Analysis model of flood spatiotemporal change characteristics based on k-means clustering algorithm

Qianyi Chen[1,2], Wei Zhang[1,2*], Weixiang Huang[1,2], Peng Liu3, Shan Li[1,2]

[1]Electric Power Research Institute of Guangxi Power Grid Co., Ltd, Nanning, 530023, China

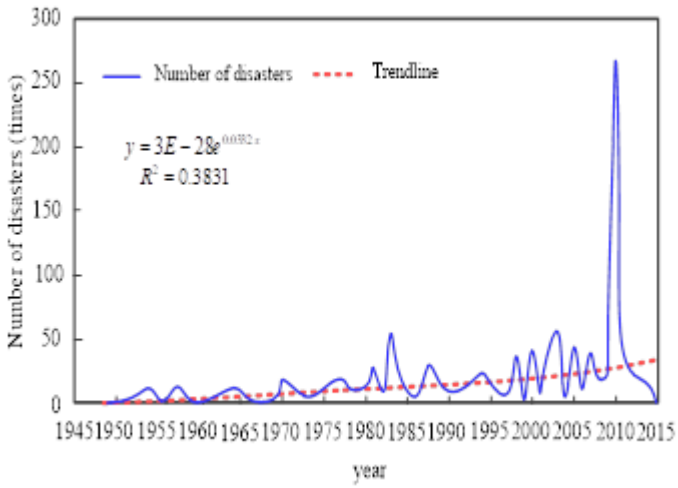
[2]Innovation Department, Guangxi Power Grid Equipment Monitoring and Diagnosis Engineering Technology Research Center, Nanning, 530023, China

[3]Guangxi Power Grid Co., Ltd, Nanning, 530023, China

[*]Corresponding author: Wei Zhang

E-mail: zhangwei_wz@outlook.com

GRAPHICAL ABSTRACT



ABSTRACT

In order to analyze the spatiotemporal change characteristics of regional flood disaster, analyze the main change characteristics and hydrological response of precipitation distribution in Guangxi, and improve the comprehensive utilization efficiency of water resources, this paper introduces the k-means clustering algorithm to design an analysis model of spatiotemporal change characteristics of flood disaster. First of all, obtain the flood data in Guangxi and the observation data of the national meteorological station within 5 km, complete the collection of the basic data of mountain

flood disasters. Based on the collected data, an analysis was conducted on the spatial and temporal distribution of flash floods in the Guangxi region. Secondly, the Kriging spatial interpolation method was used to analyze the spatial distribution of precipitation data in Guangxi. The Mann-Kendall trend test was then employed to examine the trend of precipitation-related statistical parameters over time. Additionally, wavelet theory was applied to analyze the time series of annual precipitation and precipitation with different durations in Guangxi. Subsequently, the k-means clustering algorithm was introduced to construct a model for analyzing the spatiotemporal characteristics of flood changes, determining the concentration and duration of precipitation in different years in the region. Finally, analyze the spatiotemporal change characteristics of flood events in different seasons under various indicators, and realize the analysis of flood spatiotemporal change characteristics. The research results indicate that the Frank Copula function fits the best correlation between annual precipitation and temperature, and can better characterize the correlation between the two. The Frank Copula function has the best fitting effect on the correlation between precipitation and temperature in autumn in Guilin, summer in Nanning, and summer and winter in Beihai. In Guangxi Zhuang Autonomous Region, the annual precipitation shows a gradually decreasing trend, especially at R and P stations. In summary, the Frank Copula function can effectively characterize the correlation and trend of precipitation and temperature in different seasons and regions of Guangxi.

Keywords: Mountain flood disaster, Mann-Kendall trend test, k-means algorithm, Kriging space interpolation method, Wavelet theory

1. Introduction

Flooding is caused by the rapid rise of water levels in rivers, lakes and reservoirs, causing dikes and dams to overflow or breach, resulting in the influx of water into populated areas, causing significant damage. In addition to causing severe harm to agriculture, flood disasters can also result in losses to industries, as well as to human lives and properties. They are one of the top ten natural disasters that pose a threat to human survival (Yao et al., 2022). Over the past 50 years in China, although there hasn't been a clear trend of extreme precipitation, the number of rainy days has decreased while the average precipitation intensity has increased. Taking precipitation in the Yangtze River Basin as an example, the significant increase in its concentration in both time and space is a major reason for the exacerbation of surrounding water and drought disasters (Schroers and Martin, 2022). For example, since the 1990s, the spatiotemporal concentration of precipitation in the Yangtze River basin has increased significantly. This change in the spatiotemporal distribution of precipitation has weakened the natural flood control mechanism of the Yangtze River basin, not only causing an increase in the number of occasional floods and shallow disasters, but also causing an increase in the number of drought disasters in the non-precipitation concentration period. This is consistent with the fact that the flood and drought disasters in the Yangtze River basin have been increasing in recent years (Yang et al., 2022). It is shown from some facts that the spatiotemporal concentration of precipitation is closely related to serious drought and flood events. So whether the spatiotemporal concentration of precipitation will further strengthen with global warming, and whether its change will lead to more serious drought and flood disasters, and other issues have aroused widespread concern (Wang et al., 2022; Darand, 2021). Therefore, many domestic scholars have conducted in-depth research on the spatiotemporal variability of precipitation in China.

