Problem Chosen	2021	Team Control Number
B	ShuWei Cup Summary Sheet	202111097411

Optimal prediction and quantitative analysis of precipitation in cities with extreme precipitation

In this paper, MK analysis model and wavelet analysis model were established to analyze the annual variation characteristics of precipitation characteristics in Zhengzhou, and the gauge rate and characteristics of annual precipitation in Zhengzhou were derived. Then a GM(1,1) and quadratic exponential smoothing prediction model was established for comparison to optimize the precipitation forecast for Changsha city, and the rainfall situation of Changsha in the next two years was derived.

For problem 1, data preparation was first carried out, and the pivot function of Excel was applied to the screening of annual PRCP. After getting the pre-processed data, the MK annual precipitation analysis model was established, and the MK model was programmed by MATLAB, and the PRCP data were imported for calculation can be derived that the years with higher annual precipitation in Zhengzhou are 1958, 1959, 1960, 1963, 1979, 2013, 2015, and 2021. Then a wavelet analysis model is established, and the main period of the annual precipitation time series in Zhengzhou is 12 years, with 12 years as the oscillation period, and the oscillation center is in 1979. The annual precipitation in the whole time series of Zhengzhou region shows a cyclic alternating characteristic of rising and falling. Then a quantitative analysis of the 2021 flood event in Zhengzhou was performed using the variance. It was found that the precipitation in Zhengzhou was relatively small before July, and the precipitation in July suddenly increased and then decreased rapidly, and the excessive precipitation caused the flood.

For problem 2, this paper selected monthly precipitation data from 2009 to 2020 for four cities, Changsha, Guangzhou, Nanning, and Haikou, from the http://data.cma.cn/analysis/yearbooks.html website, and used the MK model for precipitation trend analysis, and found that Changsha city has a sudden change in 2009, 2010, 2013, and 2017 have abrupt change points with a period of about 4 years, and the precipitation trend shows floating changes. Guangzhou city has sudden change points in 2009 and 2019 with a period of about 10 years, and the precipitation trend increases first and then decreases. Nanning and Haikou have one abrupt change point each in 2018 and 2010, respectively, with a period of about 5 years, and the precipitation trend is increasing and then decreasing.

For problem 3, Changsha, which has the most mutation points, was selected as the research object, and a quadratic exponential smoothing based prediction model and GM(1,1) were established, with 80% of the data as the data set and the remaining data as the test set. the mean square error and standard deviation of the exponential smoothing prediction model were 0.128 and 0.357, which were better than 0.892 and 0.479 of GM(1,1), respectively. so the application of exponential smoothing prediction to forecast the precipitation in Changsha city for the next two years, and substituting the predicted data into the MK model, it is found that Changsha will encounter heavy rainfall conditions in May 2021 and April 2022.

For problem 4, using other data from three stations in Zhengzhou City and the precipitation in Taiyuan City, Shanxi Province, a comprehensive analysis was conducted. The rainstorm in Zhengzhou City in July was characterized by the proximity of the abrupt change point, higher temperature, higher pressure, and regional characteristics; while the one in Shanxi Province was characterized by abrupt change in precipitation, thermal reflux, and high terrain. In summary, the rainstorm characteristics of the two are roughly the same. Through the comparative analysis of hazards, it can be found that the scale and degree of hazards of the July rainstorm in Zhengzhou is larger than that of the October rainstorm in Shanxi, while the high temperature during the rainstorm in Zhengzhou causes difficulties in rescue, is located in a low-lying area, has more plains, and has large areas of standing water, leading to huge losses in agriculture, tourism, and transportation, and other differences.

For problem 5, according to the above conclusions on the points and cycles of heavy precipitation and hazard analysis, the city can be built according to the specifications of sponge city in the process of urban construction to enhance the ability of the city to withstand heavy rainfall, but also to regulate the ecological environment of the city, improve the livability of the city and reduce the hazard of extreme precipitation to the city through the analysis of the city under extreme precipitation conditions.

Keywords: MK analysis; Wavelet analysis; Quadratic exponential smoothing; GM(1,1); Sponge city

Content

1. Introduction

1.1 Background

The researchers concerned introduced that, in the context of global warming, the amount, intensity, frequency and type of future precipitation in China will be directly affected. Precipitation is expected to increase by about 10% by the end of this century, with a significant increase in the probability of extreme precipitation. Due to the large land area of China, the precipitation characteristics of different cities show different characteristics due to the combined influence of various types of topography and landforms and other factors. Therefore, it is imperative to establish prediction models for cities with different potential extreme precipitation events and quantitative analysis models for these events.

