

OPEN

# Nonstationary flood coincidence risk analysis using time-varying copula functions

Ying Feng<sup>1,2</sup>, Peng Shi<sup>1,2</sup>, Simin Qu<sup>1,2\*</sup>, Shiyu Mou<sup>1,2</sup>, Chen Chen<sup>1,2</sup> & Fengcheng Dong<sup>1,2</sup>

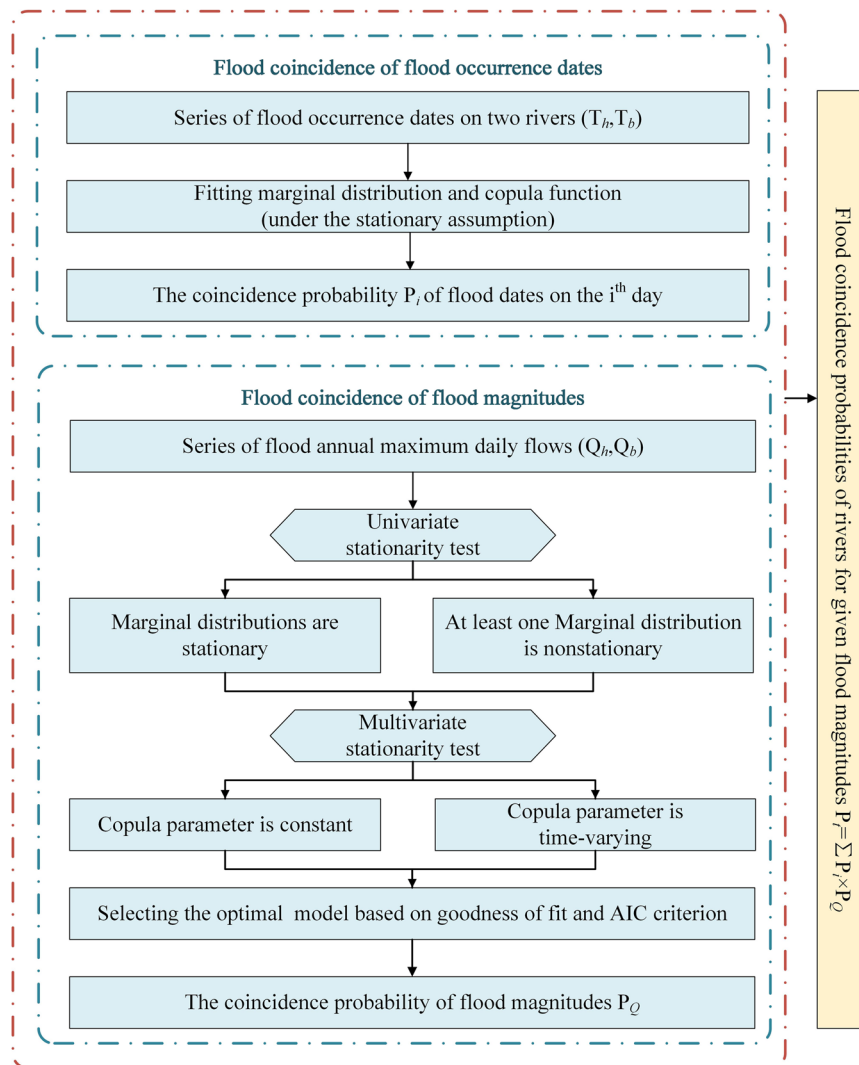
The coincidence of flood flows in a mainstream and its tributaries may lead to catastrophic floods. In this paper, we investigated the flood coincidence risk under nonstationary conditions arising from climate changes. The coincidence probabilities considering flood occurrence dates and flood magnitudes were calculated using nonstationary multivariate models and compared with those from stationary models. In addition, the “most likely” design based on copula theory was used to provide the most likely flood coincidence scenarios. The Huai River and Hong River were selected as case studies. The results show that the highest probabilities of flood coincidence occur in mid-July. The marginal distributions for the flood magnitudes of the two rivers are nonstationary, and time-varying copulas provide a better fit than stationary copulas for the dependence structure of the flood magnitudes. Considering the annual coincidence probabilities for given flood magnitudes and the “most likely” design, the stationary model may underestimate the risk of flood coincidence in wet years or overestimate this risk in dry years. Therefore, it is necessary to use nonstationary models in climate change scenarios.

The term flood coincidence is used to denote the simultaneous occurrence of floods in two (or more) rivers. The coincidence of flood flows in a mainstream and its tributaries may lead to catastrophic floods<sup>1</sup>. Therefore, assessing the risk of flood coincidence for the main river and its tributaries is critical for flood control and water project operations. As flood events are characterized by flood occurrence dates and flood magnitudes, both of these factors should be taken into account when analyzing the flood coincidence risk<sup>2</sup>. In addition, the analysis of flood coincidence involves at least two rivers. For these reasons, a multivariate hydrological analysis is needed that considers the dependence among flood variables<sup>3–5</sup>.

Traditionally, multivariate probability distributions are derived using various assumptions, e.g., the same type of marginal distribution or independence of the variables is assumed<sup>6</sup>. In addition, considering multivariate models from this traditional perspective, mathematical formulations are often complicated when more than two variables are involved<sup>7</sup>. For these reasons, a new method of determining the multivariate probability distribution based on copula functions was proposed by Sklar<sup>8</sup>. The copulas describe and model the dependence structure among random variables, independently of the margins involved. Due to their flexibility of construction, copula functions have been widely used in multivariate hydrological frequency analyses in recent years<sup>9–17</sup>, especially in flood coincidence risk analyses<sup>1,18,19</sup>.

The risk of flood coincidence has mainly been analyzed under the stationarity assumption. In other words, both the marginal distribution and the copula function are modeled with fixed moments and parameters<sup>18–20</sup>. However, climate change and anthropogenic activities have changed the statistical characteristics of hydrological series and the dependence structure of the variables<sup>21,22</sup>. As a result, increasing attention is being paid to the development of nonstationary multivariate models with copula functions<sup>23–25</sup>. In the univariate case, nonstationary models have been widely applied<sup>26–28</sup>. Liang, *et al.*<sup>29</sup> grouped nonstationary flood frequency methods into two types: indirect and direct methods. Direct methods have been widely used because they do not require the restoration of hydrological series. In the multivariate case, copulas have been used to describe the dependence structure of different series<sup>30–32</sup>. Chebana, *et al.*<sup>33</sup> discussed the use of copula functions with time-varying parameters in the case of a changing dependence structure for the investigated variables. Sarhadi, *et al.*<sup>34</sup> defined the copula parameter as a deterministic function of time. Jiang, *et al.*<sup>32</sup> compared time-varying copula models with time or a reservoir index as the covariate. However, to the best of our knowledge, few studies in the hydrological

<sup>1</sup>State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, 210098, China. <sup>2</sup>College of Hydrology and Water Resources, Hohai University, Nanjing, 210098, China. \*email: [waniy@hhu.edu.cn](mailto:waniy@hhu.edu.cn)



**Figure 1.** The procedure used to develop the flood coincidence model.

field have defined the form of the copula parameter as an autoregressive moving average model (ARMA), which can capture variation in dependence<sup>35</sup>.