Liu (2022) and others based on 3DGIS flood spatial-temporal situation deduction and dispatching simulation research, established flood spatial-temporal situation process model based on situation deduction and other theories, combined with flood characteristics and influence factors, proposed flood spatial-temporal situation visualization and dynamic deduction method under 3DGIS scenario,

and built flood spatial-temporal situation dispatching simulation program framework. Taking the flood of the Yangtze River basin in 2020 as an example, with the assistance of the dispatch calculation results for the flood process, this program framework is utilized to carry out the analysis, deduction and visualization application of various spatiotemporal situations, and visually display the flood process in the three-dimensional virtual scene. The results show that the flood spatiotemporal situation deduction and dispatching simulation method based on 3DGIS can clearly present the flood spatiotemporal situation, assist in understanding the flood evolution trend and risk, intuitively reproduce the whole flood spatiotemporal process, and provide auxiliary support for flood control dispatching consultation. Zhang et al. (2022) studied the flood simulation of small watershed in mountainous areas based on time-space information. Taking an experimental area in Zhejiang Province as the research object, flood simulation was carried out by analyzing the small watershed covered by it, and the daily model hydrological model was used for actual prediction and prediction. Meanwhile, to verify the accuracy of the daily hydrological model, the results of the HEC-HMS model are compared and analyzed. The experimental results showed that the daily model hydrological model comprehensively considered the water balance. The energy balance can effectively adapt to the small basin flood simulation in the study area, and the results are better than the HEC-HMS model, which proves that the model is applicable and provides theoretical and practical support for the subsequent basin flood simulation. Jia et al. (2022) based on Sentinel-1 SAR data and Chaohu Lake basin flood spatial-temporal dynamic change monitoring research, taking Chaohu Lake basin as the experimental area, based on Sentinel-1 Synthetic Aperture Radar (SAR) image, constructed a flood inundation identification method combining spectral relationship and threshold segmentation, which was applied to Google Earth Engine platform, obtained the flood spatial-temporal pattern of Chaohu Lake basin from 2015 to 2020, combined with land use data, The impact of flood on farmland and residential areas represented by construction land in Chaohu Lake basin is analyzed. Compared with the single-band threshold method and the simple index method, the accuracy of this method is improved by 3%~7%, and it can quickly apply remote

sensing data to extract the flood inundation range of the basin over the years. The research shows that SAR satellite data has good applicability in monitoring the flood in Chaohu Lake basin, helping to grasp the extent of damage to farmland and rural settlements caused by the flood, and is crucial for formulating relevant planning strategies in the future and strengthening the flood control in the basin and rural areas to ensure personnel and food security. However, the above methods are not effective in analyzing the spatiotemporal characteristics of flood disasters.

In view of the above problems, this paper proposes a flood spatiotemporal change feature analysis model based on k-means clustering algorithm.

2. Multi-time scale analysis of precipitation series in flood areas

2.1 Data acquisition of mountain flood disaster in Guangxi

Guangxi is located in the coastal area of South China, with a total area of 236700 km^2, accounting for 2.5% of the total land area of the country. The Tropic of Cancer, which runs through the central part of the country, belongs to the subtropical monsoon climate zone. It has abundant rainfall, with an annual average rainfall of about 1537 mm, which is about three times the national average. There are many rivers, and there are about 1450 rivers with a rainfall collection area of more than 50 square kilometers. Surrounded by three high mountains and facing the sea on one side, the area of hills accounts for 77%. The terrain is high in the north and low in the south, and the central part is low in the basin, forming a dustpan-like terrain. Due to complex topographical and geological conditions and karst development, it is easy to form local strong rainfall and strong convection in hilly areas, and the impact of human activities, resulting in frequent occurrence of mountain flood disasters in Guangxi, which is one of the most serious natural disasters in Guangxi.

The data used in this paper are mainly from the relevant departments of the national government, the research institute of the Chinese Academy of Sciences and the official data sharing website of Shaanxi Province. See Table 1 for details of data types and types.

Meteorological data mainly collected daily rainfall data from 34 meteorological stations in Guangxi Zhuang Autonomous Region from 1990 to 2015. The data is from the China

Meteorological Data Network (http://data.cma.cn). Slope data, soil type data and vegetation cover data (NDVI index) are directly provided by Shaanxi Provincial Hydrological Survey Center. Elevation (DEM) data is derived from geospatial data cloud (http://www.gscloud.cn/) DEM image with a resolution of 30 m obtained from. Land use data is the distribution of land use types in Guangxi Zhuang Autonomous Region in 2015 obtained from the National Basic Geographic Information Center. Population and socio-economic data were obtained from the Guangxi Zhuang Autonomous Region Statistical Annals, and population density data was extracted with the help of GIS tools.

