Response of flood events to extreme precipitation: two case studies in Taihu Basin, China
Feiqing Jiang, Zengchuan Dong, Yun Luo, Moyang Liu, Tao Zhou, Xinkui Wang and Zhenye Zhu

ABSTRACT

Flood events are typically triggered by extreme precipitation in rain-dominant basins. In this study, to better understand the genetic mechanisms and characteristics of floods, copula functions are used to analyze the response of flood events to extreme precipitation. The coincidence probabilities of the typical extreme flood and precipitation events are calculated; different return periods of their arbitrary combinations are calculated, whereas the dangerous domains for flood control under different return periods are identified; furthermore, flood risk analysis under different extreme precipitation scenarios is performed via their conditional exceedance probabilities. The Xitiaoxi catchment (XC) and Dongtiaoxi catchment (DC) in the Zhexi Region of the Taihu Basin are selected as the study area. The results show that in four scenarios with precipitation frequencies of 80%, 90%, 93.33%, and 95%, the probabilities of the dangerous flood are 9.72%, 10.57%, 10.86%, and 11.01% in the XC, respectively, and 5.91%, 6.31%, 6.44%, and 6.51% in the DC, respectively. This study provides a practical basis and guidance for the computation of rainstorm designs, management of flood control safety, and water resource scheduling in the Taihu Basin.

Key words | bivariate frequency, copula, extreme precipitation, flood risk analysis, return periods, Taihu Basin

HIGHLIGHTS

- This study focused on the flood risk in the main water source of Taihu Lake, i.e., the Zhexi Region, which had a direct and significant impact on the floods in the Taihu Basin.
- Extreme hydrological and meteorological events were considered as the objects of this study. Their joint probability distribution and combined effects were directly related to the degree of flood risk.
INTRODUCTION

Floods occur frequently and are widely distributed, resulting in significant casualties and economic losses worldwide, particularly in plain river network areas with developed economies and dense populations. (UNISDR 2016; Zeng et al. 2019; Thapa et al. 2020). The Taihu Basin is a typical plain river network area in China, located in the economically developed Yangtze River Delta. Although the area and population of the Taihu Basin account for only 0.4 and 3% of the country, the Gross Domestic Product (GDP) is as high as 13% of that of the country (Bao et al. 2016). Characterized by a flat terrain, extensive river networks, and numerous polder areas, the Taihu Basin is vulnerable to flooding (Zhai et al. 2020). In 2016, direct economic losses due to flood disasters in the Taihu Basin were $1.06 billion, accounting for 0.12% of its GDP (Wang 2018). Meanwhile, in 1991 and 1999, the direct economic losses due to flood disasters accounted for 6.7% and 1.6% of the GDP in the corresponding year (Xu 2009; Wang et al. 2018). Within this context, satisfying national and municipal demands for a better understanding of flood risks in the Taihu Basin is of top priority for scientific and effective disaster prevention and mitigation work (Tanaka et al. 2017).

Precipitation and discharge are closely related, and floods are typically triggered by extreme precipitation events in rain-dominant basins (Fang et al. 2018). The flood events in the Taihu Basin have occurred primarily during the flood season (from May 1st to September 30th) and have always been accompanied by extreme precipitation events (Li et al. 2013). Therefore, investigating the bivariate frequency of extreme flood and extreme precipitation during the flood season will facilitate a better understanding of the incidence characteristics and genetic mechanisms of flood events in the Taihu Basin.

The copula function, introduced by Sklar (Sklar 1959), is a commonly used tool for calculating the joint distribution of multiple random variables. It can flexibly construct the multivariate joint distribution of random variables with different marginal distributions and describe linear and non-linear correlations between them (Favre et al. 2004). The application of copula in hydrology can be categorized into two primary types. The first is to analyze the frequency of the multifeature attributes of a certain hydrological phenomenon, such as the drought duration and intensity (Kwon & Lall 2016); flood duration, peak, and volume (Daneshkhah et al. 2016); and precipitation duration, intensity, and depth (Gao et al. 2018). The other is to study the combination of hydrological events at different time and space scales, such as the flood coincidence probability analysis for the mainstream and its tributaries (Gao et al. 2018), as well as river flow simulations based on spatio-temporal joint distributions (Chen et al. 2019). A more comprehensive understanding of hydrological phenomena can be obtained based on multivariate joint frequencies.

Thus far, many scholars have used copula functions to evaluate the multivariate frequencies of flood or precipitation events. Chen et al. analyzed the risk of flooding as a result of flood coincidences by considering flood magnitudes and times (dates) of occurrence using multivariate copula functions (Chen et al. 2012). Zhang et al. investigated the probabilistic behaviors of precipitation extremes based on copulas in both space and time (Zhang et al. 2013). Saad et al. analyzed spring flood-causing mechanisms in terms of the occurrence, frequency, duration, and intensity of precipitation as well as temperature events and their combinations prior to and during floods using a proposed multivariate copula approach (Saad et al. 2015). Goswami et al. analyzed the copula-based probabilistic behavior of precipitation extremes over the eastern Himalayan region and discovered that the co-occurrences of heavy and weak precipitation will be more frequent in the future, resulting in a higher risk of floods and droughts within the same year (Goswami et al. 2018). However, the studies above have focused on the multivariate frequency analysis of different feature attributes of the same event (i.e., either the flood or precipitation event), whereas the multivariate frequency analysis of the extreme flood and extreme precipitation events has not been performed.

Frequency analysis of meteorological or hydrological events in the Taihu Basin has been performed. Hu et al.
constructed the joint distribution of the initial time of typhoons, the start time of plum rains, and the end time of plum rains in the Taihu Basin based on Gumbel copulas and concluded that the total encounter possibility of typhoons and plum rains in the Taihu Lake Basin was 9.23% (Hu et al. 2010). Liu et al. investigated the relationships between the onset of plum rains and the rainfall amount during the plum rains period in the Taihu Basin based on the copula function; the results showed that changes in floods appearing during the early onset of plum rains must be the focus (Liu et al. 2013). Han et al. applied bivariate copulas to analyze the correlation between flood peak levels and duration at two stations in the Taihu Basin and obtained the joint return periods of their arbitrary combinations (Han et al. 2018). Luo et al. calculated the joint distribution of the precipitation events of eight hydrological subregions in the Taihu Basin using Archimedean copulas considering the spatial and temporal differences in flood risks caused by plum rains and typhoons (Luo et al. 2019). However, these studies focused on the different feature attributes of a single meteorological or hydrological event in the Taihu Basin, whereas the joint distribution analysis of extreme meteorological and hydrological events has not yet been investigated. Meanwhile, the aforementioned studies pertaining to the Taihu Basin considers the average situation of the entire Taihu Basin as the study object; however, a targeted research regarding the Zhexi Region, i.e., the main water source of the Taihu Lake, has not been conducted.

