

Temporal Analysis and the Joint Probability Distribution of Rainfall Indexes in the Taihu Lake Basin, China

Manqiu Hao

Hohai University <https://orcid.org/0000-0003-4555-0658>

Cheng Gao (✉ gchohai@163.com)

Hohai University <https://orcid.org/0000-0003-1622-0190>

Jian Chen

Shanghai Construction Group

Research Article

Keywords: Rainfall characteristics, Joint probability distribution, Three-dimensional copula function

Posted Date: June 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-563694/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Temporal Analysis and the Joint Probability Distribution of Rainfall Indexes in the Taihu Lake Basin, China

Manqiu Hao, Cheng Gao*, Jian Chen

Abstract

Taking the Taihu Lake Basin as an example, in this study, the characteristics of the rainfall factors in the study area were analyzed using daily rainfall data from 1955 to 2018. Three factors, i.e., the contribution rate of the rainfall in the flood season, the rainfall frequency, and the maximum daily rainfall, were selected to determine the optimal probability distribution function for each single factor. Furthermore, the root mean square error (RMSE) goodness of fit test was used to determine the optimal copula function for the three-dimensional joint probability distribution characteristics of the rainfall factors. The research results show that the three-dimensional copula joint probability method contains much more information than the results of the single variable probability calculation. The copula function can be used to analyze the multi-dimensional joint distribution of rainfall factors, which can fill the gap in research on multiple rainfall factors.

Keywords: Rainfall characteristics; Joint probability distribution; Three-dimensional copula function.

1 Introduction

The impacts of climate change have gradually increased, and coupled with the aggravation caused by human activities, they have resulted in remarkable changes in the hydrological processes around the world (Liu, 2020). The Intergovernmental Panel on Climate Change's (IPCC) Fifth Assessment Report (AR5) shows that the economic losses caused by extreme climate events will increase gradually (Brunner et al., 2017). Rainstorms are one of many extreme weather events, and they have features such as shorter durations and higher intensities (Li et al., 2019). The frequent occurrence of rainstorms increases the probability of urban waterlogging, and the disaster loss caused by urban waterlogging increases at an unprecedented rate (Alcamo et al., 2000; Fischer et al., 2016).

Over the last 100 years, the global rainfall situation has changed significantly. In particular, the extreme rainfall in the middle and high latitudes has increased significantly. According to research conducted by Chinese and foreign scholars, the maximum daily rainfall in most parts of the world exhibits an increasing trend, and the daily rainfall intensity is also increasing. Research on rainfall characteristics can directly reflect the temporal and spatial variations and periodic characteristics of regional rainfall. The commonly used methods include the sliding t-test, the Mann-Kendall trend test, and rescaled range (R/S) analysis (Lu et al., 2019). Sayemuzzanman et al. (2014)

*Corresponding author: Cheng Gao. E-mail: gchohai@163.com

analyzed the seasonal and annual rainfall trends in North Carolina in the United States. Their results revealed that there was a significant mutation in North Carolina from 1960 to 1970. Duhan et al. (2013) used statistical analysis to study the spatial and temporal variation characteristics of the annual rainfall in Madhya Pradesh, India. Their results revealed that the rainfall data from almost all of the rainfall stations from 1901 to 1978 exhibit a decreasing trend.

At present, most of the studies conducted on rainstorms have focused on the temporal and spatial distributions of rainfall and simulations of the waterlogging caused by urban rainstorms. In terms of rainstorm risk, increasing attention has been paid to the multivariate joint risk analysis method, which describes rainstorms from multiple perspectives (Pedro-Monzonís et al., 2015). Correlations generally exist between the internal attributes of hydrological events, such as the duration and intensity of rainstorms, the flood peak and volume, and the duration and intensity of droughts (Yin et al., 2018). Copula functions have been widely used in the fields of rainfall multivariate joint probability analysis and flood joint risk analysis because of their advantages, including a flexible structure and an unlimited marginal distribution type (Wang, 2017). The theory and application of two-dimensional copula functions are relatively mature (Qian et al., 2017).

Ganguli et al. (2013) described the joint probability distribution of two variables and three variables (i.e., the peak discharge, flood volume, and flood duration) using various types of connection functions, and they discussed the characteristics of two return periods and tail correlation. Yang et al. (2019) analyzed the risk posed by the duration and accumulated rainfall based on a bivariate copula distribution model for heavy rainfall variables. Salvadori et al (2004) used two-dimensional Frank copula functions to simulate the joint probability distribution of the rainfall duration and rainfall intensity, deduced the isoline of the joint return period, and introduced the concept of the second return period. Gao et al. (2019) used two-dimensional copula functions for the joint statistical analysis of the maximum daily rainfall and its occurrence date.

The theory of three-dimensional and higher-dimensional copula functions is more complex due to the expansion of the dimensions and the nondependence among the variables, so is relatively lacking research in this area. In this study, several factors of rainfall were selected, and copula functions were used to determine the two-dimensional and three-dimensional joint distributions. Three indexes (i.e., the contribution rate of the rainfall in the flood season, the maximum daily rainfall, and the rainstorm frequency) were used to reveal the inherent probability distribution characteristics.

2 Study area and database

The Taihu Lake Basin is located in the southern part of the Yangtze River Delta in China (30°05′–32°08′N, 119°08′–121°55′E). The Taihu Basin is dominated by plains, accounting for two thirds of the total area. The basin covers four provinces: Jiangsu, Zhejiang, Shanghai, and Anhui. It has a total area of 36900 km². The climate of the basin is a typical semi-tropical monsoon

climate. The annual mean temperature in the basin is about 15–17°C and increases from north to south. The general situation of the study area is shown in Figure 1.