Table 1 Main data sources

Data name	Data content	Data sources
Meteorological data	Daily rainfall data of 34 meteorological stations in Guangxi Province from 1990 to 2015	China Meteorological Data Network
Landform data	Slope data of Guangxi Zhuang Autonomous Region	Guangxi Zhuang Autonomous Region Hydrological Survey Center
Soil type data	Soil type data of Guangxi Zhuang Autonomous Region	Guangxi Zhuang Autonomous Region Hydrological Survey Center
DEM data	30m resolution DEM image	Geospatial data cloud
Vegetation cover data	NDVI value of vegetation index in Guangxi Zhuang Autonomous Region	Hydrological Survey Center of Guangxi Zhuang Autonomous Region
Land use data	Land use distribution map of Guangxi Zhuang Autonomous Region in 2015	National Basic Geographic Information Center
Demographic data	Using ArcGIS technology to extract population density from the population data of Guangxi Zhuang Autonomous Region in 2015	Statistical Yearbook 2016
Socio-economic data	GDP data of Guangxi Zhuang Autonomous Region in 2015	Statistical Yearbook 2016

In this regard, the collection of basic data of mountain torrent disasters is completed, and the temporal and spatial distribution of mountain torrent disasters in Guangxi is analyzed based on the collected data.

2.2 Study on temporal and spatial distribution of mountain flood disaster

Based on the historical flood investigation reports of various districts and counties in Guangxi Zhuang Autonomous Region, the mountain flood disaster situation in the region from 1949 to 2015,

spanning 67 years, was compiled. There were more than 1,300 flash floods during this period, with an average of approximately 19 per year. The annual frequency of mountain flood disasters and the changing trend of disasters over the years are illustrated in Figure 1.

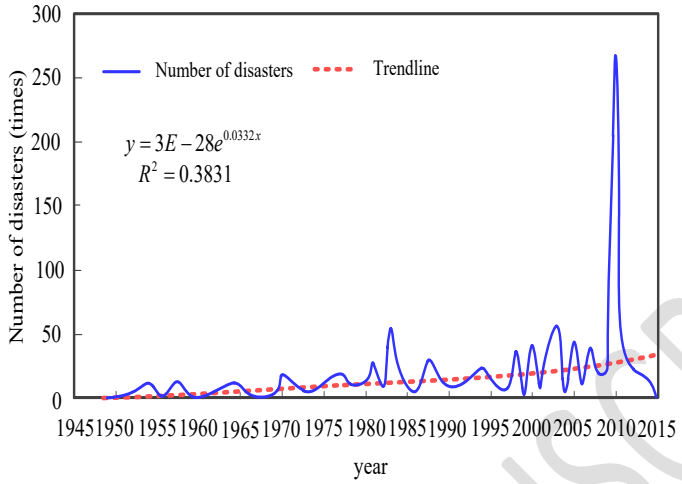


Figure 1. Inter-annual changes of mountain flood disasters

It can be seen from the figure that the number of mountain flood disasters in Guangxi Zhuang Autonomous Region generally presents a trend of stable annual growth. Through multiple item fitting of the annual disaster frequency during the study period, it is found that the periodic change of mountain flood disasters is characterized by an upward trend of fluctuations year by year. According to the records of rainfall data over the years in the study area, there were 7 years with abnormal values of disaster number in these 67 years, most of which had rainstorm records. For example, 2010 was the year of the highest concentration of flash floods, with severe disasters occurring in many places. 122 villages and communities in M District were affected, involving a population of 227.83 million people (including 5947 emergency transfers), 3215 houses collapsed, 2.11km of dykes were destroyed, and 15031.5 mu of crops were affected. At the same time, the flood destroyed many public facilities such as highway bridges and convenience bridges, resulting in high cliff collapse and other disasters, leading to serious casualties and losses.

According to the above analysis results of spatiotemporal characteristics, the spatial distribution data of precipitation in Guangxi is analyzed with Kriging spatial interpolation method.

2.3 Kriging space interpolation method

Spatial interpolation is a method to predict the data of unknown points in the region from known points by interpolation or extrapolation based on some data function relationship or theory. According to different mathematical principles, spatial interpolation is usually divided into two categories: deterministic interpolation and geostatistical interpolation. According to the research, the ordinary Kriging interpolation method in geostatistical interpolation has a good effect in rainfall interpolation calculation. Therefore, when drawing the spatial distribution of precipitation related statistics in Guangxi, the ordinary Kriging interpolation method (Hou et al., 2022) is used to determine the precipitation prediction model in Guangxi, and the expression is:

$$Z(X_o) = \sum_{i=1}^{n} \lambda_i Z(x_i) \tag{1}$$

Where, $Z(x_i)$ is the measured value of annual precipitation, $Z(X_o)$ is the predicted value of annual precipitation, n is the number of stations, λ_i is the weight, and λ_i is determined by Kriging equation group:

$$\begin{cases} \sum_{i=1}^{n} \lambda_i C(X_i, X_j) - u = C(X_i, X_o) \\ \sum_{i=1}^{n} \lambda_i = 1 \end{cases} \tag{2}$$

Where, $C(X_i, X_j)$ is the covariance between the annual precipitation samples, $C(X_i, X_o)$ is the covariance between the annual precipitation measured value and the samples, and u is the Lagrange multiplier (Lee et al., 2022). The analysis model of precipitation in Guangxi is obtained as follows:

$$r(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (Z(S_i) - Z(S_i + h))^2 \tag{3}$$

Where, $Z(S_i + h)$ is the semi-variance, $2N(h)$ is the data pair, and $Z(S_i)$ is the measured value.

