

Extreme rainfall forecasting based on wavelet analysis and M-K test

Zijian Wang *

School of Water Conservancy and Engineering Zhengzhou University Henan, China

* Corresponding Author Email: chunshuiruolanke@126.com

Abstract. In recent years, many places in China have encountered extremely rare heavy rainfall weather, and the north has also encountered rare blizzard weather, which poses a serious threat to the personal and property safety of local people. On July 20th, 2021, with the maximum hourly rainfall reaching 201.9 mm, Zhengzhou suffered the extremely heavy rainfall once in a thousand years. The 7.20 heavy rain in Zhengzhou caused many communities and roads in Zhengzhou to be flooded by the rain and thus drew attention to and recognized the extraordinary precipitation. In this paper, a correlation analysis is made on the interannual variation of precipitation characteristics in Zhengzhou area, and concrete quantitative analysis is made on flood events; In addition, the characteristics and losses of precipitation in Shaanxi and Zhengzhou are compared. Finally, put forward the long-term planning and construction of cities under the condition of extreme heavy rainfall in the future.

Keywords: Wavelet analysis; M-K test; Extreme rainfall.

1. Introduction

Quantitative analysis is an important method to analyze the amount and intensity of future rainfall in China. Based on the analysis of rainfall data of 3 hydrological stations in Zhengzhou and annual data in Wuhan and Shanxi, this paper establishes the correlation analysis and prediction model of rainfall characteristics, makes abrupt change analysis, periodic analysis and prediction of urban rainfall characteristics, and puts forward the long-term urban construction plan under extreme rainfall conditions.

Firstly, we quantitatively analyze collected data and calculate the annual average rainfall. Next, we set up a correlation analysis model, and make Pearson correlation analysis on the characteristics of rainfall, and conclude that the factors most closely related to rainfall are the highest temperature and the lowest temperature. Then, M-K test [1] and wavelet analysis [2] are used to analyze the abrupt change and periodicity of rainfall respectively, and it is concluded that the years with higher rainfall are 1973, 1976 and 1978.

Through M-K test, wavelet analysis and prediction analysis, it is concluded that the rainstorm characteristics in 2021 are the same in two areas. In this paper, the losses caused by Zhengzhou rainstorm and Shanxi rainstorm are quantitatively described and analyzed from many aspects, and the difference of rainfall between the two areas is obtained.

In the end, this paper takes Zhengzhou as an example, analyzes its main problems in long-term urban construction planning under the condition of extreme rainfall, and puts forward some suggestions such as improving the forecast level of extreme precipitation weather and strengthening the standards of urban infrastructure construction.

2. Model assumptions and notation

2.1. Model assumptions

In the establishment of the model, the influence of some special mutation values is not considered;
Interpolation fitting of hydrological monitoring data in missing years;
Select some key pilot cities to analyze, establish and forecast the model;

In the process of model analysis and prediction, the time series change factors of extreme rainfall are mainly considered, while some secondary factors are ignored.

2.2. Explanation of symbols

Explanation of symbols which in this paper is shown in TABLE 1.

Table 1. Explanation of symbols.

Symbols	Explanation
$\rho_s(X_i, X_j)$	Covariance between X_i and X_j sum of random variables
$\sigma(X_i)$	Standard deviation of X_i random variables
$\sigma(X_j)$	Standard deviation of X_j random variables
$W_f(a, b)$	Wavelet coefficient
$W_f(a)$	Wavelet variance
S	Test statistical variable of normal distribution
x_i, x_j	Two series with different distributions in the same sample
σ_s	Standard deviation
$sign$	Operational symbol
n	Total number of samples
Z	Test value of the sample
m	Cumulative number of i sample $x_i > x_j (1 \leq j \leq i)$
$E(c_k), \sigma(c_k)$	Mean and variance of c_k

3. Model construction and example analysis

3.1. Correlation analysis of precipitation

3.1.1. Problem analysis

In the process of correlation analysis of the annual variation characteristics of precipitation in Zhengzhou area, firstly, qualitative analysis and regular exploration are carried out on the collected data, and various factors related to rainfall (including maximum temperature, minimum temperature, sea level pressure, station pressure, wind speed, etc.) are preprocessed.

In order to further analyze the annual variation characteristics of precipitation, it is necessary to calculate the annual average precipitation in advance, so as to carry out correlation analysis and evaluation. In the process of correlation analysis and evaluation, we can evaluate and synthesize from two aspects: catastrophe analysis and periodic relationship analysis. Then, according to the specific situation of correlation coefficient, the closely related factors related to precipitation are further obtained, and these correlation factors are quantitatively analyzed and evaluated [3].