Therefore, the main purpose of this study is to investigate the response of flood events to extreme precipitation in the Zhexi Region of the Taihu Basin through the joint risk analysis of extreme flood and extreme precipitation events. The main contents of this study include: (1) establishment of a copula-based joint distribution model of the extreme flood and precipitation events for flood risk analysis, (2) calculation of the coincidence probabilities of typical extreme flood and precipitation events, (3) analysis of different return periods of arbitrary combinations of extreme flood and precipitation events and the identification of dangerous domains for flood control under different return periods, and (4) flood risk analysis under different extreme precipitation scenarios via their conditional exceedance probabilities. The technical roadmap of this study is shown in Figure 1.

**MATERIALS AND METHODS**

**Materials**

**Study area**

The Xitiaoxi catchment (XC) and Dongtiaoxi catchment (DC) are located in one of the eight hydrological subregions
of the Taihu Basin, i.e., the Zhexi Region. The upstream area of the Hengtangcun hydrometrical station in the XC and the upstream area of the Pingyao hydrometrical station in the DC were selected as the study area (Figure 2), whose drainage areas measured 1,359 km² and 1,420 km², respectively. The Zhexi Region is located upstream of the Taihu Lake and is the subregion of the Taihu Basin with the highest rainfall. The inflow of the Taihu Lake from the Zhexi Region accounts for approximately 50% of the total inflow of this lake (Wang et al. 2019). The steep slopes and fast flowing rivers resulted in a high risk of floods in the XC and DC, which are the main source of floods in the Taihu Basin.

The XC (119°14′E–19°3′E, 30°23′N–30°53′N) lies in the subtropical monsoon climate zone, which is mild and humid. The multiyear average temperature and precipitation are 15.5 °C and 1,465.8 mm, respectively. The distribution of the precipitation is uneven during the year, with the most precipitation occurring in the flood season, accounting for 75% of the annual precipitation. The runoff in the XC is generated by the precipitation, and its occurrence and quantity are dependent on the characteristics of the precipitation. Therefore, the within-year variation in the runoff is significant, and the runoff during the flood season accounts for 45–54% of the annual runoff, making the XC prone to floods. The average annual inflow of the Taihu Lake from the XC is $26.8 \times 10^8$ m³, accounting for 27.7% of the total inflow of this lake (Liu et al. 2009). Since the mid-20th century, extreme floods have occurred frequently in the XC. For example, as a result of a typhoon, the maximum observed discharge at the Hengtangcun station was 1,770 m³/s in 2013. High mountainous and hilly areas are distributed in the southwest with a maximum elevation of 1,578 m, whereas low alluvial plains lie in the northeast (Lv et al. 2015). The dominant land use types are woodland and cropland.

The DC (119°30′E–120°06′E, 30°08′N–30°42′N) has a subtropical monsoon climate, with an annual average temperature and precipitation of 15.8 °C and 1,640 mm,
respectively. The precipitation varies significantly by year, and its distribution is uneven during the year. May to July is the flood season characterized by plum rains, with the precipitation ranging from 450 to 510 mm; August to September is the flood season characterized by typhoons, with the precipitation ranging from 190 to 800 mm. Furthermore, the precipitation in the flood season accounts for approximately 75% of the annual precipitation. Similar to the runoff in the XC, the runoff in the DC is generated by precipitation, and the distribution of the runoff during the year is consistent with the precipitation. The average annual inflow of Taihu Lake from the DC is $15.4 \times 10^8$ m$^3$, accounting for 18.4% of its total inflow (Ma & Wang 2011). The maximum flood discharge at Pingyao station was 795 m$^3$/s, which occurred in 1984. The geomorphology from northwest to east varies from mountainous to hilly and ends in a plain area, and the corresponding height ranges from 1,500 to 3 m. The major land use types are woodland and cropland, and the DC has more urban and construction land compared with the XC.

**Dataset**

The data used in this study included:

1. Daily precipitation data during the flood season (from 1965 to 2016) of 14 representative stations in the XC, including the Hengtangcun station, and 12 representative stations in the DC, including the Pingyao station.
2. Daily discharge data during the flood season from 1965 to 2016 of the Hengtangcun station in the XC and the Pingyao station in the DC.

The locations of each station are shown in Figure 2. The areal average precipitations in the two catchments were calculated using the Thiessen polygon method. All data were obtained from the ‘Annual Hydrological Report P.R. China’ (Lin et al. 1965-2016), which is a hydrological yearbook compiled by the Ministry of Water Resources of the People’s Republic of China. The data in the hydrological yearbook have been reviewed by the authors to guarantee the reliability, including the completeness, accuracy, and consistency. Therefore, the data used here are reliable.

**Determination of marginal distribution**

The marginal distributions of the four series were determined: (1) the maximum flood discharge during the flood season (MF) at the Hengtangcun station, (2) the areal average precipitation during the flood season (Pr) in the XC, (3) the MF at the Pingyao Station, and (4) the Pr in the DC.

**Autocorrelation test**

Autocorrelation refers to the degree of correlation of the same variable between two successive time intervals. Autocorrelation must be tested when analyzing a set of historical data to evaluate its randomness. The autocorrelation coefficient measures the relation between the lagged and original versions of the value in a time series and is expressed as

$$
\gamma_h = \frac{\sum_{t=1}^{n-h} (x_t - \bar{x})(x_{t+h} - \bar{x})}{\sum_{t=1}^{n} (x_t - \bar{x})^2}
$$

where $\{x_t\}$ is the time series, $n$ the sample size, $\bar{x}$ the mean of the time series, and $h$ the lag.

In the autocorrelation plot, the upper and lower bounds for autocorrelation with significance level $\alpha$ are obtained as follows:

$$
B = \pm z_{n_a/2}SE(\gamma_h)
$$

where $z$ is the cumulative distribution function (CDF) of the standard normal distribution, $\alpha$ the significance level, $SE$ the standard error, and $\gamma_h$ the autocorrelation coefficient. If the autocorrelation is higher (lower) than this upper (lower) bound, the null hypothesis that no autocorrelation exists at and beyond a specified lag is rejected at a significance level of $\alpha$.