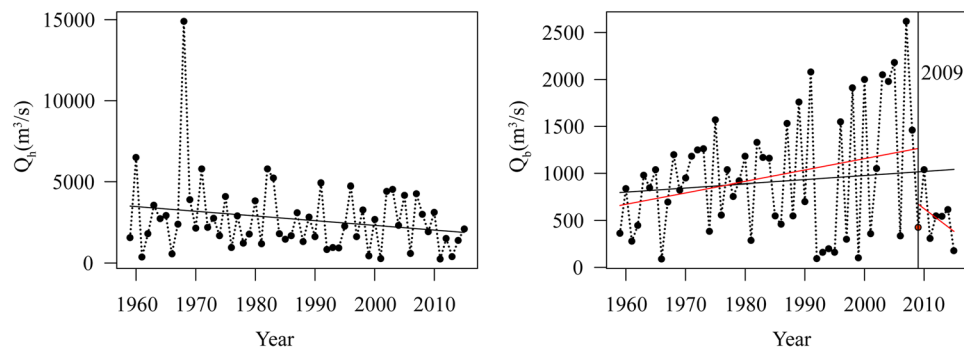
The objective of this study is to apply a time-varying copula to analyze the flood coincidence risk of rivers, considering the nonstationarity of flood magnitudes and the dependence structure among variables (Fig. 1). For this purpose, the mixed von Mises distribution was used as the marginal distribution of flood occurrence dates. Then, the Generalized Additive Models for Location, Scale, and Shape (GAMLSS) model<sup>36,37</sup> (with rainfall as the covariate in this study) was selected to obtain the marginal distribution of flood magnitudes. Finally, a static copula and a time-varying copula were chosen as candidates to obtain the joint distribution of flood occurrence dates and flood magnitudes. In this study, we (1) assess the nonstationarity of the flood magnitudes and the dependence structure among variables; (2) fit the marginal distributions of flood occurrence dates and flood magnitudes; (3) develop a joint distribution of flood occurrence dates and flood magnitudes; and (4) discuss the risk of flood coincidence by considering annual coincidence probabilities for given flood magnitudes and the “most likely” design<sup>38</sup>.

### Study Area and Data

The Huai River Basin (N30°55′–36°36′, E111°55′–121°25′), which is composed of many tributaries, is the sixth largest river basin in China. The basin is located in the transitional zone between semiarid and semihumid climates (Fig. 2). Previous studies have shown that extreme rainfall has increased at most stations in the Huai River Basin<sup>39–41</sup>, resulting in more turbulent flows in the main stream and its tributaries during the flood season. Therefore, it is necessary to analyze the risk of flood coincidence in the Huai River Basin.

The study area refers to the upper reaches of the Huai River Basin above the Wangjiaba station, with an area of  $2.82 \times 10^4$  km<sup>2</sup> (Fig. 2). The proposed method was applied for flood coincidence analysis for two rivers within the region: one is the upper reach of the main stream of the Huai River, with a catchment area of  $1.58 \times 10^4$  km<sup>2</sup>, and the other is the Hong River, which has a catchment area of  $1.15 \times 10^4$  km<sup>2</sup>. The length of the channel from Huaibin to Wangjiaba is 27 km, which drains an area of 900 km<sup>2</sup>. There are seven rainfall stations in the





**Figure 3.** Time series of  $Q_h$  and  $Q_b$  data. The vertical solid line indicates the possible mean change point, and the solid red lines indicate the trends before and after the change point.

Gauging station	Parameters										$p$ -Value
	$\mu_1$	$\mu_2$	$\mu_3$	$k_1$	$k_2$	$k_3$	$p_1$	$p_2$	$p_3$		
Huaibin	3.01	5.06	/	1.33	3.22	/	0.93	0.07	/	0.75	
Bantai	0.34	3.03	4.36	1.13	4.17	2.42	0.28	0.42	0.30	0.81	

**Table 2.** Parameters and goodness-of-fit results for the mixed von Mises distribution. The  $p$ -value is the approximate Monte Carlo goodness-of-fit test  $p$ -value (based on Kolmogorov–Smirnov statistics).

The results of the change point test show that the mean of  $Q_h$  is constant, and the variance of  $Q_h$  displays an abrupt change in 2009. For  $Q_b$ , the change point in the mean occur in 2009, and the change point in the variance is observed in 1985. In addition, a change point in the dependence structure between  $Q_h$  and  $Q_b$  occur in 2010. According to the MK test, the series of  $Q_h$  presents a significant downward trend at the 0.05 significance level. For  $Q_b$ , there is a significant upward trend (Fig. 3).

The above analyses demonstrate that the flood magnitudes in the two rivers and the dependence structure are nonstationary. In additions, the flood occurrence dates in the two rivers and the dependence structure between them are both stationary.

**Estimation of marginal distribution.** *Marginal distribution of flood occurrence dates.* In this study, the mixed von Mises distribution with constant parameters was selected as the marginal distribution of the flood occurrence dates. The parameters of the mixed von Mises distribution were estimated by the maximum likelihood method. Table 2 summarizes the values of parameters and the goodness-of-fit results for the mixed von Mises distribution of flood occurrence dates for the two rivers: The  $p$ -values of the KS test were both larger than 5% (Table 2), which supports the validity of the assumed models.

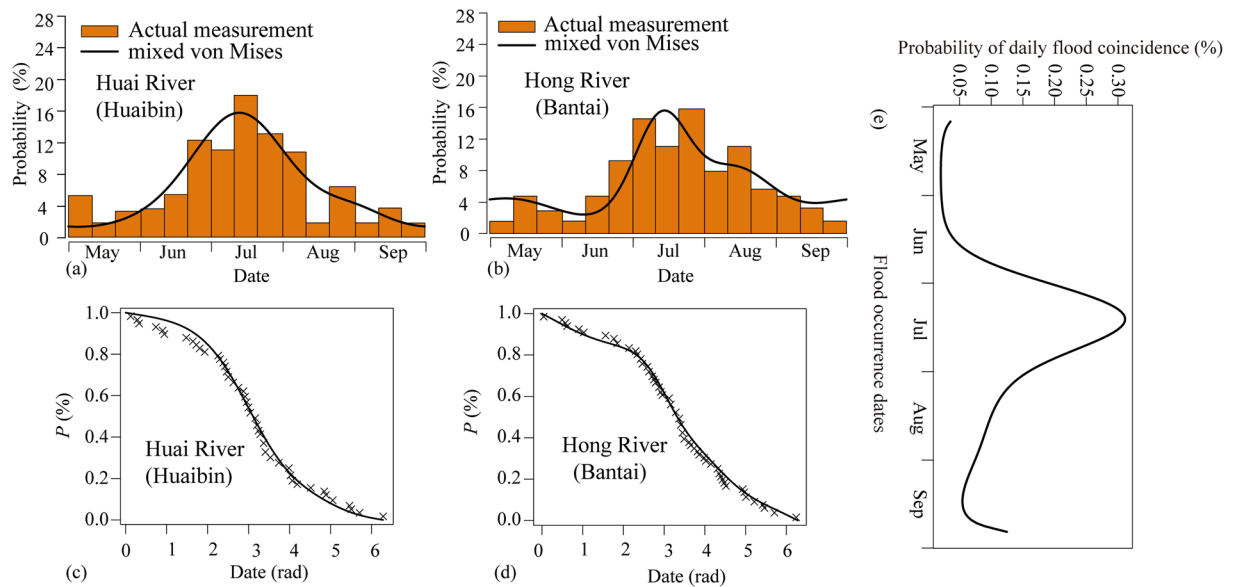
The frequency histograms of flood occurrence dates fitted by the mixed von Mises distribution are presented in Fig. 4(a,b). The marginal cumulative distribution function (CDF) curves of the flood occurrence dates are shown in Fig. 4(c,d), in which the lines (theoretical distribution) intersect with the observed empirical frequencies. Figure 4(a–d) indicate that the theoretical distribution fits the observed data well. In additions, the highest relative frequencies for the two rivers both occur in July, indicating the floods in these two rivers are more likely to occur during this period.