3.1.2. Selection of correlation coefficient type

In the existing literature related to correlation analysis, Pearson correlation coefficient, Spearman correlation coefficient, Kendall correlation coefficient etc., whose correlation coefficients are widely used, are introduced in the following list:

a) Pearson correlation coefficient

Pearson correlation coefficient is used for bivariate normal distribution data to measure whether two data sets are on the same line. Or used to measure the linear relationship between distance variables.

b) Spearman correlation coefficient

Spearman rank correlation is the most commonly used nonparametric correlation analysis (rank correlation). When two variables do not meet the assumption of bivariate normal distribution,

Spearman rank correlation should be used to describe the relationship between variables. At this time, the distribution pattern of scatter points on the scatter chart cannot fully describe the correlation between two variables, so it is generally unnecessary to draw the scatter chart.

c) Kendall correlation coefficient

Kendall correlation coefficient is a rank correlation coefficient, which is an index to reflect the consistency of classification variables. It can only be used when multiple columns of variables belong to orderly classification.

Through analysis and comparison, among the various factors related to precipitation, the factor variables are not simply nonparametric linear correlation, and the two variables do not conform to the assumption of bivariate normal distribution, so the linear relationship between the two variables cannot be described by using simple scatter chart. Therefore, we choose Spearman rank correlation coefficient for analysis and make correlation index graph.

3.1.3. Model establishment

a) Pearson correlation coefficient analysis

Pearson correlation coefficient can only reflect the linear relationship between random variables. When random variables meet the normal distribution, it can also describe the nonlinear relationship between random variables. However, when any random variable obeys the non-normal distribution, Pearson correlation coefficient cannot accurately describe the correlation between random variables, so Spearman rank correlation coefficient needs to be derived from $\rho_s(X_i, X_j)$.

To solve the problem that the relationship between precipitation and other factors belongs to non-normal distribution random variables, Spearman rank correlation coefficient can better describe the correlation [4]. This coefficient does not measure the correlation between the true values of random variables, but first arranges the samples of random variables from low to high, and then calculates their respective ranks $\rho_s(X_i, X_j)$.

Pearson correlation coefficient between sum of random variables can be expressed as:

$$\rho(X_i, X_j) = \frac{\text{cov}(X_i, X_j)}{\sigma(X_i)\sigma(X_j)} \quad (1)$$

Type.

$\rho_s(X_i, X_j)$: Covariance between X_i and X_j sum of random variables;

$\sigma(X_i)$: Standard deviation of X_i random variables;

$\sigma(X_j)$: Standard deviation of X_j random variables.

According to the data of each factor, N groups of sample data of random variables are available, and then the Pearson correlation coefficient between them can be expressed as $\hat{\rho}(X_i, X_j)$.

$$\hat{\rho}(X_i, X_j) = \frac{\sum_{n=1}^N (x_{in} - \bar{X}_i)(x_{jn} - \bar{X}_j)}{\sqrt{\sum_{n=1}^N (x_{in} - \bar{X}_i)^2} \sqrt{\sum_{n=1}^N (x_{jn} - \bar{X}_j)^2}} \quad (2)$$

Type.

$\bar{X}_i = \left(\sum_{n=1}^N x_{in} \right) / N$: Represents the average value of sum of random variables X_i in N groups of sample data;

$\bar{X}_j = \left(\sum_{n=1}^N x_{jn} \right) / N$: Represents the average value of sum of random variables X_j in N groups of sample data.

b) Wavelet analysis

Wavelet analysis is a new time-frequency localization analysis method developed on the basis of Fourier analysis. Using wavelet analysis to explore this topic can better solve the multi-scale, multi-level and multi-factor, and it can better simulate the geographical characteristics of rainfall distribution [5].

On the basis of synthesizing various factors, the wavelet transform function formula is obtained:

$$W_f(a, b) = \int_R f(x) \bar{\psi}\left(\frac{x-b}{a}\right) dx = \frac{1}{\sqrt{|a|}} \int_R f(x) \bar{\psi}\left(\frac{x-b}{a}\right) dx \quad (3)$$

Then calculate the wavelet variance.

$$W_f(a) = \int_R |W_f(a, b)|^2 db \quad (4)$$

Type.

$W_f(a, b)$: Wavelet coefficient

$W_f(a)$: Wavelet variance

Based on wavelet function and wavelet variance, it can better reflect the geographical distribution of rainfall with various factors.