**Parameter estimation**

The generalized extreme value (GEV), Gumbel, Pearson type III (P-III), gamma, normal, and log-normal distributions, which are widely used in the distribution of extreme hydrological events, were selected to fit the marginal distribution of the four series. We applied the maximum likelihood (ML) estimator to estimate the parameters, which were both intuitive and flexible. The
log-likelihood function is
\[ L(\theta) = L(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \ldots, \theta_k) = \frac{1}{n} \sum_{i=1}^{n} \ln[f(x_i; \theta_1, \theta_2, \ldots, \theta_k)] \]
(3)
where \( x_1, x_2, \ldots, x_n \) are a number of observations; \( \theta_1, \theta_2, \ldots, \theta_k \) are parameters of the marginal distribution; and \( f(\cdot) \) is the marginal distribution probability distribution function (PDF). The function \( L(\theta) \) is maximized over the parameter space \( \Theta \), which corresponds to solving
\[ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \ln[f(x_i; \theta_1, \theta_2, \ldots, \theta_k)]}{\partial \theta_j} = 0, \quad j = 1, 2, \ldots, k \]
(4)

### Goodness-of-fit test

The Kolmogorov–Smirnov (K–S) goodness-of-fit test was used to evaluate the performance of the candidate marginal distributions, whose statistic is expressed as
\[ D_n = \sup_{x \in \mathcal{X}} |F(x) - F_e(x)| \]
(5)
where \( F(x) \) is the estimated CDF; \( F_e(x) \) is the empirical CDF, which is calculated using the Gringorten plotting position formula expressed as
\[ F_e(x) = \frac{k - 0.44}{n + 0.12} \]
(6)
where \( n \) is the sample size, and \( k \) is the ranking of the data set in the increasing order. The \( p \)-value for the K–S test was estimated using Miller’s approximation (Miller 1956).

Moreover, the mean absolute error (MAE), probability plot correlation coefficient (PPCC), and deterministic coefficient (DMC) were used to further select the most suitable distribution. These methods are expressed as follows:
\[ \text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |a_i - b_i| \]
(7)
\[ \text{PPCC} = \frac{\sum_{i=1}^{n} (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^{n} (a_i - \bar{a})^2 \sum_{i=1}^{n} (b_i - \bar{b})^2}} \]
(8)
\[ \text{DMC} = 1 - \frac{\sum_{i=1}^{n} (a_i - b_i)^2}{\sum_{i=1}^{n} (a_i - \bar{a})^2} \]
(9)

where \( a_i \) and \( b_i \) are the empirical CDF and estimated CDF, respectively; \( a \) and \( b \) are the means of the empirical CDF and estimated CDF, respectively; and \( n \) is the number of samples. The smaller the value of MAE, the better is the fitting performance. The larger the values of the PCC and DMC, the better is the fitting performance.

### Establishment of copula model

The copula model used to describe the dependence structure of the MF and Pr series in the study areas was constructed by joining their marginal distributions.

### Dependence evaluation

The Spearman’s rank correlation coefficient and Kendall rank correlation coefficient were used to measure the dependence between the MF and Pr series; they were computed as
\[ r_s = \rho_{rgX,rgY} = \frac{\text{cov}(rgX, rgY)}{\sigma_{rgX}\sigma_{rgY}} \]
\[ \tau = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{n(n-1)/2} \]
(10)
(11)
where \( r_s \) is the Spearman’s rank correlation coefficient; \( \rho \) is the Pearson correlation coefficient of the rank variables; \( rgX \) and \( rgY \) are ranks; \( \text{cov}(rgX, rgY) \) is the covariance of the rank variables; \( \sigma_{rgX} \) and \( \sigma_{rgY} \) are the standard deviations of the rank variables; \( \tau \) is the Kendall rank correlation coefficient; and \( n \) is the sample size.

### Copula function and parameter estimation

A copula function is a multivariate CDF function that can solve the marginal distribution of each variable and connection function separately. For \( d \)-dimensional random variables \( U_1, U_2, \ldots, U_d \) with marginal CDFs \( F_1, F_2, \ldots, F_d \), a \( d \)-dimensional copula \( C \) is defined as
\[ H(u_1, u_2, \ldots, u_d) = C[F_1(u_1), F_2(u_2), \ldots, F_d(u_d); \theta], \]
(12)
where \( H \) is the multivariate CDF; \( F_i(u_i) = F_i(U_i \leq u_i) \) is the marginal CDF of \( U_i; \theta \) is the copula parameter vector.
The density \( h(\cdot) \) of the multivariate distribution is expressed as

\[
h(u_1, u_2, \ldots, u_d) = c[F_1(u_1), F_2(u_2), \ldots, F_d(u_d); \Theta] \prod_{i=1}^{d} f_i(u_i)
\]

where \( f_i(u_i) = f_i(U_i \leq u_i) \) is the marginal PDF of \( U_i \); \( c \) is the copula PDF, expressed as

\[
c(u_1, u_2, \ldots, u_d) = \frac{\partial C(u_1, u_2, \ldots, u_d)}{\partial u_1 \partial u_2 \cdots \partial u_d}
\]

The copula is unique if the marginals \( F_i(u_i) \) are continuous.

The typically applied Archimedean copulas in hydrology, including four single-parameter copulas and four double-parameter copulas whose structures are more flexible, were employed in this study. They are defined in Table 1.

We used the inference function for margins (IFM) estimator to estimate the parameters of the copula model, which addressed the computational inefficiency of the ML estimator by performing the estimation in two steps. The first was the estimation of the marginal parameters, as described in the section on Determination of marginal distribution. Subsequently, the estimated marginal parameters were substituted into the log-likelihood function to obtain the estimated copula parameter. The log-likelihood function of the joint distribution is

\[
L_c(\Theta_c) = \sum_{i=1}^{n} \ln c[F_1(u_{1i}; \Theta_1), F_2(u_{2i}; \Theta_2), \ldots, F_d(u_{di}; \Theta_d); \Theta_c]
\]

where \( n \) is the sample size; \( \Theta_c \) is the parameter vector of the copula; and \( \Theta_1, \Theta_2, \ldots, \Theta_d \) are estimated parameter vectors of the marginal distribution.