*Marginal distribution of flood magnitudes.* In this section, both stationary models and nonstationary models with a rainfall covariate were applied to build the marginal distributions of flood magnitudes in the two rivers. Five probability distributions, including the gamma, Weibull, lognormal, Gumbel, and general extreme value (GEV) distributions, were selected as the candidate marginal distributions (Table S1).

Under the stationary assumption, the results of the five distributions in fitting the two series are presented in Table S2. According to the Akaike information criterion (AIC) minimization method, the gamma distribution with constant parameters is the optimal distribution for  $Q_h$ , and the Weibull distribution with constant parameters is the best distribution for  $Q_b$ .

Under the nonstationary assumption, a series of statistical analyses (Table S3) indicate that the Weibull distribution with a rainfall covariate is the best-fitted distribution for  $Q_h$  and that the gamma distribution with a rainfall covariate is the best choice for  $Q_b$ .

Table 3 summarizes the performance of the four optimal distributions in fitting the two series. According to AIC minimization method, nonstationary models provide a better fit than stationary models for both  $Q_h$  and  $Q_b$ . Consequently, the nonstationary models with a rainfall covariate are selected as the marginal distributions of flood magnitudes. The worm plots for the selected models [Fig. 5(a,b)] show that all the points fall in the 95% confidence interval (i.e., the upper and lower gray dotted lines). In the quantile–quantile (QQ) plots [Fig. 5(c,d)], all the points are basically distributed along a straight line at a 45 degree angle. Figure 5 indicates that the actual residuals of the selected models are in good agreement with the theoretical residuals.



**Figure 4.** Fitting plots of the mixed von Mises function (a–d) and the coincidence probabilities for flood occurrence dates (e).

Series	Type	Distribution	Distribution parameters		p-Value	AIC
			$\mu$	$\theta$		
$Q_h$	Stationary	GA	1856	1.455	0.73	1010
	Nonstationary	WEI	$\exp(7.894)$	$\exp(0.607 - 0.065P_h)$	0.89	946
$Q_b$	Stationary	WEI	1018	1.486	0.57	884
	Nonstationary	GA	$\exp(5.440 + 0.308P_b)$	$\exp(-0.024 - 0.154P_b)$	0.32	845

**Table 3.** Performance of the four optimal distributions in fitting the two series. The *p*-value is the approximate Monte Carlo goodness-of-fit test *p*-value (based on Kolmogorov–Smirnov statistics).

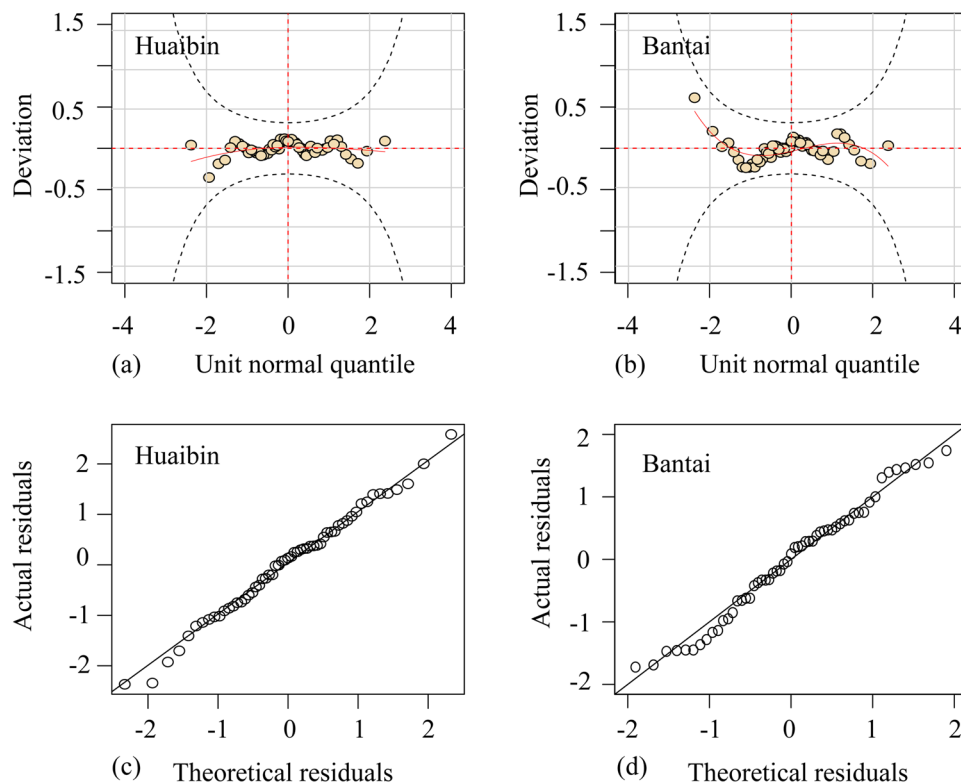
*Estimating the copula parameter.* Before the selection of copula functions, the dependence structures among the flood magnitudes and occurrence dates for the two rivers were explored. The Kendall coefficient and Spearman coefficient were used to perform the independence test<sup>48</sup>, as shown in Table 4. The results indicate that  $Q_h$  and  $Q_b$  are significantly correlated at the 5% significance level. However, the coefficient between the flood magnitude and occurrence date in the same river is close to zero, which suggests that these two variables are independent.

Three candidate copulas with constant parameters were used for modeling the dependence structure of flood occurrence dates ( $T_h$  and  $T_b$ ), including the Gumbel, Frank, and Clayton copulas. Table 5 summarizes the values of parameters and the goodness-of-fit results for these copulas. According to the goodness-of-fit test results (based on Cramér–von Mises statistics), the Clayton copula is not adequate for modeling the bivariate features of the flood occurrence dates (*p*-value is less than 0.05). In terms of the minimum AIC, the Gumbel copula with  $\theta_c = 2.077$  is the best choice.

Considering the nonstationarity of the dependence of flood magnitudes, the time-varying copulas were used as candidates as Eq. (1). The parameters of time-varying copulas were described and estimated as Eqs. (8–9). Table 6 summarizes the values of parameters and the goodness-of-fit results for the optimal copulas under stationary and nonstationary assumptions for modeling the dependence structure of the flood magnitudes ( $Q_h$  and  $Q_b$ ). As Table 6 shows, the P-KS values of  $Z_1$  and  $Z_2$  and the P-Kendall values are larger than 0.05, which supports the validity of the assumed models from a statistical perspective. The time-varying Frank copula was selected to model the dependence structure of flood magnitudes by comparing the AIC values. Figure 6 shows the worm plots for the goodness-of-fit test of the Frank copula with time-varying parameters; notably, all points fall within the 95% confidence interval, indicating satisfactory fitting performance for the selected copula model.