1.2 Work

Annex 1 shows the daily precipitation observations for the past 70 years from three weather stations near Zhengzhou City. Try to use mathematical modeling to solve the following problems.

Question 1: To correlate the annual variation characteristics of precipitation characteristics in the Zhengzhou area and screen out the years with more precipitation. Also, conduct a specific quantitative analysis of the Zhengzhou 2021 inundation event.

Question 2: Collect and compile precipitation data for more cities in China for many years and analyze the precipitation trends in these cities. (Data sources and how they were obtained should be indicated)

Question 3: Using the collected weather data of cities, we will analyze the prediction of cities that may experience extreme rainfall in the future based on different methods and compare the prediction results.

Question 4: Do you think the characteristics of the rainstorm in Zhengzhou in July 2021 are the same as those of the rainstorm in Shanxi in October 2021? What is the difference in the damage caused.

Question 5: Can you propose a long-term construction plan for cities under extreme precipitation conditions in the future, in which an in-depth analysis of typical cities in China is particularly necessary?

2. Problem analysis

2.1 Analysis of Problem 1

To address question 1, the data were first prepared by applying Excel's pivot function to calculate the annual data for each indicator to obtain the annual data for each indicator. In order to analyze the changing characteristics of the annual precipitation characteristics in Zhengzhou, the M-K analysis model was applied to analyze the data from the three meteorological stations for abrupt changes. The analysis was mainly to study in which years the more obvious changes occurred, and the M-K method has the characteristics of wide detection range and less artificiality, which is more suitable for this question. In order to derive the amount of temporal variation in precipitation water, wavelet cycle analysis is subsequently performed to analyze what kind of variation characteristics the annual average precipitation has on the time scale. The annual precipitation mutation point and period relationship of Zhengzhou were derived, and then 2021 data were selected to visualize the image as shown in the figure below with a daily period unit, and then all 2021 data were analyzed using ANOVA for a specific quantitative analysis of the 2021 flood event in Zhengzhou[1].

2.2 Analysis of Problem 2

For question 2, monthly precipitation data from 2009 to 2020 for four cities, namely Changsha, Guangzhou, Nanning, and Haikou, were selected from the http://data.cma.cn/analysis/yearbooks.html website as our research object to study and analyze the relationship between precipitation and practice changes in these four cities, i.e., to study their abrupt change points and cycle The pattern of changes. The data of these four cities are substituted into the MK model of Problem 1 for analysis, and the precipitation trends of the selected cities are finally obtained.

2.3 Analysis of Problem 3

For question 3, according to question 2, it can be concluded that a city has more sudden change points in precipitation, indicating that this city experiences more extreme rainfall conditions, so for this question we choose to study the data for this city. Since only the precipitation for each month of the 12 years was collected, the sample data is relatively small and also in the time domain condition, so we apply the quadratic exponential smoothing prediction model and the gray prediction model to make predictions, and then compare the error rate of the test data and choose the model with a smaller error rate to make the prediction of extreme rainfall situations. Using monthly precipitation as the metric, 80% of the data is applied as the data set, and the remaining 20% of the data is used as the test set to compare the accuracy of their predictions. The model with the smaller error rate is selected to substitute all the data for the prediction, which ultimately results in the precipitation amount for 2021 and 2022, and then MK analysis is performed to finally arrive at the intense precipitation situation for this city.

2.4 Analysis of Problem 4

For question 4, based on the above conclusions, the characteristics of heavy rainfall in Zhengzhou, i.e., sudden change points, cycles, and predicted values, are derived. Using these characteristics, together with the values of other indicators, the causes of the flooding in Zhengzhou in July and the damages caused are analyzed; then a comparison is made with Shanxi to finally conclude whether they are the same.

2.5 Analysis of Problem 5

In response to question 5, recommendations for city construction are made based on the conclusions drawn from the previous 4 questions.

3 Symbol and Assumptions

3. 1 Symbol Description

Table 3.1 Description of the meaning of symbols[1]

Symbols	Meaning
STATION	Meteorological station number
DATE	Date
DEWP	Dew point temperature
FRSHTT	Various indicators
GUST	Gusts of wind
MAX	Maximum temperature
MIN	Lowest Temperature
MXSPD	Maximum wind speed
PRCP	Precipitation
SLP	Sea level pressure
SNDP	Snow depth
STP	Air pressure of this site
TEMP	Temperatures
VISIB	Visibility
WDSP	Wind direction and speed

Note: The specific meaning of the remaining symbols has been specified in the article.

3.2 Fundamental assumptions

1. Assumes that the data in the annexes are fully consistent with objective laws

and have practical relevance.