2.4 Mann-Kendall non-parametric trend test

The Mann-Kendall trend test method (Khavse and Chaudhary, 2022) can be used to examine relatively stationary sequences. Therefore, in this study, the method was employed to analyze the trend of precipitation-related statistical parameters in Guangxi over time.

When Mann-Kendall method is used for mutation test, the meaning of each statistic is different. First, construct rank sequence S_k for precipitation time series X. Secondly, calculate statistics UF_k, where $E(S_k)$ represents the mean of precipitation rank series S_k and $VAR(S_k)$ represents the variance of S_k.

$$UF_k = \frac{S_k - E(S_k)}{\sqrt{VAR(S_k)}} \tag{4}$$

Thirdly, after the time sequence of precipitation sequence X is reversed, the statistic UB_k is constructed by the same method, then $UF_k = -UB_k$. Finally, draw UF and UB statistics curves. Current statistics. At $UF > 0$, the sequence showed an upward trend. When the statistic is $UF < 0$, the sequence shows a downward trend. If the UF or UB curve exceeds the critical line of significant value, it indicates that the sequence has a significant upward or downward trend, and the time period corresponding to the exceeded part is defined as the mutation time period. If curves UF and UB intersect with two adjacent boundaries at a certain point, it indicates that this point is the start time of mutation.

2.5 Multi-time scale analysis of precipitation series based on wavelet analysis

Wavelet analysis is a time-frequency localization analysis method based on Fourier transform proposed by Morlet in the 1980s (Hirai et al., 2022). This paper will use wavelet theory to analyze the time series of annual precipitation and precipitation of different durations in Guangxi. Its basic function is Morlet wavelet of complex wavelet (Belaid, 2021), which is defined as follows:

$$\Psi(t) = e^{iw_0 t} e^{-t^2/2} \tag{5}$$

And wavelet function $\Psi(t) = L^2(R)$ meets:

$$\int_{-\infty}^{+\infty} \Psi(t)dt = 0 \tag{6}$$

At the same time, function $\Psi(t)$ can form a family of function systems:

$$\Psi_{a,b}(t) = |a|^{-1/2} \overline{\Psi}(\frac{t-b}{a}) \tag{7}$$

For signal $f(t) \in L^2(R)$, wavelet transform is defined as (Li et al., 2021):

$$W_f(a,b) = |a|^{-1/2} \int_{-\infty}^{+\infty} f(t) \overline{\Psi}(\frac{t-b}{a}) dt \tag{8}$$

Where, $W_f(a,b)$ is the binary function of the continuous wavelet transform coefficient, and its value changes with the change of a and b, representing the wavelet change coefficient. Parameter a is the scaling factor of wavelet period length. Parameter b is the time factor of time translation. The wavelet change coefficient map is a two-dimensional contour map about $W_f(a,b)$, reflecting the time-frequency change characteristics of the signal. The wavelet coefficient modulus square of the scale factor is integrated in the time domain to obtain the wavelet variance (Chen et al., 2021), so as to determine the multiple time scales of the precipitation series. The expression is:

$$var(a) = \int_{-\infty}^{+\infty} |W_f(a,b)|^2 db \tag{9}$$

The change process of multiple time scales $var(a)$ of precipitation series with wavelet coefficient modulus can be used to analyze the change and development trend of precipitation structure in different time scales.

3.Determining the concentration and duration of precipitation in flood-prone areas

3.1 Analysis model of flood spatiotemporal change characteristics based on k-means clustering algorithm

Given sample observation data matrix

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \tag{10}$$

Among them, each row of X is a sample (or observation), and each column is n observation values of a variable, that is, X is a matrix composed of observation values of p variables of n samples $(x_1, x_2, ..., x_n)$.

Fuzzy clustering is to divide n samples into c categories and $2 \le c \le n$, and record $V = \{v_1, v_2, ..., v_c\}$ as the cluster center of c categories, of which $v_i = \{v_{i1}, v_{i2}, ..., v_{ip}\}(i = 1, 2, ..., c)$. In k-means clustering, each sample is not strictly assigned to a particular cluster but rather has a certain degree of membership in a given cluster. Let u_{ik} indicate that the k sample x_k belongs to category i, where $0 \le u_{ik} \le 1$, $\sum_{l=1}^{c} u_{ik} = 1$. Define the analysis model of flood spatiotemporal change characteristics:

$$J(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}^{m} d_{ik}^{2} \tag{11}$$

Where, $U = (u_{ik})_{c \times n}$ is the membership matrix, $d_{ik} = \|x_k - v_i\|$. Obviously, $J(U,V)$ represents the sum of the weighted square distance from the sample to the cluster center in each category, and the weight is the m power of the membership degree of sample x_k belonging to the i category. The clustering criterion of k-means clustering algorithm is to find $J(U,V)$, so that $J(U,V)$ can get the minimum value. The specific steps of k-means clustering algorithm are as follows:

(1) Determine the number of classes c, power index $m > 1$ and initial membership matrix $U^{(0)} = (u_{ik}^{(0)})$. The usual method is to take the uniformly distributed random number on [0,1] to determine the initial membership matrix $U^{(0)}$. Let $l = 1$ represent the first iteration.