In order to prevent the signal loss in the process of wavelet transform, the wavelet transform is inverted, and the wavelet inversion formula is:

$$f(x) = \frac{1}{C_\psi} \iint_{R \times R^+} W_f(a, b) \psi_{(a,b)}(x) \frac{da db}{a^2} \quad (5)$$

From the point of view of time distribution characteristics, wavelet transform is no longer used to reflect the changes at an accurate "time point", but to better reflect the changes of various factors in a certain "time period".

c) M-K analysis

Nonlinear test

In order to test the change trend of rainfall elements, the M-K nonlinear trend test method is used to test various meteorological elements such as temperature and air pressure. This method assumes that the time data series $(x_1, x_2, x_3, \dots, x_n)$ is an independent and random sample series with the same variable distribution, and the calculation equation of the test statistical variable S is as follows:

$$S = \sum_{i=2}^n \sum_{j=2}^{i-1} \text{sign}(x_i - x_j) \quad (6)$$

When

$$(x_i - x_j) \begin{cases} > 0 \\ = 0 \\ < 0 \end{cases}, \text{sign}(x_i - x_j) \begin{cases} -1 \\ 0 \\ 1 \end{cases} \quad (7)$$

$$\sigma_s = \sqrt{\frac{n(n-1)(2n+5)}{18}} \quad (8)$$

$$\text{When } S = \begin{cases} > 0 \\ = 0 \\ < 0 \end{cases}, Z = \begin{cases} \frac{S-1}{\sigma_s} \\ 0 \\ \frac{S+1}{\sigma_s} \end{cases} \quad (9)$$

Type.

S : Test statistical variable of normal distribution

x_i, x_j : Two series with different distributions in the same sample

σ_s : Standard deviation

sign —Operational symbol

n : Total number of samples

Z : Test value of the sample

Mutation analysis

Define a statistic c_k to analyze n independent sample data series $(x_1, x_2, x_3, \dots, x_n)$ with the same distribution of random variables. The calculation equations are as follows:

$$c_k = \sum_{i=1}^k m_i \quad (10)$$

$$E(c_k) = \frac{k(k-1)}{4} \quad (11)$$

$$\sigma(c_k) = \frac{k(k-1)(2k+5)}{72} \quad (12)$$

$$UF_k = \frac{(c_k - E(c_k))}{\sqrt{\sigma(c_k)}} \quad (13)$$

Type.

m : Cumulative number of i sample $x_i > x_j (1 \leq j \leq i)$

$E(c_k), \sigma(c_k)$: Mean and variance of c_k

UF_k : Standardization of c_k

3.1.4. Model solution

Evaluate the Pearson correlation coefficient of the collected data, and get the key elements closely related to precipitation. The correlation results are as follows.

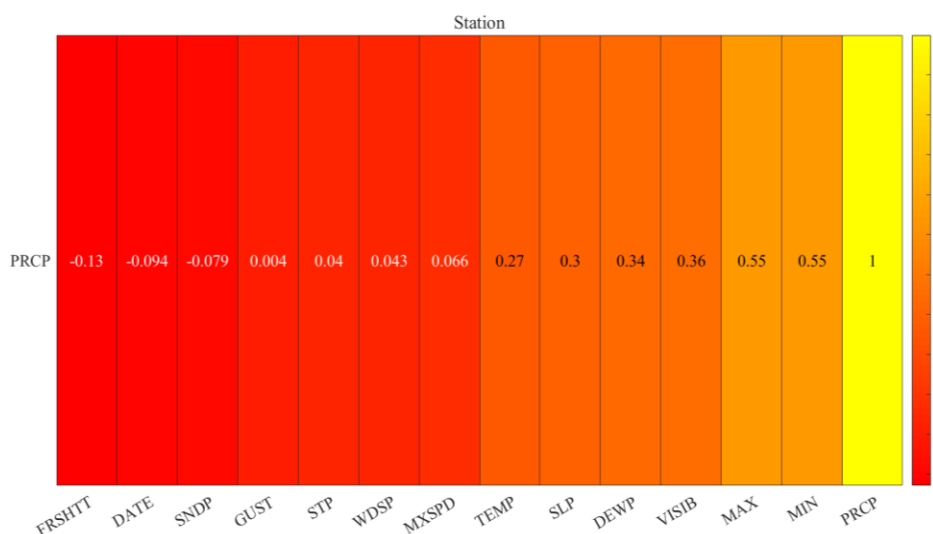


Fig. 1 Pearson correlation coefficient heat diagram.

Pearson correlation coefficient heat diagram is shown in Fig. 1. Through the analysis of correlation coefficient, it is concluded that the factors most closely related to precipitation are the highest temperature and the lowest temperature, namely MIN and MAX.