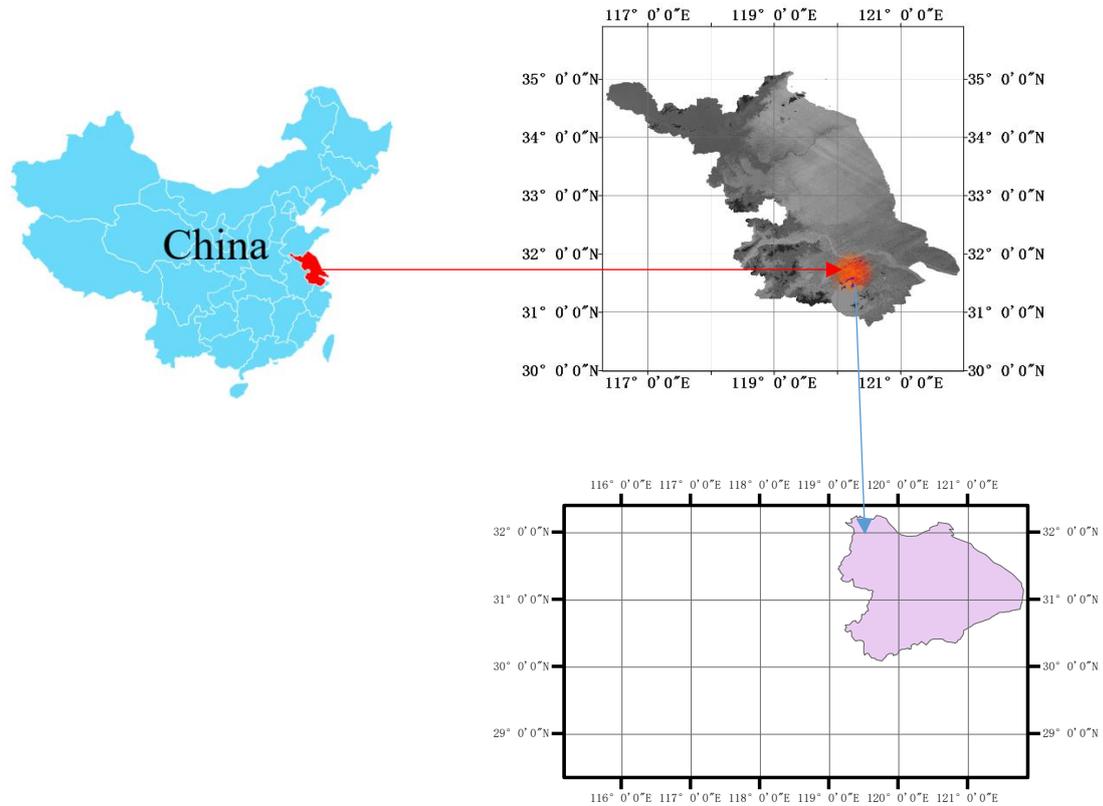


Fig. 1. Geographical location of the study area

The daily rainfall datasets from 1955 to 2019 were obtained from the China Meteorological Data Shared Services System. Data from 11 stations were used in this study, and the information for the stations is presented in Table 1.

Table 1. Geographic coordinates of the rainfall stations

Station	Longitude (E)	Latitude (N)	No
Shangpei	119°13′	31°27′	1
Hufu	119°47′	31°13′	2
Hekou	119°15′	31°23′	3
Guanlin	119°43′	31°30′	4
Jiulipu	119°46′	31°52′	5
Caoqiao	119°58′	31°31′	6
Xukou	120°28′	31°13′	7
Liqiao	120°16′	31°33′	8
Changshou	120°22′	31°48′	9
Wangting	120°25′	31°27′	10
Jinjiaba	120°46′	31°6′	11

Wangyuzha	120°49'	31°46'	12
Jiashan	120°55'	30°51'	13
Zhapu	121°4'	30°51'	14
Pingyao	119°58'	30°25'	15
Dixi	120°1'	30°40'	16
Meixi	119°45'	30°48'	17
Guanshang	119°52'	30°46'	18
Miaoxi	120°0'	30°46'	19
Xiaomeikou	120°7'	30°55'	20
Henghu	119°47'	30°27'	21
Tianjintang	119°15'	30°31'	22
Fushishuiku	119°28'	30°37'	23
Laoshikan	119°30'	30°34'	24
Changzhou	119°58'	31°46'	25
Liqiao	120°16'	31°33'	26
Wangting	120°25'	31°27'	27
Zhitang	120°25'	31°27'	28
Rulin	119°34'	31°37'	29
Ganlu	120°34'	31°33'	30

3 Methodology

3.1 Mann-Kendall trend test

The Mann-Kendall trend test is a nonparametric testing method that does not require that the samples follow a certain distribution, and the results are not easily affected by outliers, so it is more suitable for revealing the change trend of hydrological time series (Hamed et al., 2009).

Suppose X_1, X_2, \dots, X_n is the variable of the rainfall series, and n is the length of the rainfall series, i.e., the statistical year. The Mann-Kendall method defines the statistic S as

$$S = \sum_{i < j} a_{ij}. \quad (1)$$

$$a_{ij} = \text{sgn}(x_j - x_i) =$$

$$\begin{cases} 1 & (x_i < x_j) \\ 0 & (x_i = x_j). \\ -1 & (x_i > x_j) \end{cases} \quad (2)$$

When the number of samples n is relatively large, the statistic S is generally exhibits a normal distribution and the mean value is 0. The variance is as follows:

$$\text{Var}(S) = n(n-1)(2n+5)/18. \quad (3)$$

The Mann-Kendall statistic Z can be obtained as follows:

$$Z = \begin{cases} (s-1) / \sqrt{\text{var}(S)} & (S > 0) \\ 0 & (S = 0) \\ (s+1) / \sqrt{\text{var}(S)} & (S < 0) \end{cases} \quad (4)$$

In the bilateral test, for the statistic Z , if $Z > 0$, the time series exhibits an upward trend; if $Z < 0$, the time series exhibits a downward trend. Given the confidence level α , if $|Z| \geq Z_{1-\alpha/2}$, there is an obvious upward or downward trend in the confidence level α . In hydrological statistics, the confidence level α is usually selected as 0.05 (Lab, 2005).

3.2 Mann-Kendall mutation test

Suppose X_1, X_2, \dots, X_n is the time series variable and n is the length of the time series. It can be defined as:

$$S_k = \sum_{j=1}^k (X_j - \bar{X}) \quad (k=2,3,\dots,n). \quad (5)$$

When $X_j > X_i$, $r_i = 1$; and when $X_j \leq X_i$, $r_i = 0$ ($j = 1, 2, \dots, i$).

The statistic UF_k is defined as

$$\begin{aligned} UF_k &= [S_k - E(S_k)] / [\text{Var}(S_k)]. \\ E(S_k) &= n(n-1)/4. \\ \text{Var}(S_k) &= n(n-1)(2n+5)/72. \end{aligned} \quad (6)$$

The time series X_n, X_{n-1}, \dots, X_1 is calculated again to obtain the statistical sequence UF'_k , and to make $UB_k = UF'_{n-k}$ ($k = 1, 2, \dots, n$).

The time of the mutation can be determined by analyzing statistical sequences UF_k and UB_k . If $UF_k > 0$, the sequence exhibits an upward trend; if $UF_k < 0$, the sequence exhibits a downward trend. If UF_k exceeds the critical value, the change trend is significant. If there is an intersection between UF_k and the curve and the intersection is between the critical lines of the 0.05 confidence level, then the time corresponding to the intersection is the start time of the mutation.

3.2 Definition of a copula function

A copula function is a type of connection function that describes the correlation between variables by constructing the dependency structure between the variables. It is an effective function for connecting the marginal distribution and the joint distribution of variables. One of the advantages of the copula function is that it is not subject to the marginal distribution of the selection restrictions (Zhang et al., 2014.).

If the random variables X_1, X_2, \dots, X_n obey the marginal distribution of $F_{x_1}, F_{x_2}, \dots, F_{x_n}$, the joint distribution function is $H(X_1, X_2, \dots, X_n)$. According to the Sklar theorem, there is copula joint function that combines n variables, that is,

$$H(x_1, \dots, x_n) = C(F_{x_1}(x_1), \dots, F_{x_n}(x_n)) = C(u_1, \dots, u_n), \quad (7)$$

where the marginal distribution function of the random variable X_i is $F_{x_i}(x_i) = u_i$, ($i=1, 2, \dots, n$). If $F_{x_1}, F_{x_2}, \dots, F_{x_n}$ are all continuous functions, then the copula function C is unique. In contrast, if C

is an n-dimensional copula function, and $F_{x_1}, F_{x_2}, \dots, F_{x_n}$ are distribution functions, then the function H must be an n-dimensional joint probability. The distribution has a marginal distribution of $F_{x_1}, F_{x_2}, \dots, F_{x_n}$.