(2) Calculate the cluster center $V^{(l)}$ in step l by the following formula:

$$V_i^{(l)} = \frac{\sum_{k=1}^{n} (u_{ik}^{(l-1)})^m X_k}{\sum_{k=1}^{n} (u_{ik}^{(l-1)})^m} \tag{12}$$

(3) Modify the membership matrix $U^{(l)}$ and calculate the model value $J^{(l)}$ of flood spatiotemporal change characteristics analysis.

$$u_{ik}^{(l)} = \frac{1}{\sum_{j=1}^{c} (d_{ik}^{(l)} / d_{jk}^{(l)})^{\frac{2}{m-1}}}, i = 1,2,...,c; k = 1,2,...,n \tag{13}$$

Among them, $d_{ik}^{(l)} = \left\| x_k - V_i^{(l)} \right\|$.

(4) For the given membership degree termination tolerance $\varepsilon_u > 0$ (or the flood spatiotemporal change characteristic analysis model termination tolerance $\varepsilon_J > 0$, or the maximum iteration step size $L_{\max}$), when $\max\{(\left| u_{ik}^{(I)} - u_{ik}^{(I-1)} \right|)\} < \varepsilon_u$ ($l > 1, \left| J^{(l)} - J^{(l-1)} \right| < \varepsilon_J$, or $l \geq L_{\max}$), stop the iteration, otherwise, turn to (2).

After the iteration of the above steps, the final membership matrix U and cluster center V can be obtained to minimize the value of the flood spatial-temporal change feature analysis model $J(U,V)$. The attribution of all samples can be determined according to the values of the elements in the final membership matrix U. When $u_{jk} = \max\{u_{ik}\}$, sample x_k can be classified as category j.

3.2 Determination of precipitation concentration and concentration

3.2.1 Daily precipitation concentration index (CID)

The data below 1.0 mm of daily rainfall for a specific year is excluded and grouped every 1 mm. The grouping stops when the maximum daily rainfall for that year is reached. Then, the rainfall days (N_i) and total days (P_i) are calculated for each group, resulting in the correlation between the cumulative rainfall day percentage (Y) and the cumulative rainfall percentage (X). The exponential relationship between X and Y can be described by the following equation:

$$Y = aX \exp(bX) \tag{14}$$

Where a and b are constants obtained through least squares fitting. Based on equation (14), the distribution of the indicator function is defined as the Lorenz curve. The concentration of

rainfall is represented by the region A, enclosed by the Lorenz curve and the quadrant bisectors. In cases where the value of A is larger, the trend is closer to the x-axis, indicating a higher concentration of daily rainfall.

The region A can be calculated using the following approach:

$$A = 10000/2 - \int_0^{100} ax\exp(bx)dx \tag{15}$$

Then CID is:

$$CID = 2A/10000 \tag{16}$$

The CID ranges from 0 to 1, and the smaller the value of CID, the more uniform the distribution of daily rainfall.

3.2.2 Monthly precipitation concentration index (CIM)

The calculation method of the monthly precipitation concentration index is expressed as follows:

$$CIM = 100\sum_{i=1}^{12} p_i^2 / (\sum_{i=1}^{12} p_i)^2 \tag{17}$$

Where p_i represents the precipitation in month i. In a given year. When CIM is lower than 10, it indicates that the precipitation in each month of that year is relatively evenly distributed around the average level. When CIM is between 11 and 20, it suggests that the precipitation in that year exhibits seasonal variations. When CIM exceeds 20, it signifies an abnormally concentrated distribution of rainfall throughout the year, with significant differences in monthly precipitation patterns.

3.2.3 Decadal precipitation concentration (PCD) and precipitation concentration period (PCP)

Precipitation concentration duration (PCD) and precipitation concentration period (PCP) are methods based on vector analysis to describe the temporal distribution characteristics of rainfall in a certain region. PCD is used to describe the spatiotemporal distribution characteristics of rainfall in the area. The value of PCD ranges from 0 to 1 during rainfall events. The closer the PCD value is to 1, the more intense the precipitation is. Within a certain time range, when the PCD value approaches 0, it indicates a relatively smooth spatial distribution of rainfall. PCP is used to describe

the timing of the maximum rainfall occurrence within a year. The calculation method is as follows:

$$PCD = \sqrt{R_{x_i}^2 + R_{y_i}^2} / R_i \tag{18}$$

$$PCP = \arctan(R_{xi} / R_{yi}) \tag{19}$$

Where: $R_{xi} = \sum_{j=1}^{n} r_{ij} \times \sin\theta_j$, $R_{yi} = \sum_{j=1}^{n} r_{ij} \times \cos\theta_j$, and R_i represent the total precipitation at different observation stations in the region. r_{ij} refers to the precipitation for each decadal period investigated. θ_j represents the azimuth corresponding to each decadal period within the specified time cycle. i represents the year, and j represents the sequence of decadal periods within the study period.