After that, the hydrological monitoring data of three hydrological stations are predicted and analyzed by M-K, and then the maximum rainfall year is observed.

Rainfall Station 1:

Observe and analyze the hydrological monitoring data of hydrological station 1, and analyze the change trend and abrupt change of rainfall. In the process of abrupt change detection, the order sequence of sequential time series and the order sequence of reverse time series are analyzed and calculated respectively. Finally, given the significance level, the time regions of abrupt changes of two abrupt time series curves are found.

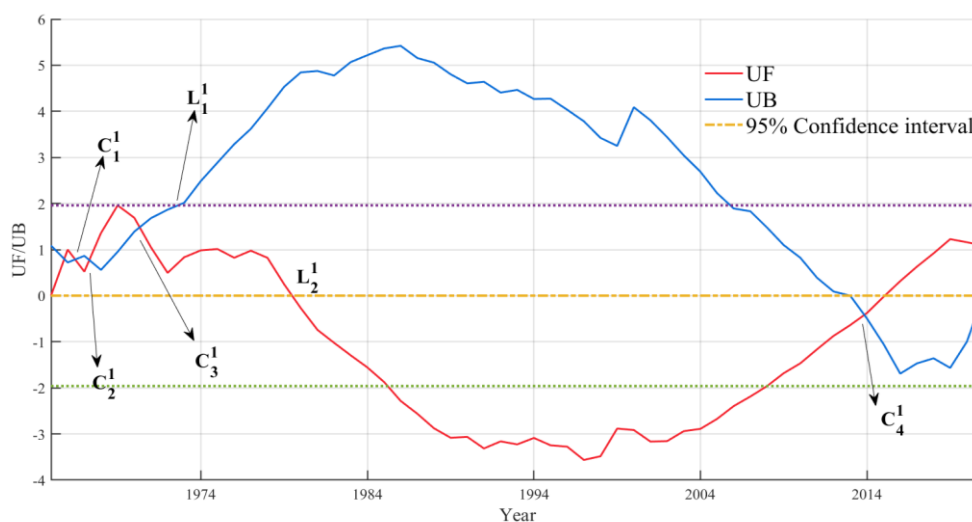


Fig 2. Hydrological station 1 M-K series forecast map.

Hydrological 1M-K series forecast map is shown in Fig. 2. By analyzing and comparing the M-K positive and negative sequence, extreme rainfall may occur at the intersection point. Therefore, extreme rainfall weather may have occurred in the period of C_1^1 (1956–1957), C_2^1 (1959–1960), C_3^1 (1960–1961) and C_4^1 (2013–2014). At the same time, when the range of its order column exceeds

$(-2, 2)$, there will also be sudden changes, that is, there may also be extreme rainfall weather, so there may be extreme rainfall weather at points $L_1^1(1972-1973)$ and $L_2^1(1981-1982)$.

Rainfall Station 2:

Observe and analyze the hydrological monitoring data of hydrological station 2, and analyze the change trend and abrupt change of rainfall. In the process of abrupt change detection, the order sequence of sequential time series and the order sequence of reverse time series are analyzed and calculated respectively. Finally, given the significance level, the time regions of abrupt changes of two abrupt time series curves are found.

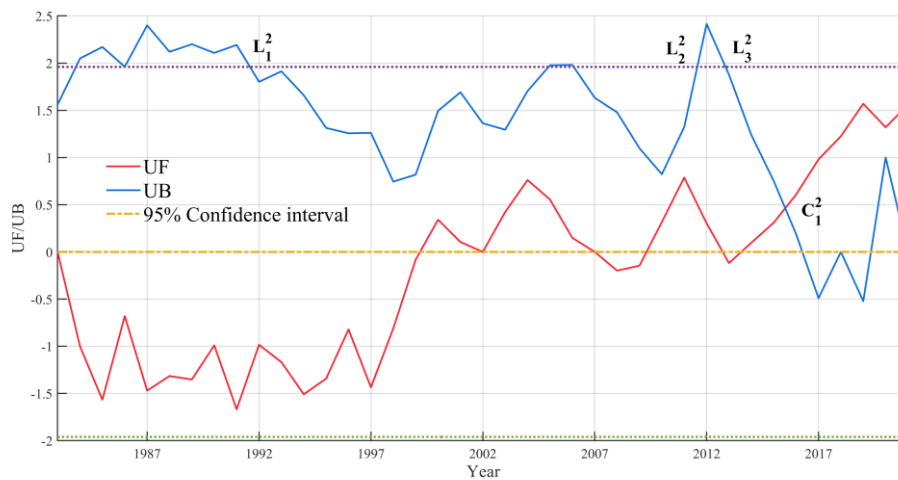


Fig 3. Hydrological station 2 M-K series forecast map.