Several common copula functions in the field of hydrology include the Archimedean copula, the elliptical copula, and the empirical copula functions. In hydrological analysis, the main method used is the Archimedean copula function. Its clear expression, few parameters, and simple structure make it superior (Yu et al., 2014).

When the Archimedean copula function contains only one parameter, it is a symmetric copula function. A n-dimensional copula $C^n: [0,1]^n \rightarrow [0,1]$ can be defined as

$$C(u_1, u_2, \dots, u_n) = \varphi^{-1} \left[\sum_{k=1}^n \varphi(u_k) \right] = \varphi^{-1} \left[\varphi(u_1) + \varphi(u_2) + \dots + \varphi(u_n) \right],$$

$$u_k \in [0,1], k = 1, 2, \dots, n, \quad (8)$$

where $\varphi(\cdot)$ is the generating function of the Archimedean copula function and φ^{-1} is the inverse function of φ .

The Archimedean copula function does not necessarily have a density function. When the Archimedean copula function is an absolute continuous function, there is a density function. The expression is as follows:

$$c(u_1, u_2, \dots, u_n) = \varphi^{-1(n)} \left[\varphi(u_1) + \varphi(u_2) + \dots + \varphi(u_n) \right] \prod_{k=1}^n \varphi'(u_k), \quad (9)$$

where $\varphi^{-1(n)}$ is the Nth derivative of the inverse function of the generating function and $\varphi'(\cdot)$ is the first derivative of φ . Common functions include the G–H copula, Clayton copula, and Frank copula functions. Their function expressions are shown in Table 2 (Latif et al., 2020).

Table 2. Three Archimedean copula function forms

Copula	$\varphi(t)$	$C(u_1, u_2, \dots, u_n)$	parameter
Gumbel	$(-\ln t)^\theta$	$\exp \left\{ - \left[\sum_{i=1}^n (-\ln u_i)^\theta \right]^{1/\theta} \right\}$	$\theta \geq 1$
Clayton	$\frac{1}{\theta} (t^{-\theta} - 1)$	$\left[\left(\sum_{i=1}^n u_i^{-\theta} \right) - n + 1 \right]^{-1/\theta}$	$\theta > 0$
Frank	$-\ln \frac{e^{-\theta t} - 1}{e^{-\theta} - 1}$	$-\frac{1}{\theta} \ln \left[1 + \frac{\prod_{i=1}^n (e^{-\theta u_i} - 1)}{(e^{-\theta} - 1)^{n-1}} \right]$	$R \setminus \{0\}$

The forms of the three-dimensional symmetric Archimedean copula function are shown in Table

3.

Table 3. Three-dimensional symmetric Archimedean copula function

Copula	copula function forms
Gumbel	$C(u_1, u_2, u_3) = \exp\left(-\left[(-\ln(u_1))^\theta + (-\ln(u_2))^\theta + (-\ln(u_3))^\theta\right]^{\frac{1}{\theta}}\right)$, $\theta \in (0, \infty)$
Clayton	$C(u_1, u_2, u_3) = (u_1^{-\theta} + u_2^{-\theta} + u_3^{-\theta} - 2)^{-1/\theta}$, $\theta \in (1, \infty)$
Frank	$C(u_1, u_2, u_3) = -\frac{1}{\theta} \ln\left(1 + \frac{(\exp(-\theta u_1) - 1)(\exp(-\theta u_2) - 1)(\exp(-\theta u_3) - 1)}{(\exp(-\theta) - 1)}\right)$

3.3 Parameter estimation of a copula function

3.3.1 Correlation index method

The parameters of a copula function can be estimated via kernel density estimation, the correlation index method, the inference for margins (IFM) method, and the exact maximum likelihood (EML) method (Akbari et al., 2019). In this study, the parameters were obtained by establishing the correlation between the Kendall rank correlation coefficient τ and the copula function parameter θ . The unified expression is as follows:

$$1. \quad \tau = 4 \int C(u, v) dC(u, v) - 1 \quad (10)$$

Dependence formula for the Gumbel Copula:

$$\tau = 1 - \frac{1}{\theta}; \theta \in [1, \infty). \quad (11)$$

Dependence formula for the Clayton Copula:

$$\tau = \frac{\theta}{2} + \theta; \theta \in (0, \infty). \quad (12)$$

Dependence formula for the Frank Copula:

$$\tau = \frac{4}{\theta} \left[\frac{1}{\theta} \int_0^\theta \frac{\tau}{\exp(\tau) - 1} \right]; \theta \in [-1, 1]. \quad (13)$$

3.3.2 Goodness of fit test

To verify the fitting degree between the theoretical calculation curve of the copula function and the empirical cumulative probability of the measured samples, it is necessary to test the goodness of fit of the constructed model. The evaluation of the goodness of fit is an important standard for selecting the optimal function of the distribution line. The commonly used indexes are the Root Mean Square Error (RMSE), the Akaike Information Criterion (AIC), and the Bayes Information Criterion (BIC).

1) RMSE method

In terms of testing the fitting effect of the copula function, the root mean square error method is a common. The smaller the RMSE value, the better the fitting effect of the copula function. The formula is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (F_{emp} - C)^2}, \quad (14)$$

where F_{emp} is the empirical probability value, C is the theoretical probability value, and N is the length of the series. The smaller the RMSE value, the better the fit of the model.

The formula for the three-dimensional empirical frequency is

$$\begin{aligned} F_{emp}(x_{i1}, x_{i2}, x_{i3}) &= P(X_{i1} \leq x_{i1}, X_{i2} \leq x_{i2}, X_{i3} \leq x_{i3}) \\ &= \left(\sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n n_{jkl} - 0.44 \right) / (n + 0.12), \end{aligned} \quad (15)$$

where n_{jkl} is the number of joint observations when $X_{i1} \leq x_{i1}, X_{i2} \leq x_{i2}, X_{i3} \leq x_{i3}$ is satisfied at the same time (Filipova et al., 2018).

2) AIC method

The AIC method is an algorithm based on information measurement, which is suitable for testing the copula model obtained using the maximum likelihood estimation method. The formula is as follows:

$$\text{AIC} = n \log(\text{MSE}) + 2m, \quad (16)$$

where

m is the number of parameters and n is the sample capacity. The smaller the AIC value, the better the fitting effect of the copula function.

3) BIC method

Similar to the AIC method, the BIC method also includes two parts: the fitting deviation and the instability caused by the number of parameters. The formula is as follows:

$$\text{BIC} = n \log(\text{MSE}) + m \log(n), \quad (17)$$

where

m is the number of parameters and n is the sample capacity. The smaller the BIC value, the better the fitting effect of the copula function.

3.3 Marginal distribution selection

In this study, the Poisson distribution (POISS), the general extreme value (GEV), the exponential distribution (EXP), the log-normal distribution (Ln2), and the Pearson type III distribution (P-III) were selected for the marginal distribution fitting (Poulin et al., 2007).

1) Poisson distribution

The Poisson distribution is a classical probability model that describes the frequency distribution of rare events. It is suitable for describing the number of random events per unit time. It can be regarded as the limit distribution of a binomial distribution.

