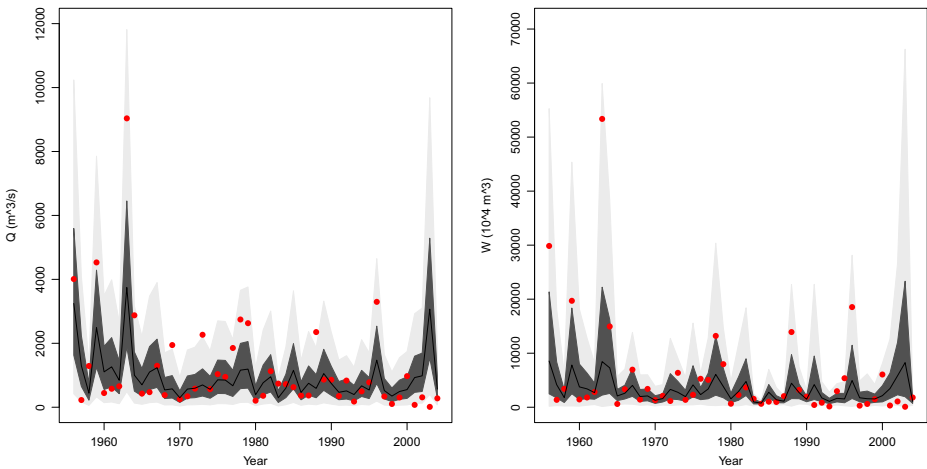
A summary of the associated results of nonstationary models are shown in Fig. 3. Ignoring the outliers in the tail of sequence caused by relatively small flood events, the undulating behavior tended to fit the flood data reasonably well. For the 95% flood quantile, the Q ranged from a minimum value of 832 m<sup>3</sup>/s in 1983 to a maximum value of 11,468 m<sup>3</sup>/s in 1963, while the W ranged from a minimum value of 1167·10<sup>4</sup> m<sup>3</sup> in 1983 to a maximum value of 59,818·10<sup>4</sup> m<sup>3</sup> in 1963. This undulating behavior and the variation of a certain flood quantile can be explained by the changes in precipitation patterns caused by the phase transition of oceanic-atmospheric patterns from 1970 to 1990, and the changes in watershed runoff yield and concentration caused by the construction of small reservoirs and check dams upstream the Wangkuai reservoir from 1980 to 2000.

### 4.2 Nonstationary Model of Bivariate-Joint Flood Frequency Analysis

In this section, taking the two optimal distributions mentioned in Section 4.1 as the marginal distributions, three nonstationary models of bivariate-joint frequency analysis were constructed through copulas. The copula parameters were estimated and each model evaluated for goodness-of-fit. As shown in Table 6, the K-S test statistic D should be less than 0.1943 at a 5% significance level for the sample size of 49, and thus only the G-H copula passed the K-S test. Although the Clayton copula had the minimum value of OLS and AIC, it failed the K-S

**Table 5** The residuals of optimal distributions

Flood series	Optimal distribution	Mean	Variance	Coefficient of skewness	Coefficient of kurtosis	Filliben coefficient
Q	Weibull	0.002	1.056	−0.303	3.637	0.9841
W	Weibull	−0.048	1.051	−0.166	2.051	0.9868



**Fig. 3** Summary of results of nonstationary models with GAMLSS implementation. The solid black line represents 0.5 quantile; the dark grey region is the area between 0.25 and 0.75 quantiles; while the light grey region is the area between 0.05 and 0.95 quantiles

test. Therefore, the G-H copula with parameter  $\theta = 1.6110$  was considered as the optimal copula for the joint distribution of the flood peak and volume.