*The risk of flood coincidence.* On the basis of marginal distributions and copula functions, the risk of flood coincidence was analyzed from the three following three perspectives.

*Coincidence probabilities for flood occurrence dates.* With the mixed von Mises distribution and copula function, the joint distribution of flood occurrence dates was determined in order to calculate the daily probability of flood coincidence for flood occurrence dates, as described in Eq. (10). The annual coincidence probability of the flood



**Figure 5.** Fitting plots of residual detection at two stations with the GAMLSS model under the nonstationary assumption (a–b) and linear regression normal QQ diagram (c–d).

Coefficients	$(T_h, T_b)$	$(Q_h, Q_b)$	$(T_h, Q_h)$	$(T_b, Q_b)$
Pearson	0.66	0.43	0.0046	0.0490
Kendall	0.48	0.56	-0.0123	0.0437
Spearman	0.62	0.78	-0.0191	0.0605

**Table 4.** The pairwise correlation coefficients of flood magnitudes and flood occurrence dates.

Copula	Parameters(s.e.)	AIC	p-Value
Gumbel	2.077(0.267)	-37.51	0.382
Clayton	1.421(0.357)	-24.47	0.004
Frank	5.498(0.944)	-28.09	0.141

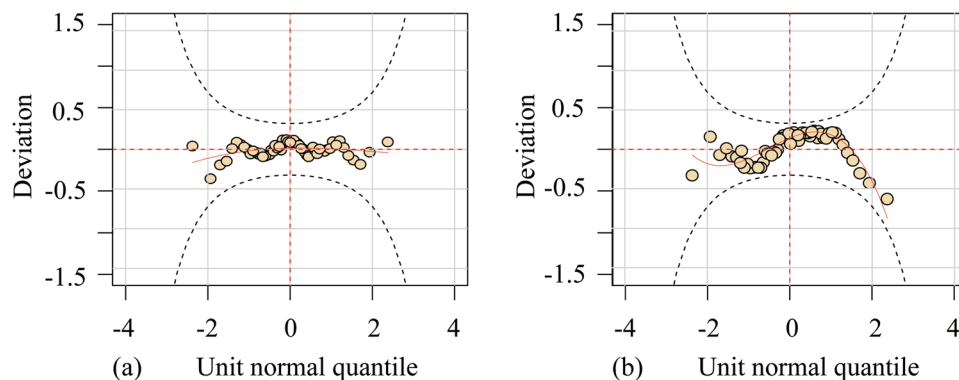
**Table 5.** Parameters and goodness-of-fit test for candidate copulas in modeling the dependence structure of flood occurrence dates. The values in parentheses indicate estimated standard errors; the approximate  $p$ -values (via a multiplier method) of the Cramér–von Mises goodness-of-fit test for copulas are also shown.

Type	Copula	Parameter form	Parameters	AIC	$p$ -KS of $Z_1$	$p$ -KS of $Z_2$	$p$ -Kendall
Stationary	Frank	$\theta_c$	7.039	-42.86	0.73	0.54	0.319
Nonstationary	Frank	$[\omega, \beta, \alpha]$	[1.68, -3.25, 2.90]	-43.98	0.89	0.48	0.332

**Table 6.** Parameters and goodness-of-fit results for the Frank copulas in modeling the dependence structure. The  $p$ -KS ( $Z_1$ ) and  $p$ -KS ( $Z_2$ ) are  $p$ -values of the KS test for the two Rosenblatt’s probabilities integral transformations  $Z_1$  and  $Z_2$ , which should be uniformly and independently distributed on  $[0, 1]$ . The  $p$ -Kendall is the  $p$ -value of the Kendall rank correction test for  $Z_1$  and  $Z_2$ .

occurrence dates  $P_T = 0.152$  was obtained by Eq. (11). According to the observed data, nine floods occurred simultaneously in the two rivers during the 57-year study period. The statistical coincidence probability of the measured data is 0.175, which is close to the model calculation result.





**Figure 6.** Worm plots of the goodness-of-fit for the time-varying Frank copula: (a) worm plot of Rosenblatt's probabilities of integral transformation for  $Z_1$ ; (b) worm plot of Rosenblatt's probabilities of integral transformation for  $Z_2$ .

The daily probability of flood coincidence for flood occurrence dates was plotted as shown in Fig. 4(e). The highest probabilities of flood coincidence occur in mid-July, which suggests a high combination risk. Hence, the Wangjiaba flood diversion sluice needs to discharge the floods to Mengwa for flood control during this period. The coincidence probabilities are close to zero from May to mid-June, which indicates that the flood coincidence is extremely low in the two rivers during this period. Therefore, it is possible to open the Caotaizi escape sluices to discharge the flow in the Mengwa flood detention basin during this period.

**Annual coincidence probabilities for given flood magnitudes.** In this section, the annual coincidence probability for given flood magnitudes is calculated. According to the investigation of the pairwise dependence structures, the occurrence dates and magnitudes of floods are independent of each other. Hence, we employ Eq. (12) to describe the joint distribution of the flood magnitudes. Moreover, the annual flood coincidence probability for given flood magnitudes was calculated by Eq. (13).

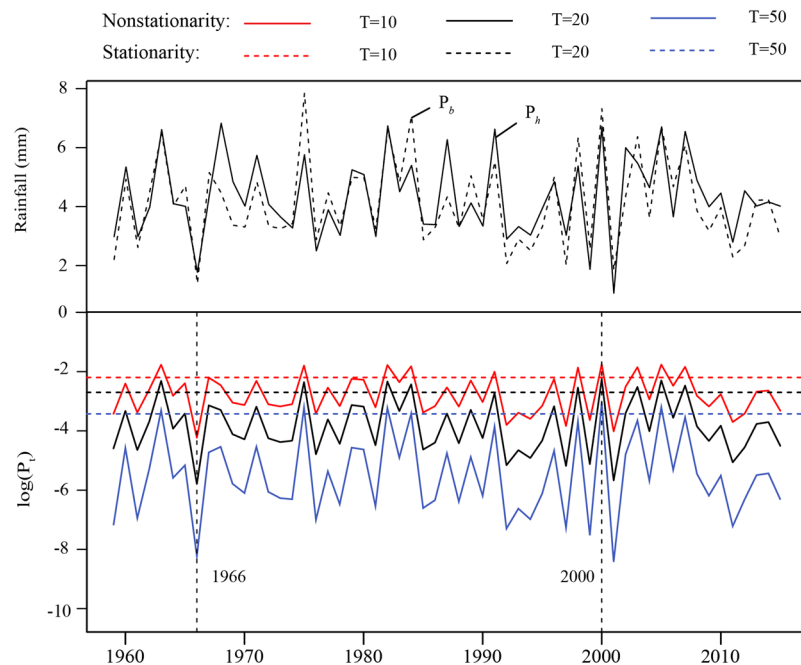
To compare the difference in flood coincidence under stationary and nonstationary conditions, the coincidence probability  $P_T$  in Eq. (13) was calculated under the stationary and nonstationary hypotheses. The values of  $q_a$  and  $q_b$  of the two rivers were the design flows for the  $t$ -year univariate return period under the stationary assumption. Here, we used  $T = (10, 20, 50)$  years as examples, and the annual coincidence probability for a given flood magnitude and the corresponding variations are presented in Fig. 7. For ease of visualization, the vertical coordinate was set to logarithm of the coincidence probabilities.