The function \( L_c \) is maximized over the parameter space \( \Theta_c \) to obtain the estimated copula parameter \( \hat{\Theta}_c \), which corresponds to solving

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{\partial \ln c[F_1(u_{1i}; \Theta_1), F_2(u_{2i}; \Theta_2), \ldots, F_d(u_{di}; \Theta_d); \Theta_c]}{\partial \Theta_c} = 0
\]

**Goodness-of-fit test**

The Cramér–von Mises test, which proved to be more effective than the K–S test (Mesfioui et al. 2010), was used to perform the goodness-of-fit test. The Cramér–von Mises statistic for copulas is expressed as

\[
S_n = n \int_{[0, 1]^d} (C_e - C_0)^2 \, dC_e
\]

**Table 1 | Notation and properties of bivariate Archimedean copula families**

<table>
<thead>
<tr>
<th>Name</th>
<th>Expression ( C_e(u, v) )</th>
<th>Generator function ( \varphi_\theta(t) )</th>
<th>Parameter range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gumbel</td>
<td>( \exp\left[-\left(\ln u\right)^\theta + \left(\ln v\right)^\theta\right]^{1/\theta} )</td>
<td>( (-\ln t)^\theta )</td>
<td>( \theta \geq 1 )</td>
</tr>
<tr>
<td>Clayton</td>
<td>( (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta} )</td>
<td>( \frac{1}{\theta} (t^{-\theta} - 1) )</td>
<td>( \theta &gt; 0 )</td>
</tr>
<tr>
<td>Frank</td>
<td>( -\frac{1}{\theta} \ln \left[1 + \frac{\left(\left(\frac{1}{\theta} - 1\right)\left(\frac{1}{\theta} - 1\right)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table><p>ight)}{e^{\theta} - 1}\right] ) | ( -\ln \left[\frac{e^{\theta} - 1}{e^{\theta} - 1}\right] ) | ( \theta \in \mathbb{R} \setminus {0} ) |
| Joe             | ( 1 - \left(\left(1 - u\right)^\delta + \left(1 - v\right)^\delta - \left(1 - u\right)^\delta\left(1 - v\right)^\delta\right)^{1/\delta} ) | ( -\ln \left(1 - (1 - \delta)^\theta\right) ) | ( \theta &gt; 1 ) |
| Clayton–Gumbel  | ( \left[1 + \left(\frac{\left(1 - u\right)^\delta + \left(1 - v\right)^\delta - 1\right)^{1/\delta} - 1\right)^{-1/\theta} \right]^{-1/\theta} ) | ( (t^{-\theta} - 1)^\theta ) | ( \theta &gt; 0, \delta \geq 1 ) |
| Joe–Gumbel      | ( 1 - \left(\left(1 - u\right)^\delta + \left(1 - v\right)^\delta - \left(1 - u\right)^\delta\left(1 - v\right)^\delta\right)^{1/\delta} ) | ( -\ln \left(1 - (1 - \delta)^\theta\right) ) | ( \theta &gt; 1, \delta \geq 1 ) |
| Joe–Clayton     | ( 1 - \left(\left(1 - u\right)^\delta + \left(1 - v\right)^\delta - \left(1 - u\right)^\delta\left(1 - v\right)^\delta\right)^{1/\delta} ) | ( \left(1 - (1 - \delta)^\theta\right)^\delta - 1 ) | ( \theta \geq 1, \delta &gt; 0 ) |
| Joe–Frank       | ( \frac{1}{\theta} \left{1 - \left(1 - \frac{1}{1 - (1 - \delta)}\left(1 - (1 - \delta u)^\theta(1 - \delta v)^\theta\right)^{1/\delta}\right)^\delta\right} ) | ( -\ln \left(\frac{1 - (1 - \delta)^\theta}{1 - (1 - \delta)^\theta}\right) ) | ( \theta \geq 1, \delta \in (0, 1] ) |</p>
where \( n \) is the sample size; \( C_0 \) is a specified parametric family of copulas; and \( \Theta \) is the estimated parameters derived from the pseudo-observations \( U_1, U_2, \ldots, U_d \); \( C_e \) is the empirical copula expressed as
\[
C_e = \frac{1}{n} \sum_{i=1}^{n} 1(U_{1i} \leq u_1, U_{2i} \leq u_2, \ldots, U_{di} \leq u_d)
\]  

(18)

The corresponding \( p \)-values were obtained via Monte Carlo methods, the detailed procedures of which have been provided by Genest et al. (2009).

Moreover, the Akaike information criterion (AIC) and Bayesian information criterion (BIC) were applied to select the best-fitting copula model. The formulas for them are as follows:
\[
\text{AIC} = -2\ln(L_e) + 2k
\]  

(19)
\[
\text{BIC} = -2\ln(L_e) + k \ln(n)
\]  

(20)

where \( L_e \) is the likelihood function of the joint distribution; \( k \) is the number of parameters; and \( n \) is the sample size. Generally, the smaller the AIC and BIC values, the better the distribution fits.

**Return periods and dangerous domains**

The return period indicates the average time between the occurrence of two hydrological events and is critical for flood prevention and mitigation work. Three types of bivariate return periods, including the joint return period \( (T_{or}) \), co-occurrence return period \( (T_{and}) \), and secondary return period \( (T_{sec}) \), are discussed here.

\( T_{or} \) refers to one of the MF \( X \) or Pr \( Y \) that is greater than or equal to a certain value, as well as the average interval required for each occurrence. \( T_{and} \) refers to both the MF \( X \) or Pr \( Y \) that are greater than or equal to a certain value, as well as the average interval required for each occurrence. \( T_{or} \) and \( T_{and} \) are also known as the primary return periods, whose formulas are expressed as
\[
T_{or} = \frac{\mu_T}{P([X \geq x] \cup [Y \geq y])} = \frac{\mu_T}{1 - C[F_X(x), F_Y(y)]}
\]  

(21)
\[
T_{and} = \frac{\mu_T}{P([X \geq x] \cap [Y \geq y])} = \frac{\mu_T}{1 - F_X(x) - F_Y(y) + C[F_X(x), F_Y(y)]}
\]  

(22)

where \( x \) and \( y \) are thresholds of \( X \) and \( Y \), respectively; \( F_X(x) \) and \( F_Y(y) \) are the marginal CDFs of the MF and Pr, respectively; \( C(\cdot) \) is the copula CDF; \( \mu_T \) is the average inter-arrival time between two successive events (\( \mu_T = 1 \) for the maximum annual events).

\( T_{sec} \), also known as the Kendall’s return period, describes the probability of occurrence of an event in the area over the copula level curve of value \( t \) (Salvadori & Michele 2004), and it can be expressed as
\[
T_{sec} = \frac{\mu_T}{1 - K_C(t)}
\]  

(23)

where \( t \in I \) is the probability level; \( K_C \) is Kendall’s distribution, which is the distribution of random variable \( C[F_X(x), F_Y(y)] \). For Archimedean copulas, \( K_C \) is expressed as
\[
K_C(t) = t - \frac{\varphi(t)}{\varphi'(t^+)}, \quad 0 < t \leq 1
\]  

(24)

where \( \varphi'(t^+) \) is the right derivative of the additive generator function \( \varphi(t) \), as presented in Table 1.