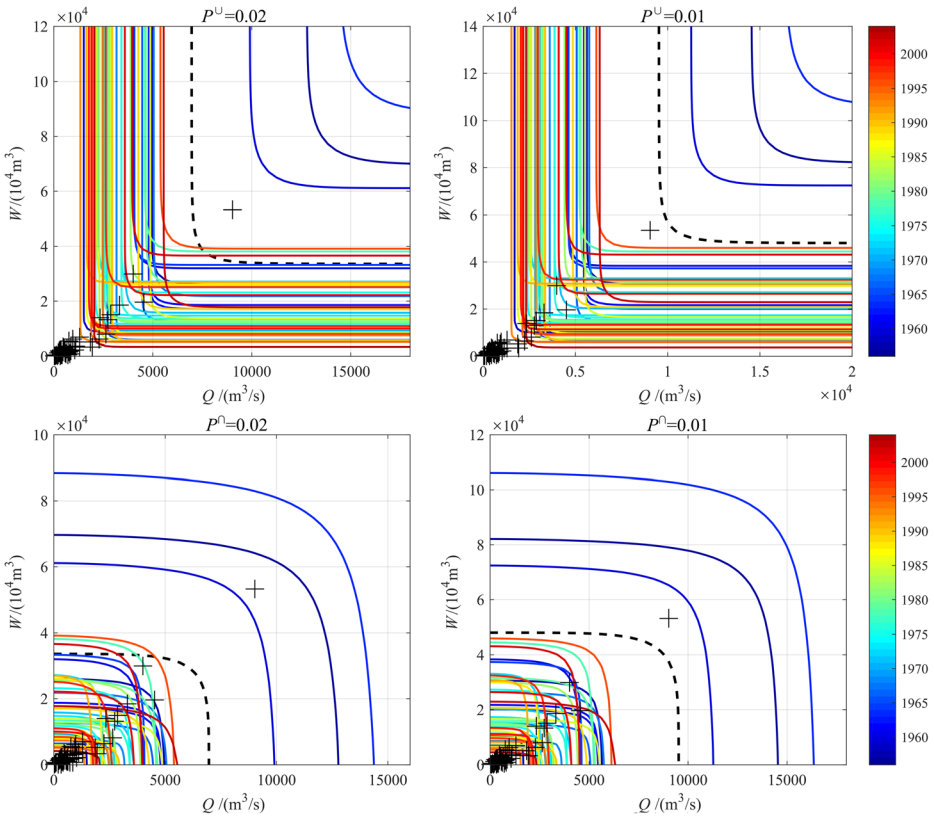
In this study, the focus was on events of the OR-joint exceedance probability of  $P^{\cup} = 0.01, 0.02$  and the AND-joint exceedance probability of  $P^{\cap} = 0.01, 0.02$ , as described in Eq. (9) and Eq. (10). Figure 4 showed the results by illustrating different exceedance probability isolines from 1956 to 2004 (color solid line). For the sake of comparison, the probability isolines derived from a constant parameter of the marginal distributions (i.e. stationary condition) are shown in Fig. 4 (black dotted line) as well.

Under the multiple effects caused by the phase transition of the climate patterns and the implementation of water and soil conservation, the isolines crossed each other – that was the occurrence of the flood events (including flood peak and volume) would be the same, which made the flood frequency more complicated to analyze. In addition, the flood events in the large flood years (i.e. 1956, 1959, 1963) were obviously greater than that estimated under the stationary condition, while they were the opposite in other years. Therefore, under such multiple effects, it tended to be risky in the large flood years upstream of the Wangkuai reservoir, while it might be rather safe in the small flood years.

The “most likely” design events derived from Eq. (11) had their maximum value of likelihood function on each probability-isoline, and all the combinations of Q and W were illustrated in Fig. 5. They had the similar undulating behavior as the univariate models shown in Fig. 3. Under the same exceedance probability, the most likely design events were smaller

**Table 6** The estimation of the copula parameters and the associated goodness-of-fit test and model selection

Copulas	Copula parameter ( $\theta$ )	Indices of goodness-of-fit and model selection		
		D (K-S)	OLS	AIC
Gumbel-Hougaard Copula	1.6110	0.1567	0.0137	-418.0851
Clayton Copula	0.8306	0.2313	0.0127	-425.8435
Frank Copula	4.6885	0.1966	0.0141	-415.5480