Hydrological 2 M-K series forecast map is shown in Fig. 3. By analyzing and comparing the M-K positive and negative sequence, extreme rainfall may occur at the intersection point. Therefore, extreme rainfall weather may have occurred in the period of $C_1^2(2014-2015)$. At the same time, when the range of its order column exceeds $(-2, 2)$, there will also be sudden changes, that is, there may also be extreme rainfall weather, so there may be extreme rainfall weather at points $L_1^2(1991-1992)$, $L_2^2(2011-2012)$ and $L_3^2(2012-2013)$.

Rainfall Station 3:

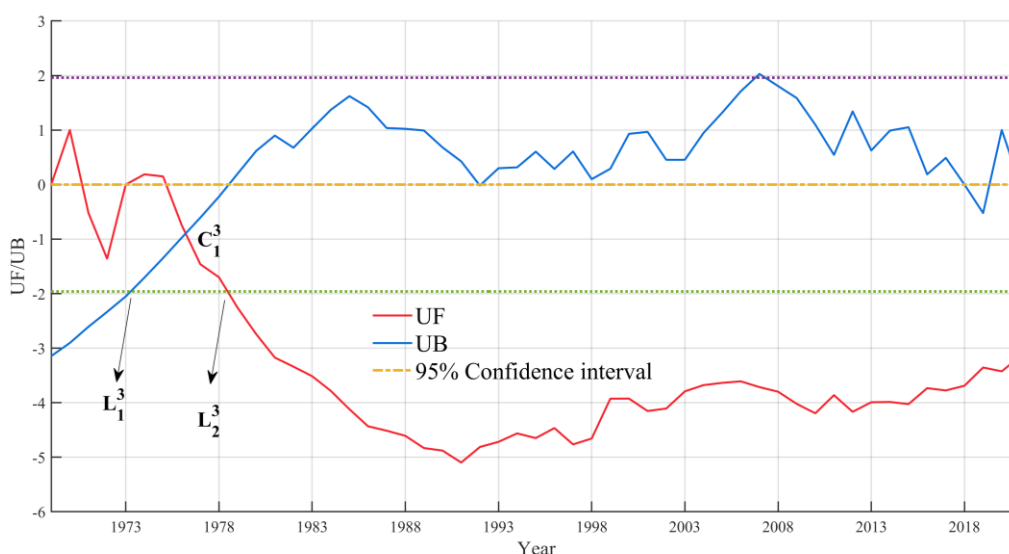


Fig 4. Hydrological 3M-K series forecast map.

Observe and analyze the hydrological monitoring data of hydrological station 3 (Fig. 4), and analyze the change trend and abrupt change of rainfall. In the process of abrupt change detection, the order sequence of sequential time series and the order sequence of reverse time series are analyzed and calculated respectively. Finally, given the significance level, the time regions of abrupt changes of two abrupt time series curves are found.

By analyzing and comparing the M-K positive and negative sequence, extreme rainfall may occur at the intersection point. Therefore, extreme rainfall weather may have occurred in the period of $C_1^3(1976-1977)$. At the same time, when the range of its order column exceeds $(-2, 2)$, there will also be sudden changes, that is, there may also be extreme rainfall weather, so there may be extreme rainfall weather at points $L_1^3(1973-1974)$ and $L_3^2(1976-1977)$.

3.2. Compare rainfall characteristics of two cities

3.2.1. Model establishment and solution

Based on the section A, the two cities follow proposed model.

Hydrological M-K series forecast map of Zhengzhou and Hydrological M-K series forecast map of Taiyuan are shown in Fig. 5 and 6.