The dangerous domain \( D^* \), the notion of which was introduced by Salvadori et al. (Salvadori et al. 2011), contains all the dangerous events that are regarded as ‘more dangerous’ than the prescribed vector of thresholds \( (x^*, y^*) \). The dangerous domain identified under different return periods are different.

(1) Under \( T_{or} \):
\[
D^*_{or} = \{(x, y) \in R^2 : x > x^* \cup y > y^*\}
\]  

(25)

where at least one of the components exceeds a prescribed threshold, as shown in the gray-shaded area in Figure 3(a).

(2) Under \( T_{and} \):
\[
D^*_{and} = \{(x, y) \in R^2 : x > x^* \cap y > y^*\}
\]  

(26)

where both the components exceed a prescribed threshold, as shown in the gray-shaded area in Figure 3(b).
(3) Under $T_{sec}$:

For a bivariate distribution $H = C(F_x, F_y)$ and $t \in (0, 1)$, the critical line $L^H_t$ of level $t$ is defined as

$$L^H_t = \{(x, y) \in \mathbb{R}^2 : H(x, y) = t\}$$  (27)

Evidently, for any vector of thresholds $(x^*, y^*)$, a unique critical line $L^H_t$ exists. Hence, the dangerous domain identified under $T_{sec}$ is

$$D_{sec}^* = \{(x, y) \in \mathbb{R}^2 : H(x, y) > t\}$$  (28)

as shown in the gray-shaded area in Figure 3(c).

The definition of the multivariate return periods $T_{or}$ and $T_{and}$ is limited in the identification of dangerous domains, whereas $T_{sec}$ based on the Kendall measure is more rational. An advantage of $T_{sec}$ is that thresholds lying over the same critical line always generate the same dangerous domain; however, this does not apply when considering $T_{or}$ and $T_{and}$, as illustrated in detail in Figure 3 and Table 2.

**Conditional exceedance probability**

If an appropriate copula function is estimated, then the conditional joint distribution can be obtained. The conditional CDF of $X \leq x$ for $Y = y$ can be expressed as

$$C_{X|Y=y}(x) = P(X \leq x | Y = y) = \frac{\partial C(F_x(x), F_Y(y))}{\partial F_Y(y)}$$  (29)

Hence, the conditional exceedance probability of $X > x$ for $Y = y$ can be expressed as

$$F_{X \times Y = y} = 1 - C_{X|Y=y}(x) = 1 - \frac{\partial C(F_x(x), F_Y(y))}{\partial F_Y(y)}$$  (30)
RESULTS AND DISCUSSION

Determination of marginal distributions

We first tested the randomness of the variables by computing their autocorrelations. The autocorrelation plots of the MF and Pr series in the XC and DC are shown in Figure 4. As shown, the autocorrelations in the four series can be disregarded at the 5% significance level. Therefore, their univariate distribution fitting can be performed.

We used six distribution models (the GEV, Gumbel, P-III, gamma, normal, and log-normal distributions) to fit the marginal distributions for the MF and Pr series in the XC and DC, and the results are shown in Figures 5 and 6. As shown in both figures, the CDFs and PDFs for the marginal distributions exhibit different fitting performances between the estimated and empirical distributions.

The K-S test, MAE, PPCC, and DMC were used to evaluate the performances of six candidate distribution models, and the results are shown in Table 3. The p-values for all

<table>
<thead>
<tr>
<th>Type of the return period</th>
<th>Thresholds</th>
<th>Value of the return period</th>
<th>Dangerous domain</th>
<th>Pros and cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_{as} )</td>
<td>( A_1(x_{11}, y_{11}) )</td>
<td>( a_1 )</td>
<td>Polygon ( A_1 )</td>
<td>As shown in Figure 3(d), the joint return period of ( B_1 ) is larger than that of ( A_1 ); but part of the dangerous domain identified by ( B_1 ) is outside that identified by ( A_1 ) (Rectangle ( A_1 )). Hence, hydrological events in Rectangle ( A_1 ) are considered dangerous by a large return period point ( B_1 ) but safe by a small return period point ( A_1 ), which is unreasonable.</td>
</tr>
<tr>
<td></td>
<td>( B_1(x_{12}, y_{12}) )</td>
<td>( b_1(b_1 &gt; a_1) )</td>
<td>Polygon ( B_1 )</td>
<td></td>
</tr>
<tr>
<td>( T_{and} )</td>
<td>( A_2(x_{21}, y_{21}) )</td>
<td>( a_2 )</td>
<td>Rectangle ( A_2 )</td>
<td>As shown in Figure 3(e), the co-occurrence return period of ( B_2 ) is larger than that of ( A_2 ); but part of the dangerous domain identified by ( B_2 ) is outside that identified by ( A_2 ) (Rectangle ( B_2 )). Hence, hydrological events in rectangle ( B_2 ) are considered dangerous by a large return period point ( B_2 ) but safe by a small return period point ( A_2 ), which is unreasonable.</td>
</tr>
<tr>
<td></td>
<td>( B_2(x_{22}, y_{22}) )</td>
<td>( b_2(b_2 &gt; a_2) )</td>
<td>Rectangle ( B_2 )</td>
<td></td>
</tr>
<tr>
<td>( T_{acc} )</td>
<td>( A_3(x_{31}, y_{31}) )</td>
<td>( a_3 )</td>
<td>Area ( A_3 )</td>
<td>As shown in Figure 3(f), any point on the same secondary return period contour line has the same dangerous domain. The larger the return period, the smaller the dangerous domain. The dangerous domain with a smaller return period necessarily covers that with a larger return period. Compared with ( T_{as} ) and ( T_{and} ), this division of dangerous domain is more reasonable and conducive to the flood risk assessment.</td>
</tr>
<tr>
<td></td>
<td>( B_3(x_{32}, y_{32}) )</td>
<td>( b_3(b_3 &gt; a_3) )</td>
<td>Area ( B_3 )</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4 | Autocorrelation plots of MF and Pr series in XC and DC. (a) Autocorrelation plot of MF series in XC; (b) autocorrelation plot of Pr series in XC; (c) autocorrelation plot of MF series in DC; (d) autocorrelation plot of Pr series in DC. The lag of each autocorrelation estimate is denoted on the horizontal axis, and each autocorrelation estimate is indicated by height of vertical bars. A pair of horizontal, dashed lines represent lag-wise 95% confidence intervals centered at zero and are used for determining the statistical significance of an individual autocorrelation estimate at a specified lag vs. a null value of zero (i.e., no autocorrelation at that lag).
Figure 5 | Comparison of different probability density estimates with observed frequency for MF and Pr series in XC. (a) Comparison of CDFs of candidate distributions and observed frequency for MF series in XC; (b) comparison of PDFs of candidate distributions and observed frequency for MF series in XC; (c) comparison of CDFs of candidate distributions and observed frequency for Pr series in XC; (d) comparison of PDFs of candidate distributions and observed frequency for Pr series in XC.
Figure 6 | Comparison of different probability density estimates with observed frequency for MF and Pr series in DC. (a) Comparison of CDFs of candidate distributions and observed frequency for MF series in DC; (b) comparison of PDFs of candidate distributions and observed frequency for MF series in DC; (c) comparison of CDFs of candidate distributions and observed frequency for Pr series in DC; (d) comparison of PDFs of candidate distributions and observed frequency for Pr series in DC.
were greater than 0.05, indicating that all the candidate distributions were suitable for fitting the distribution of the MF and Pr series at the 5% significance level. It can be concluded that the MF and Pr series in the XC fitted the gamma distribution the best, whereas the MF and Pr series in the DC fitted the GEV and gamma distributions the best, respectively.