The probability distribution of a random variable x obeying the Poisson distribution is

$$P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda} \quad (18)$$

According to the Poisson model, the probability of more than n rainstorms per year can be calculated as follows:

$$P(X \geq n) = 1 - e^{-\lambda} \sum_{k=0}^{n-1} \frac{\lambda^k}{k!} \quad (19)$$

2) Log-normal distribution

Suppose that the logarithm of the extreme value variable obeys a normal distribution; then if X is an extreme value variable, its logarithm $y = \ln X$ obeys the normal distribution, that is,

$$f(y) = \frac{1}{\sqrt{2\pi}\delta_y} e^{-(y-m_y)^2/2\delta_y^2}, \quad (20)$$

where m_y and δ_y can be estimated using the following formula (Wong et al., 2010):

$$m_y = \frac{1}{n} \sum_{i=1}^n \ln x_i, \delta_y^2 = \frac{1}{n} \sum_{i=1}^n (\ln x_i - m_y)^2 \quad (21)$$

3) General extreme value distribution

The general extreme value distribution has a good fitting effect for extreme hydrological events, and its probability density is (Sraj et al., 2015)

$$f(x) = \frac{1}{a} \left[1 + k \left(\frac{x-\mu}{a} \right) \right]^{-\frac{1}{k}-1} e^{-\left[1 + k \left(\frac{x-\mu}{a} \right) \right]^{1/k}}, \quad (22)$$

where a is the scale parameter, k is the shape parameter, and μ is the position parameter.

The distribution function is

$$F(x) = \exp \left\{ - \left[1 + k \left(\frac{x-\mu}{a} \right) \right]^{-\frac{1}{k}} \right\} \quad (23)$$

4) Exponential distribution

The probability density of the exponential distribution is

$$f(x) = \alpha e^{-\alpha(x-\beta)} \quad (\beta \leq x < +\infty), \quad (24)$$

where α and β are the distribution parameters, α is the degree of dispersion, and β is the lower limit of the distribution.

The distribution function is

$$F(x) = \int_x^{+\infty} f(x) dx = e^{-\alpha(x-\beta)}. \quad (25)$$

5) Pearson type III distribution

The P-III distribution has the following probability density function:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} (x-b)^{\alpha-1} e^{-\beta(x-b)} \quad (b \leq x), \quad (26)$$

where α is the shape parameter ($\alpha > 0$), β is the scale parameter ($\beta > 0$), and b is the position parameter.

The estimation of the fitting effect of the edge distribution parameters can be conducted using the

maximum likelihood method as follows:

$$L(\theta) = \prod_{i=1}^n F(x_i; \theta) = \prod_{i=1}^n \frac{\partial F(x_i; \theta)}{\partial x_i},$$

$$\ln[L(\theta)] = \sum_{i=1}^n \ln [F(x_i; \theta)], \quad (27)$$

where $L(\theta)$ is the likelihood function, $F(X_i; \theta)$ is the marginal distribution density function, and θ is the parameter to be estimated. Furthermore, the Kolmogorov-Smirnov 2 (K-S2) test was used to test the goodness of fit of each marginal distribution and to determine the probability distribution function suitable for each index (Subimal et al., 2010).

4 Results and discussion

4.1 Variation trends of the rainfall indexes

4.1.1 Variation trend of the contribution rate of the rainfall in the flood season

The contribution rate of the rainfall in the flood season in the Taihu Lake Basin and seven subregions were tested using the Mann-Kendall trend test. The results are shown in Table 4. As can be seen from the table, the contribution rate of the rainfall in the flood season in the Taihu Lake Basin and each subregion is decreasing; the value in West Zhejiang Province passes the 95% significance test, exhibiting a significant decreasing trend. Due to the high level of urbanization in the Taihu Lake Basin, the rain island effect is more significant, and the evaporation is also greatly increased. The Taihu Lake Basin is located in the subtropical monsoon region, which has more solar radiation and more evaporation. This reduces the rainfall in the flood season.

Table 4. Trend test table for the contribution rate of the rainfall in the flood season

Hydrological Region	Z	$Z_{\alpha/2}$	Trend	Significance
Zhexi	-2.13	1.96	decreasing	*
Huxi	-0.13	1.96	decreasing	/
Wuchengxiyu	-0.19	1.96	decreasing	/
Yangchengdianmao	-0.77	1.96	decreasing	/
Pudongpuxi	-1.07	1.96	decreasing	/
Hangjiahu	-1.45	1.96	decreasing	/
Taihu	-0.98	1.96	decreasing	/
Taihu Lake Basin	-1.1	1.96	decreasing	/

4.1.2 Variation trend of the rainstorm frequency

The Mann-Kendall trend test method was used to test the trend of the rainstorm frequency in the Taihu Basin and seven subregions. The test results are shown in Table 5. As can be seen from

the table, except for in the Taihu region, the rainstorm frequency in each subregion in the Taihu Lake Basin exhibits an increasing trend, and the rainstorm frequencies in the Wuchengxiyu and Yangchengdianmao regions pass the 95% significance test, exhibiting significant increasing trends. The Z value of the trend test for the entire Taihu Basin is 1.13, which indicates that the rainstorm frequency in the entire Taihu Basin exhibits an insignificant increasing trend.

As can be seen from the above analysis, the rainstorm frequency in each subregion in the Taihu Basin is increasing, except for in the Huxi region. This may be due to the rain island effect caused by urbanization, which increases the frequency of the rainstorm events in urban areas. However, the evaporation in the Taihu Basin is higher in the east than in the west. The Huxi region is located in the western part of the Taihu Basin, which makes the evaporation in this region lower, thus reducing the condensation nodules and water vapor condensation over the cities in this region, resulting in a lower rainstorm frequency.

Table 5. Trend test table for rainstorm frequency

Hydrological Region	Z	$Z_{\alpha/2}$	Trend	Significance
Zhexi	0.62	1.96	increasing	/
Huxi	0.42	1.96	decreasing	/
Wuchengxiyu	1.98	1.96	increasing	*
Yangchengdianmao	2.16	1.96	increasing	*
Pudongpuxi	1.34	1.96	increasing	/
Hangjiahu	0.68	1.96	increasing	/
Taihu	-0.52	1.96	increasing	/
Taihu Lake Basin	1.13	1.96	increasing	/

4.1.3 Variation trend of the maximum daily rainfall

The Mann-Kendall trend test method was used to test the trend of the maximum daily rainfall in the Taihu Basin and seven subregions. The test results are shown in Table 6. As can be seen from the table, the maximum daily rainfall in the Huxi and Pudongpuxi regions shows decreasing trend, while that in the other districts exhibits an increase

ing trend. Among them, the maximum daily rainfall in the Yangchengdianmao region passes the 95% significance test, exhibiting a significant increasing trend. The Z value of the trend test for the entire Taihu Basin is 0.41, which indicates that the annual rainfall in the entire Taihu Basin exhibits an insignificant increasing trend.

As can be seen from the above analysis, the maximum daily rainfall in the Taihu Lake Basin has obvious regional differences. Because the maximum daily rainfall mostly occurs in the flood season, the maximum daily rainfall can be caused by convective rain, and there are regional differences in the surface water entering the atmosphere through evaporation in the different areas,

so the maximum daily rainfall in each subregion is different.