4. Measurement and analysis of correlation between precipitation and temperature based on Copula function

Based on precipitation and temperature data from various stations in Guangxi from 1971 to 2012, this paper selected Beihai, Nanning and Guilin as typical regions from the coast to the interior and measured the correlation between precipitation and temperature. It is mainly analyzed from two aspects. On the one hand, the correlation between seasonal precipitation and temperature is analyzed based on data on precipitation and temperature at different seasons over the past 42 years. On the other hand, based on the data of accumulated monthly average precipitation and temperature for many years, the correlation between precipitation and temperature within the year is analyzed.

4.1 Correlation analysis of seasonal precipitation and temperature

Based on the time series data of precipitation and temperature for the four seasons from 1971 to 2012 at the three stations of Guilin, Nanning, and Beihai, the Pearson correlation coefficient, Kendall rank correlation coefficient, and Spearman rank correlation coefficient were calculated. The results are shown in Table 2. The results show that there is a certain correlation between precipitation and temperature in Guilin, Nanning and Beihai in summer, only in Guilin in autumn and only in Beihai in winter. Therefore, Copula function can be used to measure the correlation of precipitation and temperature series with correlation. Take Guilin summer as an example to

introduce the correlation measurement analysis of precipitation and temperature based on Copula function.

Table 2 Statistics of correlation between precipitation and temperature in Guilin, Nanning and Beihai□

Area	□Classification	Spring		Summer		Autumn		Winter	
Guilin	Pearson	-0.147	uncorrelated	-0.505**	correlated	-0.412**	correlated	-0.085	uncorrelated
	Kendall	-0.1	uncorrelated	-0.366**		-0.285**		-0.085	uncorrelated
	Spearman	-0.134	uncorrelated	-0.515**		-0.4051**		-0.137	uncorrelated
Nanning	Pearson	0.126	uncorrelated	-0.506**	correlated	-0.127	uncorrelated	-0.269	uncorrelated
	Kendall	0.043	uncorrelated	-0.278**		-0.043	uncorrelated	-0.164	
	Spearman	0.079	uncorrelated	-0.422**		-0.068	uncorrelated	-0.231	
The North Sea	Pearson	-0.155.	uncorrelated	-0.592**	correlated	-0.122	uncorrelated	-0.450**	correlated
	Kendall	-0.120		-0.447**		-0.058		-0.261*	
	Spearman	-0.187		-0.645**		-0.096		-0.397**	

**and * * are at 0.01 and 0.05 levels (double tail) respectively, with significant correlation.

4.2 Selection and Parameter Estimation of Copula Function

Calculate the Kendall and Spearman rank correlation coefficients corresponding to the Copula function of summer precipitation and temperature in Guilin, as shown in Table 3.

Table 3 Copula function correlation measure of summer precipitation and temperature in Guilin

Function type	τ	ρ_s
Normal Copula	-0.338	-0.489
t-Copula	-0.394	-0.563
Gumbel Copula	1.36E-06	2.05E-06
Clayton Copula	7.25E-07	1.09E-06
Frank Copula	-0.383	-0.55

Comparing the correlation measures of the Copula function in Table 2 with the Kendall and Spearman rank correlation coefficients in Table 3, we can see that the correlation measures of the normal Copula and Frank Copula functions are the closest, while the effects of other Copula

functions are poor.

After determining the edge distribution and the preliminary determination of the Copula function, you can make further selection of the Copula function, and then determine the parameters of the Copula function:

(1) Substitute the empirical distribution function into equation (15), and estimate the parameters of the normal Copula and Frank Copula functions with the semi-parametric method.

(2) Calculate the Kendall rank phase relationship value between annual average precipitation and annual average temperature, and then calculate the parameters corresponding to the three Copula functions according to the relationship formula with parameters in Table 3.

(3) According to the parameters calculated by the two methods, calculate the joint distribution of annual precipitation and annual temperature of the empirical Copula function and the two Copula functions.

(4) The AIC value and OLS value under the two parameters are calculated. The AIC value and OLS value are the smallest, indicating that the corresponding Copula function has the best fitting effect.

The copula function fitting of the joint distribution of precipitation and temperature in Guilin is shown in Table 4. The best copula function and parameter estimation method are shown in bold.

Table 4 Copula function fitting of the joint distribution of summer precipitation and temperature in Guilin

Evaluating indicator	Copula function type	Semiparametric estimation method	Kendall rank correlation coefficient estimation method
OLS	Normal Copula	0.24	0.23
	Frank Copula	0.23	0.22
AIC	Normal Copula	-119.36	-120.15
	Frank Copula	-120.84	-120.38

It can be seen from Table 4 that the OLS value and AIC value corresponding to Frank Copula function are the smallest, 0.23 and -120.84 respectively, so the goodness of fit of Frank Copula function is the best, indicating that this function can better represent the correlation between annual

precipitation and temperature.