Our results indicate that the coincidence probability is constant under the stationary condition. However, under the nonstationary condition, the coincidence probabilities for the flood magnitude fluctuate each year over a wide range. The coincidence probabilities for the flood magnitude display variational processes that are consistent with those of rainfall, reflecting the positive correlation between these factors (Fig. 7). In addition, a series of statistical analyses shows that the nonstationary multivariate model performs better than the stationary multivariate model. The probabilities under the stationary assumption may underestimate the risk of flood coincidence in wet years and overestimate this risk in dry years.

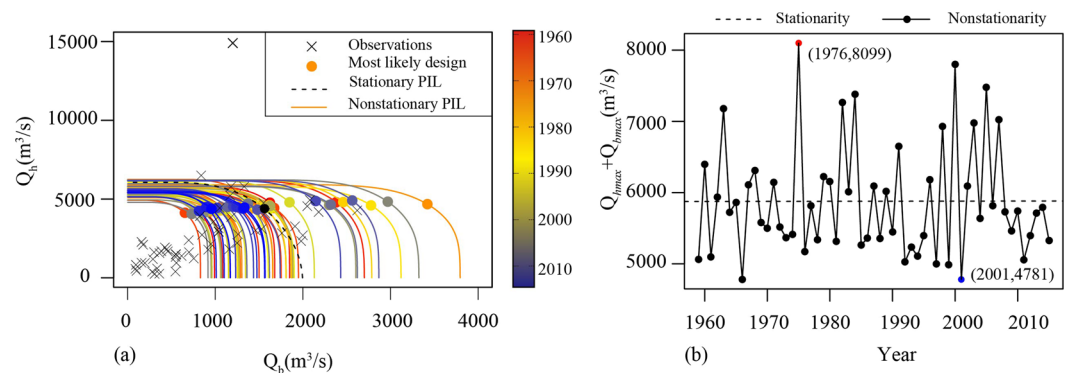
The coincidence probabilities in the case of  $T = 50$  year are close to zero under both the stationary and nonstationary assumptions, which indicates that the coincident flooding is less likely to occur for large flood events than for other events.

**Flood coincidence risk in the "most likely" design.** The "most likely" design in this section provides the most likely scenario for a specific coincidence probability. According to Eq. (13), the joint exceedance probability-isolines (PILs) for each year can be drawn assuming that the coincidence probability  $P_t$  is a definite value. Here, taking  $P_t = 0.01$  as an example, the PILs derived from the stationary model and nonstationary model are shown in Fig. 8(a). The "most likely" design for  $Q_{hmax}$  and  $Q_{bmax}$  in each year considering the definite probability can also be obtained according to Eqs. (14–15), and the corresponding results are denoted by color points in Fig. 8(a).

Under nonstationary conditions, the isoline is constantly oscillating over time while it is a fixed curve under stationary conditions. The "most likely" design of  $Q_{hmax}$  ranges from 4050  $m^3/s$  to 5035  $m^3/s$  and that of  $Q_{bmax}$  ranges from 738  $m^3/s$  to 3625  $m^3/s$ . This difference indicates that  $Q_b$  is more susceptible to environmental change than  $Q_h$ . The combined flows of  $Q_{hmax}$  and  $Q_{bmax}$  in different years are shown in Fig. 8(b) and rang from 4781  $m^3/s$  to 8099  $m^3/s$ , which may lead to flood peaks downstream. Hence, the combined flows calculated based on the "most likely" design have a certain reference significance for flood predictions downstream in Wangjiaba. Therefore, it is necessary to consider the flood coincidence between the Hong River and Huai River and take necessary measures to alleviate the pressure on downstream flood control projects.



**Figure 7.** The coincidence probabilities of the Huai River and Hong River and the corresponding rainfall data.



**Figure 8.** A comparison of the most likely scenario for a specific coincidence probability under the stationary condition and nonstationary condition: (a) the plot of PILs and “most likely” design; (b) the plot of combined flows of  $Q_{hmax}$  and  $Q_{bmax}$ .

### Conclusions

Flood coincidence analysis plays an important role in flood risk analysis. This study proposed a nonstationary multivariate model to solve the flood coincidence problem in a changing environment. The proposed model was developed using the mixed von Mises distribution, the GAMLSS model and a copula function. The main conclusions are presented as follows.

First, the probabilities of flood coincidence on two rivers show that coincident flooding is more likely to occur in mid-July than in other periods (Fig. 4e). Hence, it is possible to discharge these floods to Mengwa in this period to alleviate the downstream flood control pressure.

Second, both the flood magnitudes and their dependent structure are verified to be nonstationary, and the nonstationary multivariate model of flood magnitude performs better than the stationary model. The coincidence probability for large floods (i.e., under the situation of  $T = 50$  year) is nearly equal to zero, which indicates that coincident flood events are more likely to occur for medium-scale or small-scale floods.

Finally, the combined flow under stationary and nonstationary conditions can be obtained from the corresponding “most likely” design ( $Q_{hmax}$  and  $Q_{bmax}$ ), which provides a basis for downstream flood safety. The range of  $Q_{bmax}$  is larger than  $Q_{hmax}$ , which indicates that the nonstationarity of the Hong River is larger than that of the Huai River. Therefore, more attention should be paid to flood control planning in the Hong River Basin.

In conclusion, this study provides a reasonable approach for assessing the risk of flood coincidence in the Huai River Basin under nonstationary conditions. The trends and risks of flood coincidence can be further studied to improve flood management.



## Methods

**Framework of the copula function.** The copula function, which was first proposed by Sklar<sup>8</sup>, describes the correlation between variables. It is actually a class of function that connects joint distributions to their respective marginal distributions. Nelsen<sup>49</sup> and Joe<sup>50</sup> provided a theoretical introduction to copulas. Studies on practical approaches have also been developed<sup>6,51,52</sup>. Specifically, basic guidelines for using copulas in hydrological applications were illustrated by Favre<sup>51</sup>, Salvadori and De Michele<sup>16</sup>, and Salvadori *et al.*<sup>53,54</sup> Useful and free software routines were provided and illustrated by Hofert *et al.*<sup>55</sup>

Environmental changes can influence both the statistical characteristics of hydrological series and the dependence structure of hydrological variables. Considering these changes, a time-varying copula should be considered in such analyses. Hence, the joint distribution of the hydrological variable pair of  $(Y_1^t, Y_2^t)$  at time  $t$  can be produced as follows:

$$H(y_1^t, y_2^t) = C[F_1(y_1^t|\theta_1^t), F_2(y_2^t|\theta_2^t)|\theta_c^t] = C(u^t, v^t|\theta_c^t) \quad (1)$$

where  $H(\cdot)$  represents the joint cumulative distribution function (CDF) of  $Y_1^t$  and  $Y_2^t$ ,  $F_1$  is the cumulative marginal distribution of  $Y_1^t$ ,  $F_2$  is the cumulative marginal distribution of  $Y_2^t$ ,  $C(\cdot)$  is the copula function,  $\theta_1^t$  and  $\theta_2^t$  are the time-varying marginal distribution parameters,  $\theta_c^t$  is the time-varying copula parameter, and the marginal probabilities  $u^t$  and  $v^t$  should obey a uniform distribution in the range of  $[0,1]$ .