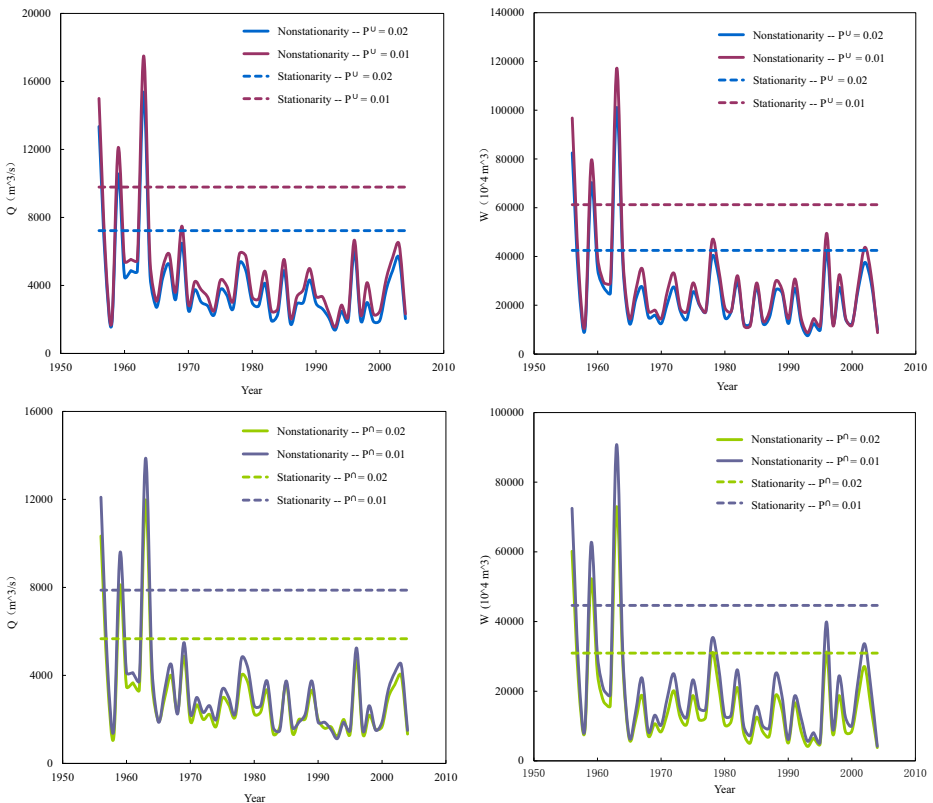
**Fig. 4** Results of the bivariate-joint probability-isolines

than that estimated by the marginal distributions and copulas both with fixed parameters (i.e. stationary condition) in most of the study period, while it was the opposite in 1956, 1959 and 1963. Take  $P^D = 0.01$  as example, the most likely events ranged from 1125 m<sup>3</sup>/s in 1993 to 13,875 m<sup>3</sup>/s in 1963 for Q and from 4200·10<sup>4</sup> m<sup>3</sup> to 90,730·10<sup>4</sup> m<sup>3</sup> for W, while it was a single value under the stationary condition, that is 7875 m<sup>3</sup>/s for Q and 44,630·10<sup>4</sup> m<sup>3</sup> for W.

### 5 Discussion

Flood is a phenomenon characterized with flood peak, volume and duration. Univariate or bivariate stationary flood frequency analysis were conducted in previous studies. However, with the climate change and land surface change, nonstationary bivariate frequency analysis is necessary for flood control.

In this paper, we made bivariate nonstationary flood frequency analysis on the basis of univariate nonstationary flood frequency analysis. The nonstationary flood peak and volume marginal probability distributions were linked by Copula function. But the copula function we applied is stationary, which means  $\theta$  in the function is a constant. This is consistent with the study of Jiang et al. (2015). Bender et al. (2014) analyzed the joint probability of flood peak and volume, considering  $\theta$  as a time-dependent parameter. The improvement in this paper



**Fig. 5** Results of the most likely design events

emphasizes that all the distribution parameters are expressed by climate indices and check dam index, which combines more variables in frequency analysis. Considering  $\theta$  as a nonstationary variable is one issue for the next work.

The aim of flood frequency analysis is to provide flood return periods for hydraulic design and flood control. This has been done in stationary flood frequency analysis including both univariate and bivariate domain. Return period concept has been presented in univariate flood characteristic such as flood peak or volume under nonstationarity (Cooley 2013). However, the concept of flood return period needs to be defined under nonstationary multi-variate flood frequency analysis. Therefore, this is another issue to be further studied.