Similarly, the copula function fitting of the joint distribution of precipitation and temperature in autumn in Guilin, summer in Nanning, and summer and autumn in the North Sea can be calculated, as shown in Table 5-7.

Table 5 Copula function fitting of the joint distribution of autumn precipitation and temperature in Guilin

evaluating indicator	Copula function type	Semiparametric estimation method	Kendall rank correlation coefficient estimation method
OLS	Normal Copula	0.25	0.25
	Frank Copula	0.24	0.24
AIC	Normal Copula	-114.84	-115.07
	Frank Copula	-116.36	-115.13

Table 6 Copula function fitting of the joint distribution of summer precipitation and temperature in Nanning

evaluating indicator	Copula function type	Semiparametric estimation method	Kendall rank correlation coefficient estimation method
OLS	Normal Copula	0.24	0.24
	Frank Copula	0.23	0.24
AIC	Normal Copula	-118.71	-117.01
	Frank Copula	-119.50	-117.29

Table 7 Copula function fitting of the joint distribution of summer and winter precipitation and temperature in the North Sea

season	evaluating indicator	Copula function type	Semiparametric estimation method	Kendall rank correlation coefficient estimation method
summer	OLS	Normal Copula	0.20	0.20
		Frank Copula	0.19	0.19
	AIC	Normal Copula	-132.16	-117.01
		Frank Copula	-135.36	-117.29
winter	OLS	Normal Copula	0.24	0.25
		Frank Copula	0.23	0.24
	AIC	Normal Copula	-116.99	-114.84
		Frank Copula	-118.28	-116.02

It can be seen from Table 5~7 that the Frank Copula function is the best fit, which shows that the Frank Copula function can better represent the correlation between precipitation and temperature in autumn in Guilin, summer in Nanning and summer and winter in Beihai.

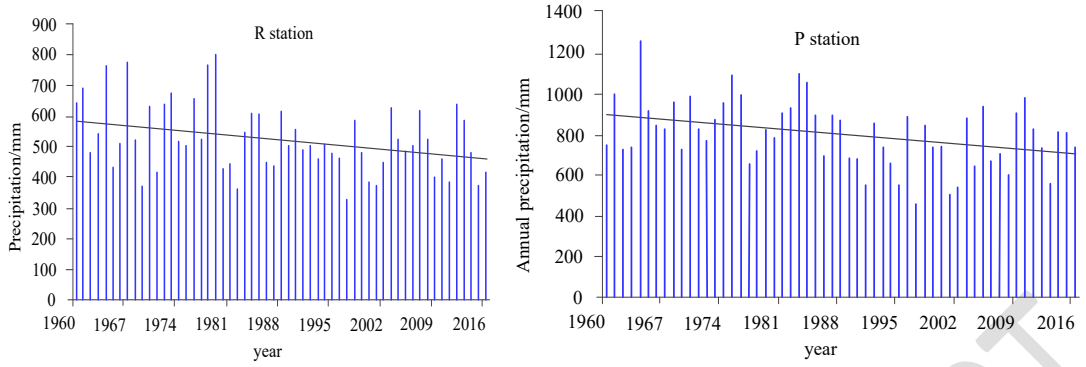
From the above analysis, it can be seen that the precipitation and temperature in Guangxi Beihai, Nanning and Guilin have a certain correlation in summer, while the precipitation and temperature in Guilin have a certain correlation in autumn, and the precipitation and temperature in Beihai in winter. Their linear correlation coefficient and rank correlation coefficient are shown in Table 2. The correlation degree and structure of precipitation and temperature in different seasons in the three regions can be measured by Frank Copula function. At the same time, it also shows that the semi-parameter estimation method is better.

4.3 The spatial and temporal changes in annual precipitation

The Mann-Kendall trend test method was used to statistically analyze the precipitation data from 19 monitoring stations in the Guangxi Zhuang Autonomous Region from 1960 to 2016 (Table 8). The spatial and temporal distribution characteristics of precipitation were studied. When the absolute value of Z is greater than or equal to 1.96, it indicates that this method has achieved a confidence level of 95%.