**Establishment of copula model**

We measured the dependence between the MF and Pr series to determine whether it is appropriate to establish their copula model. The Kendall rank correlation coefficients between the MF and Pr series in the XC and DC were 0.327 and 0.403, respectively, which were greater than the threshold at the 5% significance level, i.e. 0.273. The Spearman's rank correlation coefficients between the MF and Pr series in the XC and DC were 0.480 and 0.569, respectively, which were greater than the threshold at the 5% significance level, i.e., 0.297. This signified that the null hypothesis at the 5% significance level was rejected and a correlation existed between the two variables. Therefore, it is feasible to use the copula function to establish the joint probability distribution between them.

The parameters of eight candidate copulas estimated using the IFM method and their goodness-of-fit tests through the Cramér–von Mises test, AIC, and BIC are shown in Table 4. As shown, all the p-values were greater than 0.05, indicating that all candidate copulas can be applied to fit the dependence structure between the MF and Pr series. The Joe–Frank copula performed the best for modeling the dependence structure of the MF and Pr series in the XC, whereas the Clayton copula performed the best for modeling the dependence structure of the MF and Pr series in the DC.

**Coincidence probability**

The 80th, 90th, 93.33th, and 95th quantiles of the Pr series were selected as the typical extreme precipitation events in both the XC and the DC, and they were 1,136, 1,245, 1,304,
and 1,342 mm in the XC, respectively, and 1,084, 1,181, 1,231, and 1,265 mm in the DC, respectively. The typical extreme flood events in the XC were designated as 800, 1,000, 1,200, and 1,400 m$^3$/s, respectively. Here, 800 m$^3$/s is the safety discharge capacity of the Hengtangcun station. Safety discharge refers to the maximum discharge to pass the river safely under normal circumstances; 1,400 m$^3$/s is the warning discharge determined based on the historical flood peak discharges at the Hengtangcun station. When the flood peak reaches the warning discharge, the flood control department focuses on flood control emergencies. The values 1,000 and 1,200 m$^3$/s correspond to two extreme flood events between the safety discharge and warning discharge. Similarly, the typical extreme flood events in the DC were designated as 500, 550, 650, and 750, respectively; 500 m$^3$/s is the safety discharge capacity of the Pingyao station; 750 m$^3$/s is the warning discharge; 550 and 650 m$^3$/s are two extreme flood events between the safety discharge and warning discharge in the DC.

The situation when the MF does not exceed the safety discharge is defined as a safe flood, which is beneficial for flood control; meanwhile, the situation when the MF is greater than or equal to the warning discharge is defined as a dangerous flood, which is adverse to flood control.

Using the established copula models, the coincidence probabilities of the typical MF and Pr events in the XC and DC were calculated to study the joint risk of the two extreme events, the results of which are shown in Table 5.

The following were observed in Table 5:

1. For cases with the same MF, with the increase in the Pr, the marginal distribution probability and coincidence probability of the Pr decreased. Similarly, for cases with the same Pr, as the MF increased, the marginal distribution probability and coincidence probability of the MF decreased.

2. As the MF increased, the sensitivity of its coincidence probability to changes in the Pr increased. Similarly, as the Pr increased, the sensitivity of its coincidence probability to changes in the MF increased.

Hence, we conclude that in the XC, the coincidence probabilities of encountering dangerous floods with the Pr frequency exceeding 80%, 90%, 93.33%, and 95% were 22.80%, 13.69%, 10.69%, and 9.19%, respectively. In the DC, the coincidence probabilities of encountering dangerous floods with the Pr frequency exceeding 80%, 90%, 93.33%, and 95% were 21.84%, 12.39%, 9.26%, and 7.70%, respectively.

Table 4 | Estimated parameters and goodness-of-fit tests of the candidate copulas