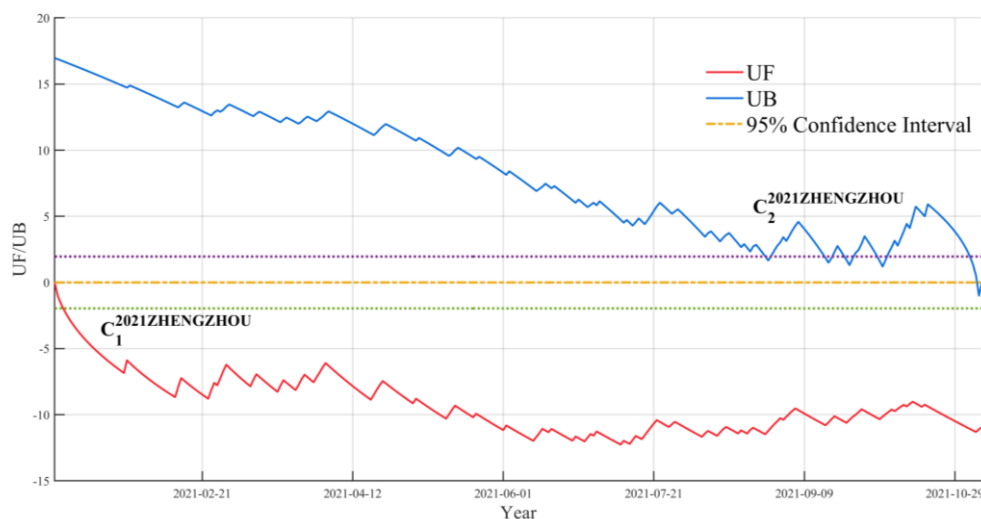


Fig. 5 Hydrological M-K series forecast map of Zhengzhou

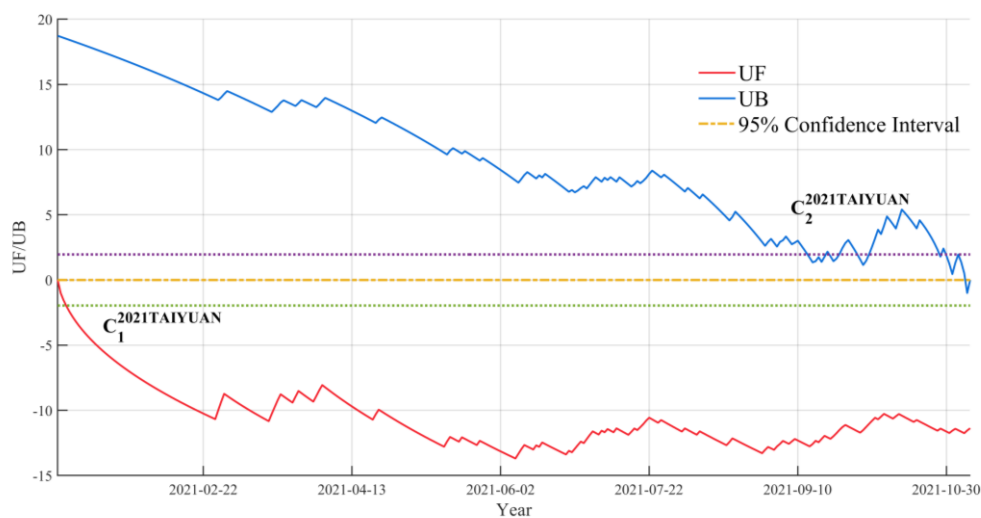


Fig. 6 Hydrological M-K series forecast map of Taiyuan

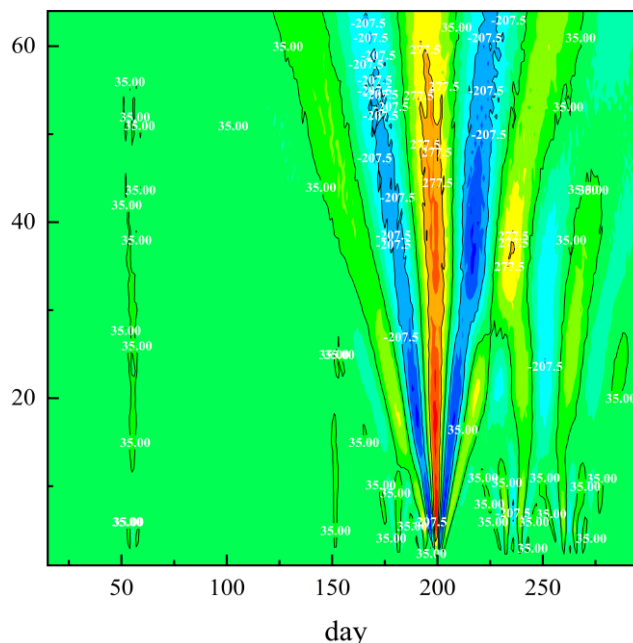


Fig. 7 Wavelet coefficient modulus square contour map of Zhengzhou

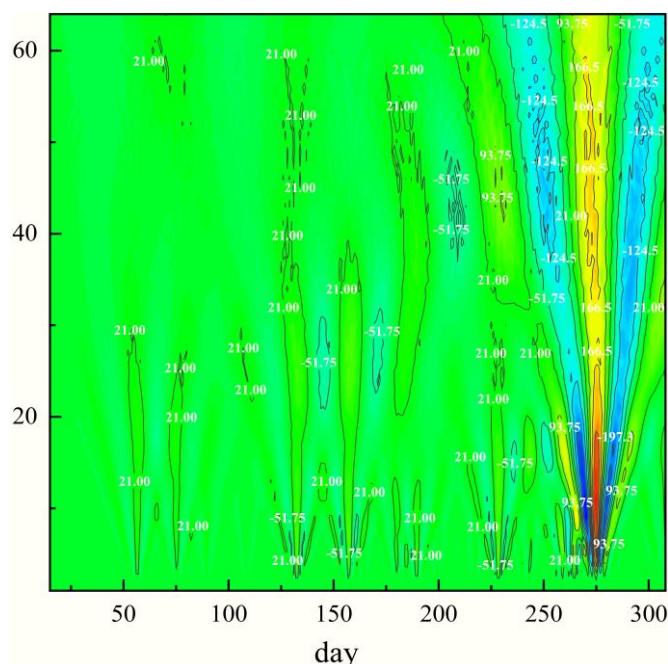


Fig. 8 Wavelet coefficient modulus square contour map of Taiyuan

Wavelet coefficient modulus square contour map of Zhengzhou and Wavelet coefficient modulus square contour map of station Taiyuan are shown in Fig. 7 and Fig. 8.

By solving the characteristics of Zhengzhou rainstorm and Shanxi rainstorm in 2021 and analyzing their correlation, it is clear that both rainstorms have the characteristics of long duration, large accumulated rainfall, wide range of heavy rainfall, concentrated periods of heavy rainfall, extreme and so on.

3.2.2. Analysis of losses caused by rainstorm in two areas

Based on the analysis of the characteristics of Zhengzhou rainstorm and Shanxi rainstorm, it is obvious that the characteristics of two cities are not the same. and thus, there is a world of difference in disaster losses. According to the collected data, the heavy rain in Zhengzhou has caused huge direct economic losses to Henan Province, especially Zhengzhou, regardless of the number of people

affected or killed, the number of collapsed houses or the area affected by crops, far exceeding Shanxi in all aspects [6].

Shanxi's geographical position is very special. It is located between Taihang Mountain and the valley in the middle reaches of the Yellow River, standing on the west side of North China Plain. Usually, Shanxi is considered as a part of the Loess Plateau; but in fact, the whole of Shanxi is a mountainous plateau with a large fluctuation of loess coverage. Throughout Shanxi Province, the terrain is diverse, but most of them are mainly mountains and hills, with few plains. However, not all mountain areas will experience landslides, but only when stimulated by the outside world. This time, there was a large-scale heavy rainfall in Shanxi, which easily saturated the soil, and eventually led to landslides and road surface collapses in the mountains. Due to the geographical factors of Shanxi, during the rainstorm in Shanxi, there were many problems, such as landslides, railway collapse and destruction of ancient city cultural relics, which were different from Zhengzhou in terms of losses. Therefore, the disaster degree in Shanxi is greater than that in Zhengzhou.