2. Assuming that all external influences in the process of forecasting are negligible and have no effect on the process and results of forecasting.

3. assuming that the data found in open source sites and databases are authentic and reliable.

4. It is assumed that the precipitation data from the three stations can fully represent the precipitation situation in Zhengzhou City.

5. Taiyuan is the capital city of Shanxi Province, so it is assumed that the climate and precipitation of Taiyuan can fully represent the climate and precipitation of Shanxi Province.

4 Modeling and Answering of Question 1

First of all, the data should be prepared, and the annual data of each index should be calculated by applying the pivot function of Excel to obtain the annual data of each index. In order to analyze the changing characteristics of the annual precipitation characteristics of Zhengzhou, the MK analysis model is applied to analyze the data from three meteorological stations for abrupt changes. The analysis is mainly to study in which years the more obvious changes occurred, and the M-K method has the characteristics of wide detection range and less artificiality, which is more suitable for the characteristics of this question. In order to derive the amount of temporal variation in precipitation water, wavelet cycle analysis is subsequently performed to analyze what kind of variation characteristics the annual average precipitation has on the time scale. The annual precipitation mutation point and period relationship of Zhengzhou were derived, and then 2021 data were selected to visualize the image as shown in the figure below with a daily period unit, and then all 2021 data were analyzed using ANOVA for a specific quantitative analysis of the 2021 flood event in Zhengzhou.

4.1 Data Preparation

Applying Excel's pivot function, the annual average of precipitation for each of the three indicators is derived and exported to a new data table for storage, which is visualized in Figure 4.1 below.

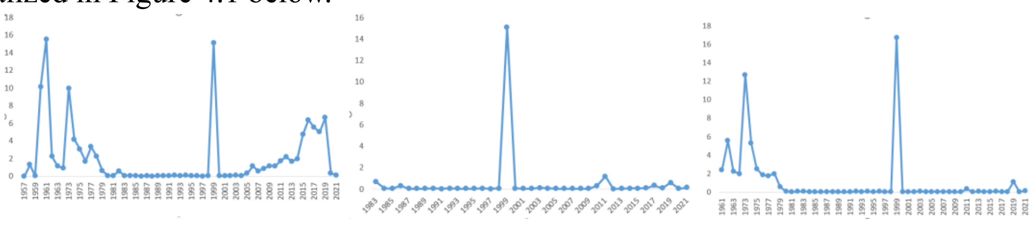

Figure 4.1 Visual image of the annual average of precipitation (left to right in station order)

From Figure 4.1, we can see that there is a regularity of annual precipitation. In order to explore the relevant characteristics and relationships, we proceed to the next step of analysis and establish the M-K model.

4.2 MK model building

In this paper, the MK test is used to analyze the annual precipitation at precipitation stations in Zhengzhou, and to obtain the annual precipitation trend and the year of sudden change in precipitation in Zhengzhou.

4.2.1 MK trend test

Step1：Constructing time series of annual precipitation at precipitation stations in Zhengzhou （$x_1, x_2, \dots, x_n$）， Assume that H_0 represents the data in this time series, and that the data are independent, random variables with the same distribution of the sample, with no linear trend[1].

Step2：Calculate the statistical variable of the test S

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn\ (x_j - x_i) \tag{4.1}$$

Step3：For 10 and more series points to get time series, this question Zhengzhou has more than 10 data points, we apply the normal approximation and the variance of the statistical value S[1,2].

$$var(S) = \frac{1}{18}[n(n-1)(2n=5) - \sum_{i=1}^{n} e_i\,(e_i - 1)(2e_i = 5)] \tag{4.2}$$

The number of samples with the same annual precipitation is i. The number of samples with different precipitation among the samples with the same value of i is e_i.

Step4：The constructed standard normal distribution variable Z is calculated as follows Equation 4.3.

$$Z = \begin{cases} \frac{S-1}{\sqrt{var(S)}}, & S > 0 \\ 0, & S = 0 \\ \frac{S+1}{\sqrt{var(S)}}, & S < 0 \end{cases} \tag{4.3}$$

Step5：At a certain α confidence level, if $Z \geq |Z_{1-\alpha/2}|$ then the original hypothesis is rejected, then it can be concluded that there is a clear trend in the annual precipitation time series at the α confidence level, then it can be concluded that the magnitude of the trend is as follows[1].

Assuming the trend size is β, Then the equation for β is

$$\beta = Jianshui\left(\frac{x_k - x_j}{k-j}\right), \forall j < k \tag{4.4}$$

When $\beta > 0$,it means an upward trend; if $\beta < 0$, it means a downward trend. After modeling the trend, we also need to detect the mutation points in the trend, so we also need to establish a mutation detection model for the annual precipitation of Zhengzhou precipitation sites.