<table>
<thead>
<tr>
<th>Catchment</th>
<th>Copula model</th>
<th>Parameters</th>
<th>p-value</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>XC</td>
<td>Gumbel</td>
<td>1.43 / 0.51</td>
<td>0.51</td>
<td>−335.44</td>
<td>−333.49</td>
</tr>
<tr>
<td></td>
<td>Clayton</td>
<td>0.77 / 0.13</td>
<td>0.12</td>
<td>−331.94</td>
<td>−331.98</td>
</tr>
<tr>
<td></td>
<td>Frank</td>
<td>3.21 / 0.27</td>
<td>0.79</td>
<td>−350.24</td>
<td>−328.29</td>
</tr>
<tr>
<td></td>
<td>Joe</td>
<td>1.64 / 1.00</td>
<td>0.12</td>
<td>−331.94</td>
<td>−328.03</td>
</tr>
<tr>
<td></td>
<td>Clayton–Gumbel</td>
<td>0.77 / 1.00</td>
<td>0.12</td>
<td>−331.94</td>
<td>−328.03</td>
</tr>
<tr>
<td></td>
<td>Joe–Gumbel</td>
<td>1.00 / 1.43</td>
<td>0.49</td>
<td>−333.44</td>
<td>−329.54</td>
</tr>
<tr>
<td></td>
<td>Joe–Clayton</td>
<td>1.00 / 0.77</td>
<td>0.13</td>
<td>−331.94</td>
<td>−328.03</td>
</tr>
<tr>
<td></td>
<td>Joe–Frank</td>
<td>3.65 / 0.68</td>
<td>0.90</td>
<td>−343.84</td>
<td>−339.94</td>
</tr>
<tr>
<td>DC</td>
<td>Gumbel</td>
<td>1.70 / 0.70</td>
<td>0.70</td>
<td>−336.98</td>
<td>−335.03</td>
</tr>
<tr>
<td></td>
<td>Clayton</td>
<td>1.26 / 0.47</td>
<td>0.07</td>
<td>−344.63</td>
<td>−342.68</td>
</tr>
<tr>
<td></td>
<td>Frank</td>
<td>4.29 / 0.15</td>
<td>0.66</td>
<td>−344.30</td>
<td>−342.35</td>
</tr>
<tr>
<td></td>
<td>Joe</td>
<td>2.08 / 0.66</td>
<td>0.35</td>
<td>−342.63</td>
<td>−338.73</td>
</tr>
<tr>
<td></td>
<td>Clayton–Gumbel</td>
<td>1.26 / 1.00</td>
<td>0.51</td>
<td>−341.41</td>
<td>−337.51</td>
</tr>
<tr>
<td></td>
<td>Joe–Gumbel</td>
<td>1.63 / 1.22</td>
<td>0.15</td>
<td>−342.63</td>
<td>−338.73</td>
</tr>
<tr>
<td></td>
<td>Joe–Clayton</td>
<td>1.00 / 1.26</td>
<td>0.59</td>
<td>−341.17</td>
<td>−337.27</td>
</tr>
<tr>
<td></td>
<td>Joe–Frank</td>
<td>2.70 / 0.93</td>
<td>0.59</td>
<td>−341.17</td>
<td>−337.27</td>
</tr>
</tbody>
</table>

Notes: P-value > 0.05 indicates that the estimated copula can be accepted at the 5% significance level. The smaller the value of the AIC and BIC, the better the distribution fits. The most appropriate copulas are indicated in boldface.
Return periods and dangerous domains

The bivariate return periods of various combinations of the flood and extreme precipitation are more effective for actual flood control and management than univariate return period analysis; therefore, the joint return, co-occurrence return, and secondary return periods of the joint distribution of the MF and Pr series in the XC and DC were calculated. The contour maps of different return periods are shown in Figure 7, and some important values are shown in Table 6.

If the MF and Pr are known, one can determine their $T_{or}$, $T_{and}$, and $T_{sec}$ based on Figure 7. Moreover, using the method described in the section on Return periods and dangerous domains the dangerous domains of the thresholds under different return periods can be identified.

The following were obtained based on Table 6 and Figure 7:

1. The return periods satisfied the following two theoretical inequalities:

$$T_{or} \leq \min(T_{un1}, T_{un2}) \leq \max(T_{un1}, T_{un2}) \leq T_{and}$$ (31)

$$T_{or} < T_{sec} < T_{and},$$ (32)

where $T_{un1}$ and $T_{un2}$ are the univariate return periods. For example, for the MF = 1,003.70 m$^3$/s, Pr = 1,136.66 mm pair in the XC (corresponding to a return period $T_{un1} = T_{un2} = 5$ years estimated using the fitted gamma distributions for MF and Pr series, respectively), $T_{or}$, $T_{and}$, and $T_{sec}$ were 3.08, 13.35, and 8.26 years, respectively. This finding is consistent with those of previous studies (Vandenberghhe et al. 2010; Requena et al. 2013).

2. The higher the return period, the greater was the differences among $T_{or}$, $T_{and}$, and $T_{sec}$.

3. The return periods of the XC and DC were not significantly different, indicating that the joint frequencies of the MF and Pr in the two catchments were the same. However, under the same univariate return period, $T_{or}$ in the XC was slightly larger than that in the DC, whereas $T_{and}$ and $T_{sec}$ were slightly smaller than those in the DC.

The contour map trend of the return periods was consistent those of previous studies (Bezak et al. 2014; Fan et al. 2016), indicating that the results were reasonable.

Flood risk under specified precipitation scenarios

In practice, water resource managers are more interested in the possibility of floods exceeding certain thresholds in certain extreme precipitation events. Hence, we calculated the conditional exceedance probabilities of the MF under different Pr scenarios. Four precipitation scenarios were presented: the 80th, 90th, 93.33th, and 95th quantiles of the Pr series. The values of the conditional exceedance probabilities are listed in Table 7, and their contour maps are shown in Figure 8.
The following were obtained based on Table 7 and Figure 8:

1. Under the same Pr scenario, as the MF increased, the conditional exceedance probability decreased.

2. For the same MF, the conditional exceedance probability in the heavy precipitation scenario was greater than that in the small precipitation scenario, i.e., the conditional exceedance probabilities of the MF increased with the amount of Pr.

---

**Figure 7** Contour maps of different return periods for XC and DC. (a) Contour map of joint return periods for XC; (b) contour map of co-occurrence return periods for XC; (c) contour map of secondary return periods for XC; (d) contour map of joint return periods for DC; (e) contour map of co-occurrence return periods for DC; (f) contour map of secondary return periods for DC. The curves represent contour lines of different return periods. Number labels near contour lines represent values of different return periods.

**Table 6** Univariate and bivariate return periods for MF and Pr series in XC and DC

<table>
<thead>
<tr>
<th>Catchment</th>
<th>Univariate</th>
<th>Bivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T_{un}$ (years)</td>
<td>MF ($m^3/s$)</td>
</tr>
<tr>
<td>XC</td>
<td>5</td>
<td>1,003.70</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1,203.64</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>1,312.28</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>1,386.69</td>
</tr>
<tr>
<td>DC</td>
<td>5</td>
<td>612.98</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>676.87</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>705.82</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>723.56</td>
</tr>
</tbody>
</table>
These conditional exceedance probabilities enabled the possibility of certain flood discharges under different precipitation scenarios to be identified. Using the XC for example, for $Pr = 1,136$ mm in the future, the probability of $MF > 800$ m$^3$/s is 61.89%.

Hence, we conclude the following:

(1) For the XC, in the four scenarios with $Pr$ frequencies of 80%, 90%, 93.33%, and 95%, the probabilities of safe floods were 38.11%, 35.74%, 35.00%, and 34.64%, respectively, and the probabilities of dangerous floods were 9.72%, 10.57%, 10.86%, and 11.01%, respectively.