## 6 Conclusions

In a changing environment, the impacts of climate change and human activities (i.e. check dam construction) have altered the hydrologic mechanisms, which makes the assumption of stationarity questionable. Likewise, for the Wangkuai watershed, the  $Q$  and  $W$  have been identified as nonstationary (Li et al. 2015), which provided further evidence that “stationarity is dead” (Milly et al. 2008).

In this paper, several univariate nonstationary models based on the GAMLSS theory were used to conduct the flood frequency analysis, taking the climate indices (NPO with a 6-month lead time, Niño3 with a 3-month lead time) and the check dam indices (CDIp, CDIV) as explanatory variables. In the nonstationary distribution model test, the Weibull distribution offered the best overall performance for the Q and W with the minimum AIC values, and tended to fit the flood data reasonably well. Thus, the flood events should be a dynamic changing process under the effects of certain climate patterns and human activities.

Using the optimal univariate models mentioned before, the copula functions were used to construct the dependence structure of the Q and W, of which the Weibull distribution was considered as the marginal distribution for the annual maximum flood peak and flood volume. The results showed that only the Gumel-Hougaard copula provided a statistically significant joint distribution. Most of the probability isolines crossed each other, which illustrated the possibility that the combinations of the Q and W are under the similar influences of the climate patterns and the soil and water conservation practices. The most likely events had similar undulating behavior as the univariate models, and the associated combinations of the Q and W were smaller than that estimated by the fixed parameters (i.e. stationary condition) in the copulas during most of the study period. However, this was opposite in the large flood years of 1956, 1959 and 1963.

**Acknowledgements** This work was supported by National Natural Science Foundation of China (No. 51209157). We are also grateful to Hydrology and Water Resource Survey Bureau of Hebei Province for providing the hydrometeorological data.

## References

- Abdollahi K, Guzman P, Huysmans M et al (2016) Rainfall-runoff modelling using a spatially distributed electrical circuit analogue. *Nat Hazards* 2:1279–1300
- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19:716–723
- Bender J, Wahl T, Jensen J (2014) Multivariate design in the presence of non-stationarity. *J Hydrol* 514:123–130
- Cancelliere A (2017) Non stationary analysis of extreme events. *Water Resour Manag* 31:3097–3110
- Cooley D (2013) Return periods and return levels under climate change: extremes in a changing climate. Springer, Netherlands, pp 97–114
- Filliben JJ (1975) The probability plot correlation coefficient test for normality. *Technometrics* 17:111–117
- Gräler B, van den Berg MJ, Vandenberghe S et al (2013) Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrol Earth Syst Sci* 17:1281–1296
- Gül GO, Aşıkoğlu ÖL, Gül A et al (2014) Nonstationarity in flood time series. *J Hydrol Eng* 19:1349–1360
- Hejazi MI, Markus M (2009) Impacts of urbanization and climate variability on floods in northeastern Illinois. *J Hydrol Eng* 14:606–616
- Jiang C, Xiong LH, Xu CY et al (2015) Bivariate frequency analysis of nonstationary low-flow series based on the time-varying copula. *Hydrol Process* 29:1521–1534
- Karmakar S, Simonovic SP (2009) Bivariate flood frequency analysis. Part 2: a copula-based approach with mixed marginal distributions. *J Flood Risk Manag* 2:32–44
- Li JZ, Liu XY, Chen FL (2015) Evaluation of nonstationarity in annual maximum flood series and the associations with large-scale climate patterns and human activities. *Water Resour Manag* 29:1653–1668
- Li JZ, Tan SM (2015) Nonstationary flood frequency analysis for annual flood peak series, adopting climate indices and check dam index as covariates. *Water Resour Manag* 29:5533–5550
- López J, Francés F (2013) Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates. *Hydrol Earth Syst Sci* 17:3189–3203
- Merz B, Vorogushyn S, Uhlemann S et al (2012) HESS opinions more efforts and scientific rigour are needed to attribute trends in flood time series. *Hydrol Earth Syst Sci* 16(5):1379–1387