3.3. Long-term construction planning

On July 20th, 2021, the rainstorm in Zhengzhou caused heavy casualties and property losses, which aroused widespread concern in the whole society. However, the problem of flood control and drainage reflected by it is worthy of attention. Even though the "sponge city" helps to relieve the pressure of urban rainwater discharge, no "sponge" can bear such intensity of precipitation [3]. Under such extreme precipitation conditions, it is an urgent task to improve the cities' resilience to disasters.

The ability of cities to cope with disasters is not only related to the construction of urban infrastructure, but also closely related to the risk early warning ability of social and economic systems, policy and market systems, response measures and so on. In such an interconnected system, the effect of "wooden barrel" has become particularly prominent, that is, the weakest link in the system has become the decisive factor of system security.

Returning to the torrential rain disaster that Zhengzhou is facing, in the construction of the city's ability to cope with the disaster, we should not only emphasize the improvement of hardware facilities, but also find out the "wooden bucket short board", and reduce the losses as much as possible by relying on the early warning and urban management ability and the systematic project of fully mobilizing people-mechanism-infrastructure.

3.3.1. Improve the forecast level of extreme precipitation weather and issue early warning information through multiple channels

In the future, more data sources such as satellite cloud images, radar echo images, remote sensing images and advanced monitoring and forecasting means should be used to make hourly rolling forecast of disastrous rainstorm weather 3 ~ 5 days in advance, refine the warning standard and issue rainstorm warnings in time more frequently. For extreme weather with heavy rain or rainfall exceeding the red warning threshold of heavy rain, the predicted value should be given, which is convenient for defense personnel to make decisions and issue warning information through multiple channels [7].