(2) For the DC, in the four scenarios with $Pr$ frequencies of 80%, 90%, 93.33%, and 95%, the probabilities of safe floods were 50.15%, 58.30%, 57.37%, and 57.45%, respectively, and the probabilities of dangerous floods were 5.91%, 6.31%, 6.44%, and 6.51%, respectively.

The contour map trend of the conditional exceedance probabilities was consistent with that of a previous study (Guo et al. 2018), indicating the results were reasonable.

The contributions of this study are as follows:

(1) We focused on the flood risk in the main water source of Taihu Lake, i.e. the Zhexi Region, which had a direct

Table 7 | Conditional exceedance probabilities of MF under given Pr scenarios in XC and DC

<table>
<thead>
<tr>
<th>Catchment</th>
<th>Conditional probabilities of MF under given Pr scenarios</th>
</tr>
</thead>
<tbody>
<tr>
<td>XC</td>
<td></td>
</tr>
<tr>
<td>MF &gt; 800 m$^3$/s</td>
<td>$Pr = 1,136$ mm 61.89%</td>
</tr>
<tr>
<td>MF &gt; 1,000 m$^3$/s</td>
<td>37.84%</td>
</tr>
<tr>
<td>MF &gt; 1,200 m$^3$/s</td>
<td>20.13%</td>
</tr>
<tr>
<td>MF &gt; 1,400 m$^3$/s</td>
<td>9.72%</td>
</tr>
<tr>
<td>DC</td>
<td></td>
</tr>
<tr>
<td>MF &gt; 500 m$^3$/s</td>
<td>$Pr = 1,084$ mm 69.85%</td>
</tr>
<tr>
<td>MF &gt; 550 m$^3$/s</td>
<td>69.43%</td>
</tr>
<tr>
<td>MF &gt; 650 m$^3$/s</td>
<td>25.65%</td>
</tr>
<tr>
<td>MF &gt; 750 m$^3$/s</td>
<td>5.91%</td>
</tr>
</tbody>
</table>

Figure 8 | Contour maps of conditional exceedance probabilities of MF under various $Pr$ scenarios in XC and DC. (a) Contour map of conditional exceedance probabilities of MF under various $Pr$ scenarios in XC; (b) contour map of conditional exceedance probabilities of MF under various $Pr$ scenarios in DC. The horizontal lines represent various $Pr$ scenarios. The curves represent contour lines of different conditional exceedance probabilities. The number labels near contour lines represent values of different conditional exceedance probabilities.
and significant impact on the floods in the Taihu Basin. This region had not been emphasized in previous studies, whereas it is vital for flood prevention and drainage work.

(2) Extreme hydrological and meteorological events (i.e., the flood peak flow and precipitation during the flood season), which were the main flood disaster factors, were considered as the objects of this study. Their joint probability distribution and combined effects were directly related to the degree of flood risk and further affected the design and management of flood prevention and drainage systems in the Taihu Basin. Moreover, considering the extreme hydrological and meteorological events simultaneously can illustrate the inherent law of flood events more effectively and enable the relationship between flood and extreme precipitation to be analyzed.

CONCLUSIONS

To study the response of flood events to extreme precipitation in the Zhexi Region of the Taihu Basin, a copula model was established in this study to analyze the bivariate frequency of the MF and Pr series in the XC and DC. The 80th, 90th, 93.33th, and 95th quantiles of the Pr series were selected as typical extreme precipitation events. The safety discharge, warning discharge, and two extreme flood events between them were selected as typical extreme flood events. The coincidence probabilities of the typical MF and Pr events were calculated. The joint return, co-occurrence return, and secondary return periods for arbitrary combinations of the MF and Pr were calculated, and their contour maps were drawn. Meanwhile, the dangerous domains under different return periods were identified. Furthermore, flood risk analysis under different extreme precipitation scenarios was performed, and contour maps of the conditional exceedance probabilities were drawn. The main conclusions were as follows:

(1) According to the K–S test, MAE, PPCC, and DMC, both the MF and Pr in the XC fitted the gamma distribution the best, whereas the MF and Pr in the DC fitted the GEV and gamma distributions the best, respectively.

Based on the Cramér–von Mises test, AIC, and BIC, the Joe–Frank copula and Clayton copula were selected as the most appropriate copulas to fit the joint distribution of the MF and Pr in the XC and DC, respectively.

(2) In the XC, the coincidence probabilities of encountering dangerous floods with the Pr frequency exceeding 80%, 90%, 93.33%, and 95% were 22.80%, 13.69%, 10.69%, and 9.19%, respectively. In the DC, the coincidence probabilities of encountering dangerous floods with the Pr frequency exceeding 80%, 90%, 93.33%, and 95% were 21.84%, 12.39%, 9.26%, and 7.70%, respectively.

(3) Different return periods for the same pair of MF and Pr satisfied the following two rules: $T_{or} \leq \min[T_{un1}, T_{un2}] \leq \max[T_{un1}, T_{un2}] \leq T_{and}$ and $T_{or} < T_{sec} < T_{and}$.

(4) In the four scenarios with Pr frequencies of 80%, 90%, 93.33%, and 95%, the probabilities of safe floods were 38.11%, 35.74%, 35.00%, and 34.64% for the XC, respectively, and 30.15%, 28.30%, 27.73%, and 27.45% for the DC, respectively. In the four scenarios with Pr frequencies of 80%, 90%, 93.33%, and 95%, the probabilities of dangerous floods were 9.72%, 10.57%, 10.86%, and 11.01% for the XC, respectively, and 5.91%, 6.31%, 6.44%, and 6.51% for the DC, respectively.

Water resources managers can identify the corresponding flood risk information for different flood and precipitation events from the contour map of return periods (Figure 7). Furthermore, the probability of a certain flood occurring under various precipitation scenarios can be assessed based on the contour map of conditional exceedance probabilities (Figure 8).

The results of this study can provide a practical basis and guidance for the computation of rainstorm designs, management of flood control safety, and water resource scheduling in the Taihu Basin. The framework designed in this study for analyzing the response of flood events to extreme precipitation can be applied to similar river basins. Furthermore, the flood generation mechanism has changed owing to climate change and human activities. Therefore, for future studies, we plan to perform flood frequency analysis under changed flood generation mechanisms to identify the effects of climate change and...
human activities on floods, as well as to provide a more practical reference for flood control.

**FUNDING**

This research was funded by the National Key Research and Development Program of China, grant number 2018YFC1508200.

**DATA AVAILABILITY STATEMENT**

All relevant data are included in the paper or its Supplementary Information.

**REFERENCES**


First received 10 July 2020; accepted in revised form 19 December 2020. Available online 4 January 2021.