Table 6. Trend test table for the maximum daily rainfall

Hydrological Region	Z	$Z_{\alpha/2}$	Trend	Significance
Zhexi	0.53	1.96	increasing	/
Huxi	-0.3	1.96	decreasing	/
Wuchengxiyu	0.58	1.96	increasing	/
Yangchengdianmao	2.24	1.96	increasing	*
Pudongpuxi	-0.2	1.96	decreasing	/
Hangjiahu	1.02	1.96	increasing	/
Taihu	0.38	1.96	increasing	/
Taihu Lake Basin	0.41	1.96	increasing	/

4.2 Mutation point detection

As can be seen from Figure 2, at the end of the 1950s, UF(k) and UB(k) intersected, and then, UF(k) exceeded the critical straight line, indicating that the rainfall in the next year had a significant mutation at the end of the 1950s. After that, the UF(k) and UB(k) curves intersected many times after the 1990s, and the rainfall exhibited a certain volatility.

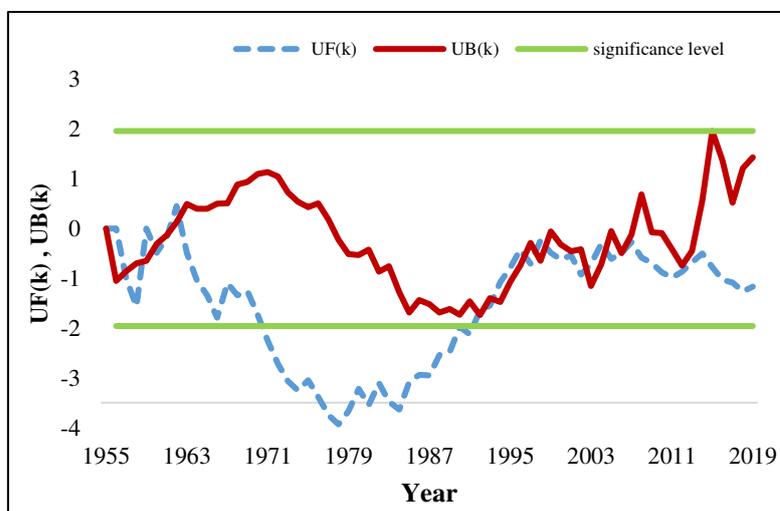


Fig. 2. Mann-Kendall mutation analysis results of the annual rainfall in the Taihu Basin

4.3 Construction of the joint distribution function

4.3.1 Measurement of the dependence and marginal distribution

The Pearson' γ , Spearman' ρ , and Kendall' τ values were used to measure the dependence between several pairs of variables: the contribution rate of the rainfall in the flood season and the maximum daily rainfall, the contribution rate of the rainfall in the flood season and the rainstorm frequency,

and the maximum daily rainfall and the rainstorm frequency. The calculation results are shown in Table 7.

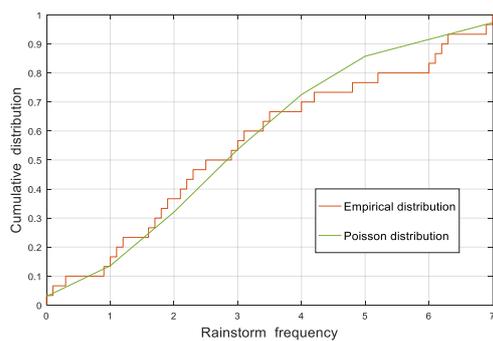
As can be seen, there are close correlations between the contribution rate of the rainfall in the flood season (C), the maximum daily rainfall (M), and the rainstorm frequency (F), so joint probability characteristic analysis of the flood variables can be carried out.

Table 7. Dependence between variables

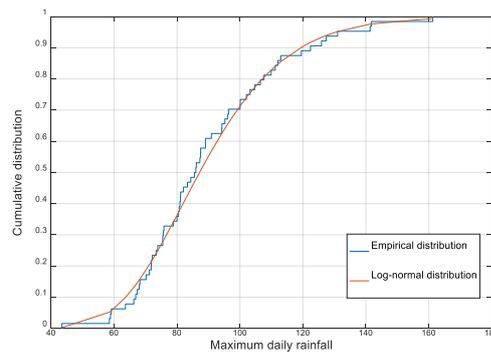
variable	Pearson' γ	Spearman' ρ	Kendall' τ
C&M	0.865	0.921	0.762
C&F	0.798	0.893	0.73
M&F	0.975	0.974	0.921

4.3.2 Selection of the marginal distribution

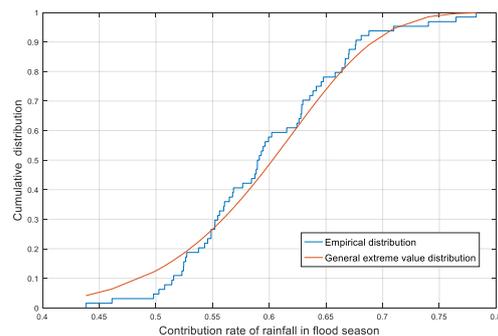
Appropriate distribution functions and parameters were used to draw the probability distribution curve for each rainfall factors (Figure 3). The tops of the cumulative probability distribution curves of the three characteristic factors have a gentle change trend, that is, the occurrence of 6–7 rainstorms in a year, a maximum daily rainfall of 140–160 mm, and rainfall contribution rates of 0.75–0.8 in the flood season are rare. The cumulative probability curve of the maximum daily rainfall is the steepest, and the rainfall is mainly concentrated between 60 and 120 mm, with a cumulative probability of 0.9. That is, the maximum daily rainfall for nearly 58 years is within the range of the 65 year research period of 1954–2018. The variation trend of the top of the cumulative probability curve of the contribution rate of the rainfall in the flood season is relatively consistent with that of the middle part, and the distribution is relatively uniform, which means there is no obvious concentrated range.



(a)



(b)



(c)

Fig. 3. Univariate distributions of the rainfall indexes: (a) the rainstorm frequency, (b) the maximum rainfall, (c) the contribution rate of the rainfall in the flood season

The maximum likelihood function was used for the parameter estimation of the probability distribution functions of the rainfall indexes, and the Kolmogorov-Smirnov test was used to evaluate whether the fitting function can pass the 0.05 significance test. If the P value is greater than the significance level, the greater the P value, the more likely it is to pass the test.

Table 8 shows that the EXP distribution only passes the test for the contribution rate of the rainfall in the flood season, and the difference between the P value and the 0.05 significance level is small, indicating that the fitting effect is weak. The Poisson distribution is suitable for fitting the rainstorm frequency. The fitting effect of the GEV distribution for the contribution rate of the rainfall in the flood season is particularly obvious. Both the P-III distribution and the Ln2 distribution pass the fitting test for the two indexes, and the fitting accuracy of the Ln2 distribution for the maximum daily rainfall is relatively high.

Table 8. K-S test of the marginal function fitting

Marginal function	C	M	F
POISS	0.0000	0.0000	0.6589*
Ln2	0.0245	0.5246*	0.0000
EXP	0.0842*	0.0235	0.0000
GEV	0.8245*	0.1107*	0.0006
P III	0.1458*	0.3589*	0.0000

Table 9 shows the values of the marginal distribution parameters of the characteristic rainfall variables.