Under the stationary assumption, all the parameters above can set as constants:

$$\theta_1^t = \theta_1; \theta_2^t = \theta_2; \theta_c^t = c \quad (2)$$

According to Eq. (1), the implementation of the time-varying copula consists of two steps. The first step is to analyze the univariate nonstationarity (including change point and trend tests) and select an appropriate marginal distribution. The second step is to analyze the nonstationarity of the dependence structure and select an appropriate copula function.

**Marginal distribution.** *Marginal distribution of flood occurrence dates.* Flood occurrence dates can be regarded as a vector with periodic changes. The mixed von Mises distribution<sup>56</sup> is a distribution commonly used to describe periodic or seasonal variables and has been proven to effectively fit flood dates<sup>12</sup>. The directional variable of flood occurrence dates  $x_i$  can be obtained from the following relation:

$$x_i = D_i \frac{2\pi}{L}, \quad 0 \leq x_i \leq 2\pi \quad (3)$$

where  $L$  denotes the length of the flood season and  $D_i$  is the flood occurrence date. The probability density function for a mixture of  $N$  von Mises distribution can be produced in the following form:

$$f(x) = \sum_{i=1}^N \frac{p_i}{2\pi I_0(\kappa_i)} \exp[\kappa_i \cos(x-u_i)] \quad (4)$$

where  $u_i$  and  $\kappa_i$  are location and scale parameters, respectively;  $p_i$  is the weight of the probability density;  $N = 2$  in this study; and  $I_0$  is the Bessel function. In this study, parameters were estimated by the maximum likelihood estimation method, and the approximate Monte Carlo goodness-of-fit test  $p$ -values (based on Kolmogorov–Smirnov statistics) were calculated to assess the validity of the assumption that the flood occurrence dates followed the selected distribution.

*Marginal distribution of flood magnitudes.* In this study, five widely used distributions, including the gamma (GA), Weibull (WEI), lognormal (LOGNO), Gumbel (GU), and generalized extreme value (GEV) distributions, were selected to describe the flood magnitudes (Table S1). Then, based on the GAMLSS model proposed by Rigby and Stasinopoulos<sup>36</sup>, the time-varying marginal distribution was determined; such distributions are widely used in frequency analyses of nonstationary hydrological series.

Taking a three-parameter model with one explanatory variable as an example, if the response variable  $y^t$  follows the distribution function  $F_y = F(y^t|\mu_t, \sigma_t, \nu_t)$  at time  $t$  ( $t = 1, 2, \dots, n$ ), then each parameter can be described as a linear function of the explanatory variable  $x^t$  via a monotonic link function  $g_k(\cdot)$  as follows:

$$g_1(\mu_t) = a_1 + b_1 x^t \quad (5)$$

$$g_2(\sigma_t) = a_2 + b_2 x^t \quad (6)$$

$$g_3(\nu_t) = a_3 + b_3 x^t \quad (7)$$

where  $a_k$  and  $b_k$  ( $k = 1, 2, 3$ ) are the parameters of GAMLSS.

In practice, only the location parameter  $\mu_t$  and scale parameter  $\sigma_t$  are considered to be associated with the explanatory covariate. In other words, either one of these variables is time varying or both are time varying. The shape parameter  $\nu_t$  is treated as a constant (i.e.,  $b_3 = 0$ ). The stationary distribution can be obtained by assuming that the parameters are independent of the explanatory variable (i.e.,  $b_1 = b_2 = b_3 = 0$ ).

The optimal distribution can be selected from the candidates by comparing the value of the Akaike information criterion (AIC). All computational processes were implemented in the R package “gamlss”.

**Copula with time-varying parameters.** In multivariate hydrological frequency analysis, the Archimedean copula is widely used due to its flexibility in structural form<sup>49</sup>. Here, three simple single-parameter copulas, including the Gumbel copula (GH), Clayton copula (CL), and Frank copula (FR), were used as candidates to model the time-varying dependence structure (Table S4).

The time-varying copula function assumes that the copula parameters obey a time-varying process and have time-varying characteristics. In this study, a process similar to ARMA (1,  $q$ )<sup>35</sup> was used to describe the time-varying nature of the parameters:

$$\tau_c^t = \Lambda(\omega + \beta\tau_{t-1} + \alpha\frac{1}{q}\sum_{j=1}^q(u_{t-j} - v_{t-j})) \quad (8)$$

where  $\Lambda(\cdot)$  is a transformation function that considers the correlation with  $\rho_t$  in the definition interval (0, 1) in this study,  $\Lambda(x) = \frac{1}{1 + \exp(-x)}$ , and  $\tau_c^t$  can be converted to  $\theta_c^t$ <sup>58</sup>.

In this paper, the ARMA model with a lag of 10 orders was adopted ( $q = 10$ ). Two-step pseudo maximum likelihood estimation was adopted for parameter estimation. First, the marginal distribution parameters were calculated, and the empirical distribution was then used instead of the marginal distribution to determine the logarithmic likelihood function of the following equation. A two-step method was used to estimate the maximum likelihood.

$$L(\theta_c^t) = \sum_i^T \ln c(\widehat{F}_X(x), \widehat{F}_Y(y); \theta_c^t) L(\theta_c^t) = \sum_i^T \ln c(\widehat{F}_X(x), \widehat{F}_Y(y); \theta_c^t) \quad (9)$$

where  $c(\cdot)$  indicates the density of the copula function and  $\widehat{F}_X(x)$  and  $\widehat{F}_Y(y)$  represent the empirical marginal distributions. The parameters can be obtained by combining Eqs. (1) and (9). Because the test method of calculating the “distance” between the fitted copula and the empirical copula is not suitable for time-varying copulas<sup>59</sup>, the Rosenblatt probability integral transformation was used to test the goodness-of-fit of the time-varying copulas<sup>60</sup>. The optimum copula was selected according to the minimum AIC.

**Analysis of flood coincidence risk.** The term coincidence refers to the simultaneous occurrence of floods in two (or more) rivers, which can be measured by the exceedance probabilities of flood events. As flood events are characterized by flood occurrence dates and magnitudes, both of these factors should be considered.

First, the flood occurrence dates were defined as the occurrence dates of the AMDF in the Huai River or Hong River. Then, the daily coincidence probabilities for the flood occurrence dates were described by the following mathematical equation:

$$P_i = P(t_i \leq T_h \leq t_{i+1}, t_i + \tau_{hb} - \Delta t \leq T_b \leq t_i + \tau_{hb} + \Delta t) \quad (10)$$

$$P_T = \sum_{i=1}^n P_i \quad (11)$$

where  $i$  indicates the  $i^{\text{th}}$  day of the flood season and  $\Delta t$  indicates the time interval between the response in the two rivers. Because the maximum daily flow in the two rivers rarely occurs on the same day, we defined  $\Delta t = 1$  as a flood coincidence event in this study, and  $\tau_{hb}$  indicates the difference in propagation time between the two stations (the difference in propagation time between the Huaibin and Bantai stations is 16 hours). Additionally,  $n$  indicates the length of the flood season, and  $P_T$  is the annual coincidence probability of flood occurrence dates.