- Milly PCD, Betancourt J, Falkenmark M et al (2008) Stationarity is dead: whither water management? *Science* 319:573–574
- Montanari A, Koutsoyiannis D (2014) Modeling and mitigating natural hazards: stationarity is immortal. *Water Resour Res* 50:9748–9756
- Nelsen RB (2006) *An introduction to copulas*. Springer
- Olsen JR, Lambert JH, Haines YY (1998) Risk of extreme events under nonstationary conditions. *Risk Anal* 18: 497–510
- Pramanik N, Panda RK, Sen D (2010) Development of design flood hydrographs using probability density functions. *Hydrol Process* 24:415–428
- Prosdocimi I, Kjeldsen TR, Miller JD (2015) Detection and attribution of urbanization effect on flood extremes using nonstationary flood-frequency models. *Water Resour Res* 51:4244–4262
- Ouarda TBMJ, El-Adlouni S (2011) Bayesian nonstationary frequency analysis of hydrological variables. *J Am Water Resour Assoc* 47:497–506
- Rigby RA, Stasinopoulos DM (2005) Generalized additive models for location, scale and shape. *J Royal Stat Soc: Ser C (Appl Stat)*. 54:507–554
- Silva AT, Portela MM, Naghettini M (2012) Nonstationarities in the occurrence rates of flood events in Portuguese watersheds. *Hydrol Earth Syst Sci* 16:241–254
- Salas J, Obeyseker J (2014) Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. *J Hydrol Eng* 19:554–568
- Serinaldi F (2015) Dismissing return periods. *Stoch Env Res Risk A* 29:1179–1189
- Shiau JT, Wang HY, Tsai CT (2007) Bivariate frequency analysis of floods using copulas. *JAWRA J Am Water Resour Assoc* 42:1549–1564
- Sklar A (1959) Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique Université de Paris* 8:229–231
- Sraj M, Bezak N, Brilly M (2015) Bivariate flood frequency analysis using the copula function: a case study of the Litija station on the Sava river. *Hydrol Process* 29:225–238
- Stasinopoulos DM, Rigby RA, Akantziliotou C (2004) Instructions on how to use the GAMLSS package in R. Technical report 02/04. STORM Research Centre, London Metropolitan University, London
- Stasinopoulos DM, Rigby RA (2007) Generalized additive models for location scale and shape (GAMLSS) in R. *J Stat Softw* 23:1–46
- Strupczewski WG, Singh VP, Mitosek HT (2001) Non-stationary approach to at-site flood frequency modelling. III. Flood analysis of polish rivers. *J Hydrol* 248:152–167
- Strupczewski WG, Kochanek K, Feluch W et al (2009) On seasonal approach to nonstationary flood frequency analysis. *Phys Chem Earth* 34:612–618
- Tramblay Y, Neppel L, Carreau J et al (2013) Non-stationary frequency analysis of heavy rainfall events in southern France. *Hydrol Sci J* 58:280–294
- Tsakiris G, Kordalis N, Tsakiris V (2015) Flood double frequency analysis: 2D-Archimedean copulas vs bivariate probability distributions. *Environ Process* 2:705–716
- Um MJ, Heo JH, Markus M et al. (2017) Performance evaluation of four statistical tests for trend and non-stationarity and assessment of observed and projected annual maximum precipitation series in major United States cities. *Water resources management* (accepted)
- Vasiliades L, Galiatsatou P, Loukas A (2015) Nonstationary frequency analysis of annual maximum rainfall using climate covariates. *Water Resour Manag* 29:339–358
- Villarini G, Smith JA, Napolitano F (2010) Nonstationary modeling of a long record of rainfall and temperature over Rome. *Adv Water Resour* 33:1256–1267
- Vogel RM, Yaindl C, Walter M (2011) Nonstationary: flood magnification and recurrence reduction factors in the unite states. *J Am Water Resour Assoc* 47:464–474
- Wilson D, Hisdal H, Lawrence D (2010) Has streamflow changed in the Nordic countries?— recent trends and comparisons to hydrological projections. *J Hydrol* 394:334–346
- Zhang L, Singh VP (2006) Bivariate flood frequency analysis using the copula method. *J Hydrol Eng* 11:150–164
- Zhang L, Singh VP (2007) Trivariate flood frequency analysis using the Gumbel-Hougaard copula. *J Hydrol Eng* 12:431–439