Table 9. Parameters and P values of the marginal distribution

variable	distribution	parameter	P value
C	GEV	$\sigma=0.6285, \mu=8.7852$	0.83
M	Ln2	$\sigma=0.2449, \mu=4.4676$	0.86
F	POISS	$\lambda=4$	0.78

Table 9 shows the fitting curves for the marginal distributions of the three variables.

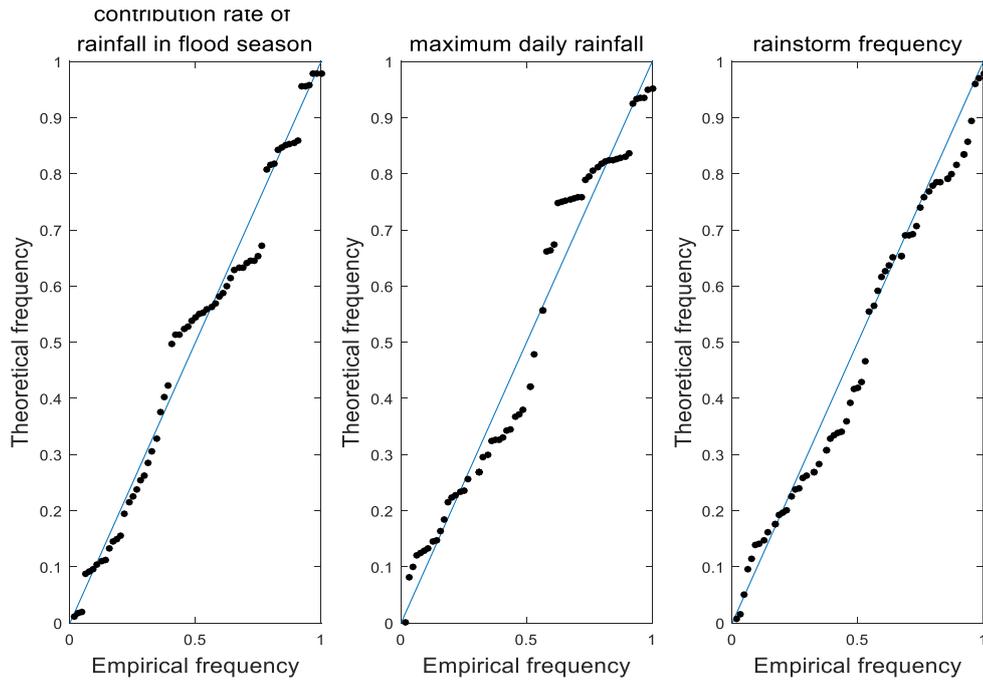


Fig. 4. Fitting curves for the marginal distributions of the three variables

4.3.3 Parameter estimation and goodness of fit test

The parameters were obtained through parameter estimation, and the theoretical frequency values of the Clayton copula, Gumbel copula, and Frank copula functions were calculated. The Copula functions with the best fitting degree were determined by comparing their the RMSE, AIC, and BIC values, which were used to construct the joint distribution of the rainfall indexes. The calculated parameter values and the goodness of fit evaluation values are shown in Table 10. As can be seen from the table, the RMSE, AIC, and BIC indexes of the Frank copula function are the smallest, so the Frank copula function was selected to build the two-dimensional joint distribution model.

Table 10. Index values for evaluating the goodness of fit

	Parameter	RMSE	AIC	BIC
Gumbel	8.43	0.0433	-112.45	-116.42
Clayton	5.94	0.0946	-118.37	-108.27
Frank	12.37	0.0372	-121.45	-119.83

4.3.4 Joint probability analysis of the rainfall indexes

The Frank copula function was used to calculate the joint distribution function for the three rainfall factors. The conditional risk probability was calculated according to the formula. The risk probability of the other two variables was calculated based on the contribution rate of the rainfall in the flood season. Figure 5 shows the conditional probability distribution map and its isoline map for the rainfall frequency and the maximum daily rainfall when the contribution rate of the rainfall in the flood season is equal to 0.65. According to Figure 5(a), the joint probability increases as the

rainfall frequency and maximum daily rainfall increase. From the joint probability contour map in Figure 5(b), the probability values of the different index value combinations can be seen more clearly. The greater the rainfall frequency and the maximum daily rainfall, the greater the joint probability value. When the rainfall frequency remains unchanged, the probability increases with increasing maximum daily rainfall. Similarly, when the maximum daily rainfall remains unchanged, the joint probability increases with increasing rainfall frequency.

Figure 6 shows the conditional probability distribution map and its isoline map for the rainfall frequency and maximum daily rainfall when the contribution rate of the rainfall in the flood season is greater than or equal to 0.65.

As can be seen from Figures 5 and 6, compared with the probability distribution under the condition that a single variable is equal to a certain designed value, the probability diagram under the corresponding condition that a single variable is less than or equal to a certain design value is taller and fatter, indicating that the probability of the risk of being greater than or equal to a certain design value under the latter condition is higher.

There is a strong correlation among the three rainfall variables. The use of the joint distribution of the multiple variables to describe the relationship between the different rainfall eigenvalues can more comprehensively describe the rainfall events and improve the analysis accuracy.

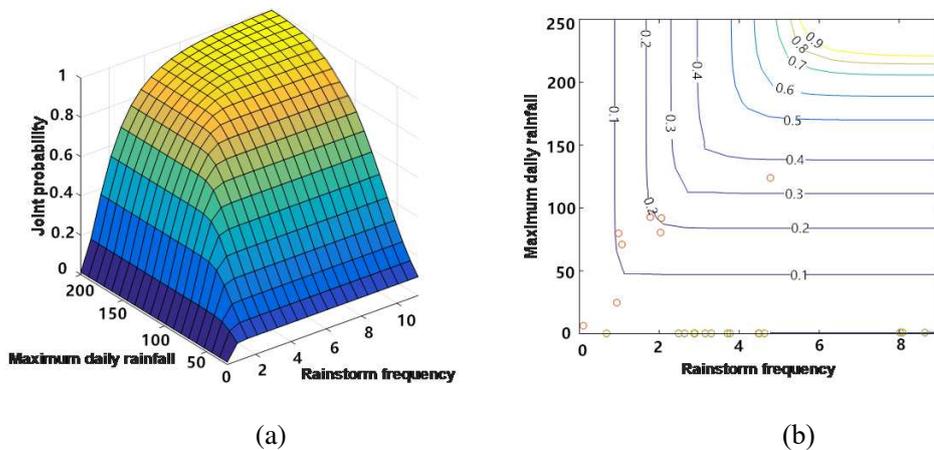


Fig. 5. Conditional risk probability when the contribution rate of the rainfall in the flood season is equal to 0.65