Table 8 Annual precipitation trend of Guangxi Zhuang Autonomous Region (1960~2016)

site	Z	trend	site	Z	trend
R	-2.086	Significant decrease	X	-1.397	-
M	-0.420	-	S	-0.929	-
N	-1.921	-	O	-1.150	-
P	-2.705	Significant decrease	J	-0.613	-
K	-1.053	-	E	-1.755	-
Q	-0.324	-	T	-1.549	-
Y	-1.108	-	L	-0.792	-
B	-1.218	-	F	-1.147	-
D	-1.425	-	V	-0.620	-
W	-0.778	-	-	-	-



(a) R station (b) P station

Figure 2. Annual precipitation trend

According to the results in Table 8, the Z value of the annual precipitation series in Guangxi Zhuang Autonomous Region is less than 0, that is, it has an insignificant downward trend, and the Z value at R and P stations is less than -1.96, which is a significant downward trend. In the entire Guangxi Zhuang Autonomous Region, the annual precipitation has a gradual downward trend, which is the same as the research results of Lei Jiang Group. According to Figure 2 (a), the maximum annual precipitation of R station was 801.5 mm, which occurred in 1979, and the minimum was 326.6 mm, which occurred in 1997. It can be seen from Figure 2 (b) that the maximum annual precipitation of P station was 1262.3 mm, which occurred in 1964, and the minimum was 465.3 mm, which occurred in 1997.

5. Conclusion

In this paper, k-means clustering algorithm is introduced to design an analysis model of flood spatiotemporal change characteristics. Obtain the flood disaster data in Guangxi and the observation data of nearby national meteorological stations, analyze the temporal and spatial distribution of mountain flood disasters in Guangxi, use the Kriging spatial interpolation method to calculate the spatial distribution data of precipitation in Guangxi, conduct the trend test on the time series of precipitation related statistics in Guangxi according to the Mann-Kendall trend test method, and build an analysis model based on Copula function that integrates the precipitation concentration index and the temporal and spatial change characteristics of flood disasters, Some achievements

have been made. The k-means clustering algorithm is introduced to build the analysis model of flood spatial-temporal change characteristics, determine the precipitation concentration degree and concentration period in different years in the region, and analyze the spatial-temporal change characteristics of flood events in different seasons under various indicators.

With the advancement of technology and the improvement of data acquisition methods, the data quality and coverage of spatiotemporal changes in flood disasters will be improved. In the future, multi-source data fusion will be considered, such as remote sensing data, meteorological data, geographic information data, etc. By utilizing the advantages of different data sources, we can more accurately capture and analyze the spatiotemporal changes of floods, thereby improving flood warning and disaster response capabilities.

Conflict of Interest

The authors report there are no competing interests to declare.

Data Availability Statement

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

References

Belaid A. The processing of resonances excited by gear faults using continuous wavelet transform with adaptive complex Morlet wavelet and sparsity measurement. Measurement, 2021, 180(1): 162-169.

Chen K., Zhang X., Liu Y. An improved denoise method based on EEMD and optimal wavelet threshold for model building of OPAX. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2021, 235(14): 3530-3544.

Darand M. Projected changes in extreme precipitation events over Iran in the 21st century based on CMIP5 models. Climate research, 2021, 26(2): 82-96s.

Hirai T. Structure of unipotent orbits and Fourier transform of unipotent orbital integrals for semisimple Lie groups. Lectures on Harmonic Analysis on Lie Groups & Related Topics, 2022, 16(9): 153-159.

Hou D., Wang L., Yan J. Vibration analysis of a strain gradient plate model via a mesh-free moving Kriging Interpolation Method. Engineering Analysis with Boundary Elements, 2022, 135(18): 156-166.

Jia J., Ma J., Shen M., et al. Research on monitoring the spatiotemporal dynamic change of flood in Chaohu Lake basin based on Sentinel-1 SAR data. Remote Sensing Technology and Application, 2022, 37 (01): 173-185.

Khavse R., Chaudhary J. L. Trend assessment in climate variable by Mann Kendall test of Bastar district of Chhattisgarh. Mausam: Journal of the Meteorological Department of India, 2022, 32(1): 73-78.

Lee H., Shin J., Lee J. Energy quadratization Runge-Kutta scheme for the conservative Allen-Cahn equation with a nonlocal Lagrange multiplier. Applied mathematics letters, 2022, 18(10): 132-145.

Li W., Xu W., Zhang T. Improvement of threshold denoising algorithm based on wavelet transform. Computer Simulation, 2021, 38 (06): 348-351+356.

Liu C., Zhang L., Fan Q., et al. Study on flood space-time situation deduction and dispatching simulation based on 3DGIS. Water conservancy and hydropower technology (Chinese and English), 2022, 53 (8): 8-15.

Schroers M., Martin E. Synoptic Connections and Impacts of 14-Day Extreme Precipitation Events in the United States. 2022, 12(32): 151-158.

Wang H., Wang H., Li H. Analysis of drainage efficiency under extreme precipitation events based on numerical simulation. Hydrological processes, 2022, 21(6): 36-38.

Yang P., Zhang S., Xia J., et al. Analysis of drought and flood alternation and its driving factors in the Yangtze River Basin under climate change. Atmospheric Research, 2022, 270(06): 106087-106092.

Yao R., Zhang S., Sun P., et al. Estimating the impact of urbanization on non-stationary models of extreme precipitation events in the Yangtze River Delta metropolitan region. Weather and Climate Extremes, 2022, 18(21): 16-25.

Zhang Q., Zheng C. Research on flood simulation of small watershed in mountainous areas based on spatiotemporal information. Electromechanical Technology of Hydropower Station, 2022, 45 (08): 85-87+91.