Second, the series of AMDFs was selected to fit the joint distribution of flood magnitudes, assuming that the random variables of AMDF in Huaibin  $Q_h$  obey the  $F_H$  distribution and the random variables of AMDF in Bantai  $Q_b$  obey the  $F_B$  distribution. Then, the exceedance probability of flood coincidence for a given flood magnitude in the  $t^{\text{th}}$  year can be defined as:

$$P_Q^t = P^t(Q_h \geq q_h, Q_b \geq q_b) = 1 - F_H(q_h|\theta_1^t) - F_B(q_b|\theta_2^t) + C(F_H(q_h|\theta_1^t), F_B(q_b|\theta_2^t)|\theta_c^t) \quad (12)$$

where  $q_h$  and  $q_b$  are the designed flows,  $C(\cdot)$  represents the copula function with time-varying parameters,  $P_Q^t$  is the exceedance probabilities for coincident flood magnitudes, and  $Q_h$  and  $Q_b$  are the flood magnitudes, namely, the AMDFs sampled by the AM method.

Then, based on the hypothesis that the flood occurrence dates and flood magnitudes are independent of each other (i.e., the Kendall coefficient is less than 0.05), the annual flood coincidence probabilities  $P_i$  for given flood magnitudes in the  $t^{\text{th}}$  year can be stated as AND-joint exceedance probabilities:

$$P_i = \sum_{i=1}^n P_i P_Q^t = P_T P^t(Q_h \geq q_h, Q_b \geq q_b) \quad (13)$$

This equation describes the probability that  $Q_h$  is greater than  $q_h$  and  $Q_b$  is greater than  $q_b$  when the occurrence dates of these two flows differ by one day.

Finally, the impacts of flood coincidence on the lower reaches of the basin were analyzed using the “most likely” design<sup>38</sup>. When  $P_i$  is fixed as a constant in the range of (0,1), there are many possible combinations of ( $Q_h$ ,  $Q_b$ ), where  $P_Q^t$  is also a constant because of the constant nature of  $P_T$ . Among these combinations, the “most likely” design reflects the combination with the highest probability, as defined in Eqs. (14–15).

$$(\mu_m^t, \nu_m^t) = \operatorname{argmax} f_c(F_H^{-1}(\mu|\theta_1^t), F_B^{-1}(\nu|\theta_2^t)|\theta_c^t) \quad (14)$$

where  $F_H^{-1}(\cdot)$  and  $F_B^{-1}(\cdot)$  are the inverse functions of the marginal distributions of  $Q_h$  and  $Q_b$  with time-varying parameters, respectively;  $f_c$  is the joint probability density with time-varying parameter  $\theta_c^t$ ; and  $\mu$  and  $\nu$  are independent of each other and subject to a uniform distribution in the range of  $[0, 1]$ . Then, the “most likely” design in the  $t^{\text{th}}$  year can be calculated by using the inverse of the CDFs of the marginal distributions:

$$Q_{hmax}^t = F_H^{-1}(\mu_m^t|\theta_1^t) \text{ and } Q_{bmax}^t = F_B^{-1}(\nu_m^t|\theta_2^t) \quad (15)$$

Received: 15 May 2019; Accepted: 10 February 2020;

Published online: 25 February 2020

## References

- Chen, L., Singh, V. P. & Guo, S. Flood coincidence risk analysis using multivariate copula functions. *Journal of Hydrologic Engineering* **17**, 742–755 (2012).
- Prohaska, S., Ilic, A. & Majkic, B. Multiple-coincidence of flood waves on the main river and its tributaries. *IOP Conf. Series: Earth and Environmental Scienc* (2008).
- Requena, A. I., Chebana, F. & Mediero, L. A complete procedure for multivariate index-flood model application. *Journal of Hydrology*, S0022169416300294.
- Mediero, L., Jiménez-Álvarez, A. & Garrote, L. Design flood hydrographs from the relationship between flood peak and volume. *Hydrology and Earth System Sciences* **14**, 2495–2505.
- Xie, H. & Tie-Guang, L. I. Effects of Dependence Structure and Marginal Distribution on Hydrological Multivariate Probability Distribution. *Journal of Irrigation & Drainage* (2011).
- Genest, C. & Favre, A. C. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering* **12**, 347–368 (2003).
- Serinaldi, F. & Grimaldi, S. Fully Nested 3-Copula: Procedure and Application on Hydrological Data. *Journal of Hydrologic Engineering* **12**, 420–430.
- Sklar, M. Fonctions de repartition a n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris*, 229–231 (1959).
- De Michele, C. & Salvadori, G. A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-Copulas. *Journal of Geophysical Research* **108** (2003).
- Zhang, L. & Singh, V. P. Bivariate rainfall frequency distributions using Archimedean copulas. *Journal of Hydrology* **332**, 93–109 (2007).
- Grimaldi, S. & Serinaldi, F. Asymmetric copula in multivariate flood frequency analysis. *Advances in Water Resources* **29**, 1155–1167 (2006).
- Chen, L., Guo, S. L., Yan, B. W., Pan, L. & Fang, B. A new seasonal design flood method based on bivariate joint distribution of flood magnitude and date of occurrence. *International Association of Scientific Hydrology Bulletin* **55**, 1264–1280 (2010).
- Reddy, M. J. & Ganguli, P. Bivariate flood frequency analysis of upper Godavari River Flows using archimedean copulas. *Water Resources Management* **26**, 3995–4018 (2012).
- Requena, A. I., Mediero, L. & Garrote, L. A bivariate return period based on copulas for hydrologic dam design: Accounting for reservoir routing in risk estimation. *Hydrology and Earth System Sciences*, **17**,8(2013-08-01) 17, 3023–3038 (2013).
- Šraj, M., Bezak, N. & Brilly, M. Bivariate flood frequency analysis using the copula function: a case study of the Litija station on the Sava River. *Hydrological Processes* **29**, 225–238 (2015).
- Salvadori, G. & De Michele, C. Frequency analysis via copula: theoretical aspects and applications to hydrological events. *Water Resources Research* **40**, 229–244 (2004).
- Segers, J. J. J. Non-parametric inference for bivariate extreme-value copulas. *Ssrn Electronic Journal* **142**, 181–203 (2008).
- Bing, J. *et al.* Flood coincidence analysis of Poyang Lake and Yangtze River: risk and influencing factors. *Stochastic Environmental Research & Risk Assessment* **32**, 879–891 (2018).
- Li, Z., Feng, P. & Yuan, X. Coincidence risk analysis for non-stationary flood peak of Yellow River and its tributaries in Ningxia Hui Autonomous Region. *Advances in Science & Technology of Water Resources* (2016).
- Chen, L., Guo, S. L., Zhang, H. G., Yan, B. W. & Liu, X. Y. Flood coincidence probability analysis for the upstream Yangtze River and its tributaries. *Advances in Water Science* **22**, 323–330 (2011).
- Zhao, X. H. & Xu, C. Auto Regressive and Ensemble Empirical Mode Decomposition Hybrid Model for Annual Runoff Forecasting. *Water Resources Management* **29**, 2913–2926 (2015).
- Zhang, Y., Zhai, X., Shao, Q. & Yan, Z. Assessing temporal and spatial alterations of flow regimes in the regulated Huai River Basin, China. *Journal of Hydrology* **529**, 384–397 (2015).
- Requena, A. I., Flores, I., Mediero, L. & Garrote, L. Extension of observed flood series by combining a distributed hydro-meteorological model and a copula-based model. *Stochastic Environmental Research and Risk Assessment* **30**, 1363–1378 (2016).
- Singh, V. P. & Zhang, L. Copula–entropy theory for multivariate stochastic modeling in water engineering. *Geoscience Letters* **5**, 6, <https://doi.org/10.1186/s40562-018-0105-z> (2018).
- Ali, H. & Mishra, V. Contrasting response of rainfall extremes to increase in surface air and dewpoint temperatures at urban locations in India. *Scientific Reports* **7**, 1228 (2017).
- Gilroy, K. L. & Mccuen, R. H. A nonstationary flood frequency analysis method to adjust for future climate change and urbanization. *Journal of Hydrology* **414**, 40–48 (2012).
- Sun, P. *et al.* Nonstationary evaluation of flood frequency and flood risk in the Huai River basin, China. *Journal of Hydrology* (2018).
- Galiatsatou, P., Anagnostopoulou, C. & Prinos, P. Modeling nonstationary extreme wave heights in present and future climates of Greek Seas. *Water Science & Engineering* **9**, 21–32 (2016).
- Liang, Z. M., Hu, Y. M. & Wang, J. Advances in hydrological frequency analysis of non-stationary time series. *Adv. Water Sci.* **22**, 864–871 (2011).
- Manner, H. & Reznikova, O. A Survey on Time-Varying Copulas: Specification, Simulations, and Application. *Econometric Reviews* **31**, 654–687 (2011).
- Almeida, C. & Czado, C. Efficient bayesian inference for stochastic time-varying copula models. *Computational Statistics & Data Analysis* **56**, 1511–1527 (2012).
- Jiang, C., Xiong, L., Xu, C. Y. & Guo, S. Bivariate frequency analysis of nonstationary low-flow series based on the time-varying copula. *Hydrological Processes* **29**, 1521–1534 (2015).
- Chebana, F., Ouarda, T. B. M. J. & Duong, T. C. Testing for multivariate trends in hydrologic frequency analysis. *Journal of Hydrology* **486**, 519–530 (2013).