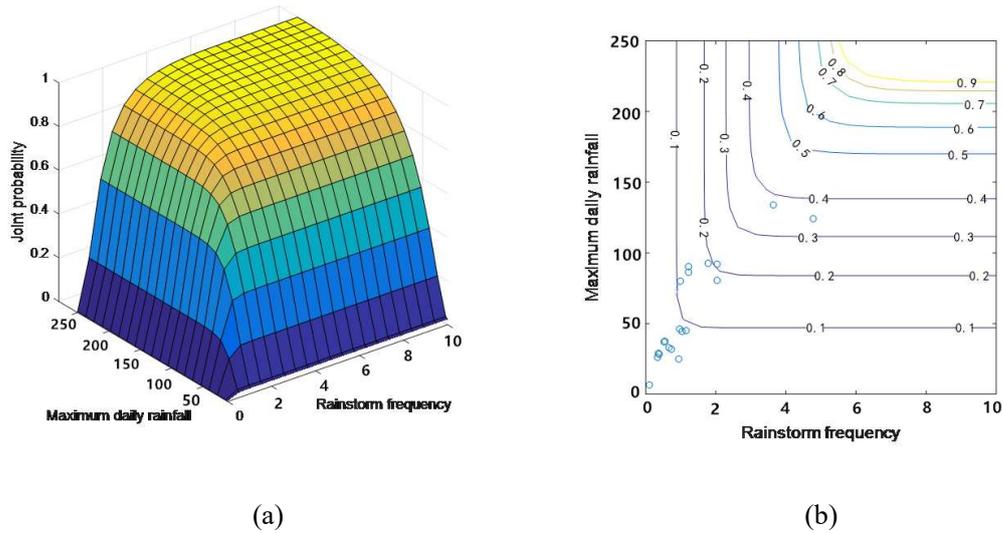


Fig. 6. Conditional risk probability when the contribution rate of the rainfall in the flood season is greater than 0.65

The Frank copula formula was used to solve the Frank copula joint probability for the study area, and then, a three-dimensional rainfall characteristic map with the rainfall factors as the x-y-z axes and the copula joint probability of the study area as the distribution value C was drawn using the surf code in MATLAB.

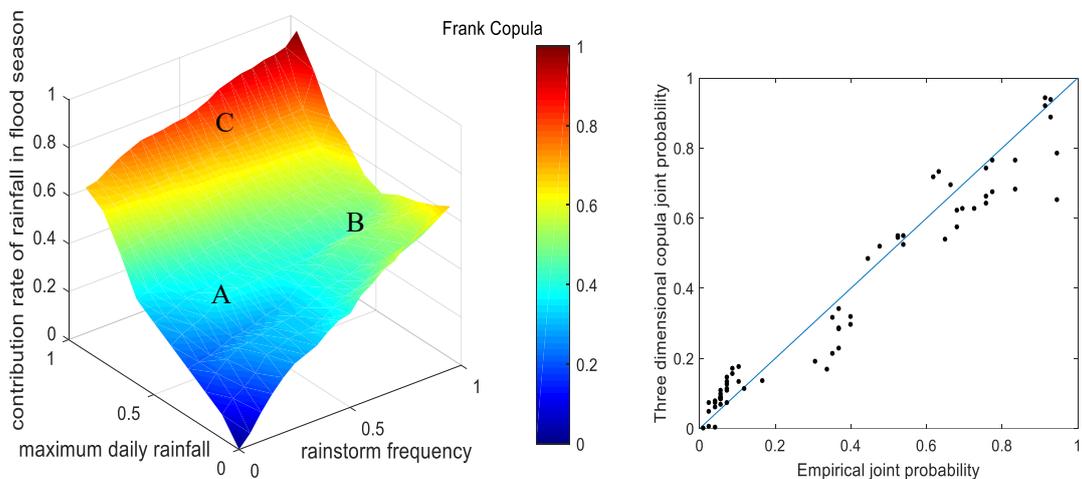


Fig. 7. Cumulative probability of the three dimensional Copula function

As can be seen from Figure 7, the Frank copula function's three-dimensional joint probability is concentrated near the 45 degree straight line, with little difference from the actual distribution, and the fitting result is reliable. When the joint cumulative probability is between 0.8 and 0.9, the deviation is relatively large and the fitting error is relatively obvious. The copula joint probability does not exhibit a synchronous change trend with the rainfall factors. For example, when X-Y-Z is (0.5,0.5, 0.5), if all of the factors have the same influence on the copula value, the copula value should be 0.5. As can be seen from the figure, the corresponding value at this point is less than 0.5. The asynchronous effect is the embodiment of the influence of each factor on the copula value in

the different intervals.

In area A, which has a Frank copula value of 0–0.38, the rainstorm frequency and maximum daily rainfall are larger in the figure. That is, when the copula value is 0.38, the values of X and Y are greater than 0.5, while C is less than 0.2, demonstrating an obvious lag and indicating that in the years with lower rainstorm frequencies, lower maximum daily rainfall values, and lower contribution rates of the rainfall in the flood season, the contribution rate of the rainfall in the flood season plays a relatively important role in the comprehensive characteristics of the rainfall.

Area A obviously protrudes into the region with a high rainstorm frequency. That is, when Y and Z (the maximum daily rainfall and the contribution rate of the rainfall in the flood season) change uniformly, a relatively large X (the rainfall frequency) is required to make the copula probability value correspond to the Y and Z values, which indicates that the maximum daily rainfall and the contribution rate of the rainfall in the flood season are the main factors affecting the copula joint probability. This also shows that if the maximum daily rainfall and the contribution rate of the rainfall in the flood season are small in a certain year, the rainfall frequency has little influence on the comprehensive characteristics of the rainfall in this year.

In area B, i.e., $X \in (0.78-1)$, $Y \in (0.14-0.46)$, $X \in (0.48-0.65)$, the variation trend of the rainstorm characteristics is that when the X value increases positively, the Y value decreases negatively, and the Z value is distributed in the range of (0.48–0.65) and approaches 0.65. Area B is triangular, with the top on the right, and the Z value and X value approach the right apex of the triangle. The corresponding actual rainfall characteristics are as follows. In a year with less rainfall, the rainfall frequency is relatively large, but the maximum daily rainfall is relatively small, which results in a small amount of rainfall compared with the historical records but accounts for a large proportion of the rainfall in this year. Therefore, the contribution rate of the rainfall in the flood season is relatively high.

As can be seen from area C, the cumulative distribution of the copula increases with as the cumulative probability value of the marginal distribution of each rainfall factor increases, which conforms to the objective fact of the cumulative probability distribution.

Area C shows that the rainstorm frequency (X) and maximum daily rainfall (Y) are relatively large, and the contribution rate of the rainfall in the flood season (Z) is relatively consistent with the copula joint probability value. This indicates that when the copula function is 0.74–0.96, the rainstorm frequency and maximum daily rainfall have the opposite effect; when one of the two factors has a weak effect on the copula function value, the other has a relatively strong effect.

5 Conclusions

Based on the indivisibility of the rainfall factors, in this study we used the copula joint distribution function to discuss the probability distribution characteristics of the rainfall factors in the study area. The conclusions of this study are as follows.

(1) In terms of the change trend, the annual rainfall and the contribution rate of the rainfall in the flood season in the Taihu Lake Basin exhibited a non-significant decreasing trend. The abrupt change point of the annual rainfall in the Taihu Lake Basin appeared at the end of the 1950s, and it fluctuated after the 1990s.