3.3.2. Strengthen the standards of urban infrastructure construction

Urban infrastructure is the lifeline system for the normal operation of the city. Once damaged by floods, there will occur a chain reaction between the systems, and the indirect loss will even exceed the direct one [8]. The improvement of urban infrastructure standards can effectively reduce the impact and losses caused by disasters. Therefore, in the future, attention should be paid to improving the defense capabilities of important infrastructure, such as important municipal infrastructure and important traffic facilities. It is suggested that the fortification standard should not be lower than the urban flood fortification standard, and independent emergency plans should be made and practiced.

3.3.3. Intensify the inspection and maintenance of underground pipeline network

Establish a large urban ecological pattern, open natural rivers, wetlands and other drainage systems, upgrade the underground pipe network, simulate emergency measures through digital models, and support emergency management and research of land and resources at the policy level. The reconstruction of the underground pipe network is the top priority, for the construction of the underground pipe network is the crux of urban waterlogging [9].

3.3.4. Strengthen the popular science propaganda and improve the ability of public to avoid meteorological disaster

Make the publicity and popularization of disaster prevention and mitigation an important task of government departments, and make knowledge of disaster prevention and mitigation cover classrooms, villages and communities, so as to normalize the knowledge of meteorological science popularization. In view of areas with high incidence of meteorological disasters, areas where people gather and villages with insufficient warning information, it is even more necessary to strengthen meteorological disaster emergency drills, try one's best to improve the public scientific literacy of disaster prevention and mitigation, and improve the public's ability of emergency handling and self-help and mutual aid in the face of meteorological disasters [10].

4. Conclusion

In this paper, a correlation analysis model based on the calculated average annual precipitation is established and Pearson correlation analysis on the characteristics of rainfall is carried out, which indicates the factors most closely related to rainfall are the maximum and minimum temperatures. Through the method of M-K test and wavelet analysis method, it is concluded that years with more rainfall were 1973, 1976 and 1978. Secondly, our work draws a conclusion that the rainstorm characteristics of the two regions are the same in 2021 through the M-K test, wavelet analysis and prediction analysis. Besides, the damages caused by heavy rainfall in Zhengzhou and Shanxi are quantitatively described and analyzed from several aspects, and the differences of rainfall in the two regions are derived. Finally, this paper analyzes the main problems and long-term urban construction planning under extreme precipitation conditions in Zhengzhou and makes suggestions to improve the level of extreme precipitation weather forecasting and strengthen urban infrastructure construction standards. However, the application scope of wavelet analysis is limited, and the calculation speed is not as fast as the threshold method, and the selection of wavelet bases is also difficult.

References

- [1] Xu Xiaoyan, Guo Chunyi. Study on the Establishment and Application of Drought and Flood Prediction Model in Hunhe River Basin Based on Grey Theory [J], Jilin water resources,2010(5):45-47.
- [2] Wang Hongyu, Ma Fang, Yang Kai, etc. Research and Application of Grey Metabolic Model in Medium and Long-term Urban Precipitation Forecast, Journal of Wuhan University,2004(6):31-34.
- [3] Zhai Qiumin, Zhang Wenjia, Anning, et al. multi-time scale analysis of climate change in southern Henan based on M-K, wavelet and R/S methods [J], Journal of Henan University,2017,47(05):532-543.
- [4] Kuoray. Simulation study on the effect of urban low-impact rain and flood [J]. Technical supervision of water conservancy,2017,25(06):99-101+128.
- [5] Zhao Linlin, Xia Letian. Improvement and Application of Grey Markov Chain Model [J], Journal of Hohai University,2007,35(4):487-490.
- [6] Mei Shuying. Prediction experiment of annual precipitation in Nanjing based on Markov chain [J], Anhui Agricultural Science, 2011(36):2 2415-22419.
- [7] Wang Yang, Wu Jiayang, Wang Min. Application of Weighted Markov Chain in the Prediction of Annual Precipitation in Tai 'an Area [J], Anhui Agricultural Science,2011(6):3577-3579.

- [8] MAKHASANA P, JOSHI G S. Trends in rainfall extremity and peak flood in Sabarmati River Basin, India [J]. Physics and Chemistry of the Earth, Parts A/B/C, 2022, 126.
- [9] NOURANI V, FARBOUDFAM N. Rainfall time series disaggregation in mountainous regions using hybrid wavelet-artificial intelligence methods [J]. Environ Res, 2019, 168: 306-18.
- [10] YIN D, CHEN Y, JIA H, et al. Sponge city practice in China: A review of construction, assessment, operational and maintenance [J]. Journal of Cleaner Production, 2021, 280.