34. Sarhadi, A., Burn, D. H., Concepción Ausín, M. & Wiper, M. P. Time-varying nonstationary multivariate risk analysis using a dynamic Bayesian copula. *Water Resources Research* **52**, 2327–2349, <https://doi.org/10.1002/2015wr018525> (2016).
35. Patton, A. J. Modelling Time-Varying Exchange Rate Dependence using the Conditional Copula. *Social Science Electronic Publishing* (2001).
36. Rigby, R. A. & Stasinopoulos, D. M. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society* **23**, 507–554 (2007).
37. Stasinopoulos, D. M. & Rigby, R. A. Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. *Journal of Statistical Software* **23**, 46 (2007).
38. Salvadori, G., De Michele, C. & Durante, F. On the return period and design in a multivariate framework. *Hydrol. Earth Syst. Sci.* **15**, 3293–3305 (2011).
39. Xia, J. *et al.* Temporal and spatial variations and statistical models of extreme runoff in Huaihe River Basin during 1956–2010. *Journal of Geographical Sciences* **22**, 1045–1060 (2012).
40. Wang, L., Wang, W. & Zhang, J. Analysis on spatial and temporal distribution characteristics of precipitation processes over main river basin in China. *Journal of Natural Disasters* **27**, 161–173 (2018).
41. Yuan, Z., Yang, Z., Zheng, X. & Yuan, Y. Spatial and temporal variations of precipitation in Huaihe River Basin in recent 50 years. *South-to-North Water Diversion and Water Science & Technology* **10**, 98–103 (2012).
42. Yamada, I. *Thiessen Polygons*. (American Cancer Society, 2016).
43. Faisal, N. & Gaffar, A. Development of Pakistan's new area weighted rainfall using Thiessen polygon method. (2012).
44. Holmes, M., I. K., J-F. Quessy Nonparametric tests for change-point detection à la Gombay and Horváth. *Journal of Multivariate Analysis*, **115**, 116–132 (2013).
45. Bücher, A., Kojadinovic, I., Rohmer, T. & Segers, J. Detecting changes in cross-sectional dependence in multivariate time series. *Journal of Multivariate Analysis* **132**, 111–128 (2014).
46. Bücher, A. & Kojadinovic, I. A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing. **22** (2013).
47. npcp: Some Nonparametric CUSUM Tests for Change-Point Detection in Possibly Multivariate Observations; R Package Version 0.1-9.; R Package: Vienna, Austria, 2017.
48. Deheuvels, P. *Non parametric tests of independence*. (Springer Berlin Heidelberg, 1970).
49. Nelsen, R. B. *An Introduction to Copulas*. (Springer, 2006).
50. Joe, H. *Dependence Modeling with Copulas*. (CRC Press, 2014).
51. Favre, A. C., Adlouni, S. E., Perrault, L., Thiémondge, N. & Bobée, B. Multivariate hydrological frequency using copulas. *Water Resources Research* **40** (2004).
52. Salvadori, G., De Michele, C., Kottegoda, N. T. & Rosso, R. *Extremes In Nature: An Approach Using Copulas*. (Springer Science & Business Media, 2007).
53. Salvadori, G., Tomasichio, G. R. & D'Alessandro, F. Practical guidelines for multivariate analysis and design in coastal and off-shore engineering. *Coastal Engineering* **88**, 1–14 (2014).
54. Salvadori, G., Durante, F., De Michele, C., Bernardi, M. & Petrella, L. A multivariate copula-based framework for dealing with Hazard Scenarios and Failure Probabilities. *Water Resources Research* **52**, 3701–3721 (2016).
55. Hofert, M., Kojadinovic, I., Mächler, M. & Yan, J. *Elements of Copula Modeling With R*. (Springer, 2018).
56. Jammalamadaka, S. R. & Mangalam, V. A general censoring scheme for circular data. *Statistical Methodology* **6**, 280–289 (2009).
57. Kendall, M. *Multivariate analysis*. (Academic Press, 1975).
58. Stoeber, J. & Brechmann, E. Parameter of a bivariate copula for a given Kendall's tau value.
59. Patton, A. J. A review of copula models for economic time series. *Journal of Multivariate Analysis* **110**, 4–18 (2012).
60. Genest, C., Rémillard, B. & Beaudoin, D. Goodness-of-fit tests for copulas: A review and a power study. *Insurance Mathematics & Economics* **44**, 199–213 (2009).

## Acknowledgements

The study is financially supported by the National Key Research and Development Program of China (2017YFC0405601), the National Natural Science Foundation of China (41730750), the UK-China Critical Zone Observatory (CZO) Program (41571130071).

## Author contributions

Y.F., P.S. and S.Q. proposed the methods and developed the model; P.S. and Y.F. did computation and simulation work; S.M., C.C. and F.D. analyzed the results; Y.F., P.S. and S.Q. wrote the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-60264-3>.

**Correspondence** and requests for materials should be addressed to S.Q.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020