(2) With regard to the selection of the marginal distribution, the Poisson distribution is suitable for the probability fitting of the rainstorm frequency, the log-normal distribution is suitable for simulating the maximum daily rainfall, and the general extreme value distribution is more accurate for fitting the contribution rate of the rainfall in the flood season.

(3) A three-dimensional joint distribution of the rainfall indexes was established, and the Frank copula function was found to be more suitable for the probability calculation based on the RMSE tests.

(4) By calculating the conditional risk probability, the probability of the risk under the condition of being greater than or equal to a certain design value is larger.

(5) The three-dimensional joint copula function reflects more of the information than the single variable distribution function and highlights the relationship between the factors, which is a part of rainfall research that must be explored further.

Acknowledgement

This study was supported by the Fundamental Research Funds for the Central Universities (Granted:2014B16814), Jiangsu Provincial Water Resources Technology Project (Granted:2017006, 2015003), and China Scholarship Council (201806715037). The authors also would like to thank the editor and reviewers for their crucial comments

Conflict of Interest: The authors declared that they have no conflicts of interest to this work.

We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Funding Statement: The research was supported by Fundamental Research Funds for the Central Universities (Granted:2014B16814), Jiangsu Provincial Water Resources Technology Project (Granted:2017006, 2015003), and China Scholarship Council (201806715037).

Author's Contribution: Methodology, formal analysis, original draft preparation, Manqiu Hao and Gao Cheng; Validation, Jian Chen.

Availability of data and material: Please contact author for data requests.

Code availability: Not applicable.

Ethics approval: This study does not violate ethics.

Consent to participate: Not applicable.

Consent for publication: Not applicable.

References

- Alcamo, J, Thomas, H.T, Rösch, T. World Water in 2025-Global Modeling Scenarios for the World Commission on Water for the 21st Century; Report A0002; Center for Environmental Systems Research: University of Kassel,Kassel, Germany, 2000.
- Brunner M, Viviroli D, Sikorska A, Vannier A, Favre A-C, Seibert J.Flood type specific construction of synthetic design hydrographs. [J].Water Resource Research.2017,53:1390–1406.
- Cailin Wang, Xuehui Ren, Ying li. Analysis of extreme rainfall characteristics in low mountain areas based on three-dimensional copulas—taking Kuandian County as an example.[J].Theoretical and Applied Climatology.2017,128:169-179.
- Duhan,D., Pandey,A., 2013. Statistical analysis of longterm spatial and temporal trends of rainfall during 1901–2002 at Madhya Pradesh, India. Atmos. Res. 122, 136–149.
- Fischer S, Schumann A, Schulte M .Characterisation of seasonal flood types according to timescales in mixed probability distributions.[J]. Journal of Hydrology.2016,539:38–56.
- Ganguli, P., Janga Reddy, M. Risk assessment of droughts in gujarat using bivariate copulas.[J]. Water Resources Management. 2012,26 (11):3301–3327.
- Ganguli P,Reddy M J. Probabilistic assessment of flood risks using trivariate copulas [J]. Theoretical and Applied Climatology,2013,111(1-2) : 341-360.
- Gao Junguo, Zhai Jinjin, Dong Sheng.Joint statistical analysis of maximum daily rainfall and its occurrence date at Qingdao port[J].Periodical of Ocean University of China,2019,49(7):125-132.
- Janga Reddy, M., Ganguli, P. Bivariate flood frequency analysis of upper godavari river flows using archimedean copulas.[J] .Water Resources Management.2012,26 (14):3995–4018.
- K H Hamed. Exact distribution of the Mann-Kendall trend test statistic for persistent data[J].Hydrology,2009(1):86-94.
- Lab D. Recent advances in wavelet analyses: Part 1. A review of concepts[J].Journal of Hydrology,2005,314(1-4):275-288.
- LI H S,WANG D,SINGH V P,et al. Non-stationary frequency analysis of annual extreme rainfall volume and intensity using Archimedean Copulas:a case study in Eastern China[J].Journal of Hydrology,2019,571: 114-131.
- LIU Y R,LI Y P,MA Y,et al.Development of a Bayesian-copula-based frequency analysis method for hydrological risk assessment: the Naryn River in Central Asia[J].Journal of Hydrology, 2020, 580:124349.
- Lu J,Zhang Y H,Gao F,et al. Spatial and temporal characteristics of extreme rainfall in Sanjiang Plain in recent 40 years[J]. Research of Soil and Water Conservation,2019,26(2):272-282.
- Yang Yin, Li Yangying. Characteristics of short-time heavy rainfall in Hedong area of Gansu based on copula function.[J].Meteorological Monthly,2019,45(5):632-640.
- Salvadori G. Bivariate return periods via 2-copulas[J].Statistical Methodology, 2004,12(1):129-144.

Shagufta Akbari, Manne Janga Reddy. Non-stationarity analysis of flood flows using copula based change-point detection method: Application to case study of Godavari river basin. [J].Science of the Total Environment,2019,43:1-10.

Pedro-Monzonis, M.; Solera, A.; Ferrer, J.; Estrela, T.; Paredes-Arquiola, J. A review of water scarcity and drought indexes in water resources planning and management. [J].Journal of Hydrology. 2015, 527: 482–493.

Poulin A, Huard D, Favre A, Pugin S. Importance of tail dependence in bivariate frequency analysis.[J]. Journal of Hydrology Engineering.2007,12(4):394–403.

Shahid Latif,Firuz Mustafa. Copula-based multivariate flood probability construction: a review.[J]. Arabian Journal of Geosciences.2020,13:132-157.

Qian, L,Wang, H,Dang, S, Wang, C. Modelling bivariate extreme rainfall distribution for data-scarce regions using Gumbel-Hougaard copula with maximum entropy estimation. [J].Hydrological Processes. 2017, 32.

Sraj M, Bezak N, Brilly M. Bivariate flood frequency analysis using the copula function: a case study of the Litija station on the Sava River. [J].Hydrology Process 2015, 29(2):225–238.

Subimal G. Modelling bivariate rainfall distribution and generating bivariate correlated rainfall data in neighbouring meteorological subdivisions using copula [J].Hydrological Process,2010,24:3558-3567.

Valeriya Filipova,Deborah Lawrence,Harald Klempe. Effect of catchment properties and flood generation regime on copula selection for bivariate flood frequency analysis.[J].Acta Geophysica. 2018,66:791-806.

Sayemuzzanman, M., Jha, M.K., 2014. Seasonal and annual rainfall time series trend analysis in North Carolina, United States. Atmos. Res. 137, 183–194.

Wong G, Lambert MF,Leonard M,Metcalf AV. Drought analysis using trivariate copulas conditional on climatic states. [J].Journal of Hydrologic Engineering . 2010,15(2):129–141.

YIN Jia-bao, GUO Sheng-lian, HE Shao-kun, et al. A Copula-based Analysis of Projected Climate Changes to Bivariate Flood Quantiles [J]. Journal of Hydrology, 2018, 566: 23-42.

Yu, Xiong, L. Derivation of low flow distribution functions using copulas. [J]. Journal of Hydrology. 2014, 508: 273–288.

Zhang D. Vine copulas and applications to the European Union sovereign debt analysis. [J]. International Review of Financial Analysis. 2014, 36: 46–56.