4.2.2 MK mutation test

Assume that the time series of annual precipitation at precipitation stations in Zhengzhou is （$x_1, x_2, \dots, x_n$）,Order

$$s_k = \sum_{i=1}^{k} r_i, k = 2,3, \dots, n \tag{4.5}$$

Among them.

$$r_i = \begin{cases} 1, & 当x_i > x_j, \\ 0, & 当x_i \le x_j, \end{cases} j = 1,2, \ldots, i \tag{4.6}$$

Then we define the statistic.

$$UF_k = \frac{[s_k - E(s_k)]}{\sqrt{var(s_k)}}, k = 1,2, \ldots, n \tag{4.7}$$

Where the mean and variance of s_k are $E(s_k)$, $var(s_k)$, respectively, and in the case of independent identically distributed time series, their expressions are[2,3]

$$\begin{cases} E(s_k) = \frac{k(k-1)}{4} \\ var(s_k) = \frac{k(k-1)(2k+5)}{72}, \end{cases} k = 2,3, \ldots, n \tag{4.8}$$

Arrange the time series x in reverse order $x_n, x_{n-1}, \ldots, x_1$,and then find UF_k by equation (4.7), while making $UB_k = -UF_k$ （$k = n，n-1，\ldots，1$）, $UB_1 = 0$.

The statistical series UB_k and UF_k were obtained, and the positive and negative values of UF_k indicated the trend of x_k. Positive values indicated an upward trend of the series, and negative values indicated a downward trend; when UB_k and UF_k exceeded the critical line of significance, the trend was significant. If the intersection point of UB_k and UF_k curves is between the critical lines, the moment corresponding to the intersection point is the time when the abrupt change occurs. At this point, the sudden change analysis model of precipitation in Zhengzhou has been established, but in order to make a comprehensive analysis of the precipitation characteristics of Zhengzhou, it is necessary to analyze the time domain, i.e., period analysis. The following wavelet analysis model will be established to analyze the period of precipitation in Zhengzhou.

4.3 Wavelet analysis model building

Morlet wavelet is a non-orthogonal complex-valued wavelet, so most of the domestic and foreign experts use Morlet wavelet for periodic analysis of weather time series.

The Morlet wavelet function and discrete wavelet transform equations are[1]

$$\varphi(t) = \pi^{-1/4} e^{-t^2/2} e^{-iabt} \tag{4.9}$$

$$\omega_f(a,b) = |a|^{-\frac{1}{2}} \sum_{i=1}^{N} f(i\delta_t) \varphi^* \left(\frac{i\delta_t - b}{a}\right) \tag{4.10}$$

Where ω_0 is the angular frequency, $\omega_f(a,b)$ is the wavelet coefficient after wavelet transform, a and b are the scale factor (indicating the wavelet period length) and translation factor (indicating the translation in time), respectively, $f(i\delta_t)$ is the precipitation time series for conducting the analysis, and δ_t is the time interval of the time series. Since the annual precipitation in Zhengzhou is studied, then the value of δ_t is 1.

According to the definition of the power spectrum, it can be derived that the power spectrum of the wavelet at this time is

$$E_{a,b} = |\omega_f(a,b)|^2 \tag{4.11}$$

$$P = \sigma^2 P_a \frac{\chi_2^2}{2} \tag{4.12}$$

Where P_a is the red noise or white noise power spectrum.

$$P_a = \frac{1-r(1)^2}{1+r(1)^2-2r(1)\cos\left(\frac{2\pi\delta_t}{1.033a}\right)} \tag{4.13}$$

$r(1)$ autocorrelation coefficient of the precipitation time series lag 1, χ_2^2 is the value of the chi-square with degree of freedom v=2 at a particular significance.

If $E_{a,b} > P$,it means that the wavelet power spectrum is significant. This problem applies the standard spectrum of red noise and white noise to test the significance of wavelet power spectrum, so as to obtain the periodic variation pattern of annual precipitation in Zhengzhou[1].

4.4 Answers to the model

4.4.1 Solution of MK model

By programming the MK model with MATLAB, the abrupt change point of annual precipitation in Zhengzhou can be derived as shown below, i.e., when the value of UF exceeds 2, it is judged to be a sudden change point, i.e., having a strong rainfall point.

The UF and UB curves plotted by the aforementioned MK mutation test method are shown in Figure 4.2, Figure 4.3, and Figure 4.4.

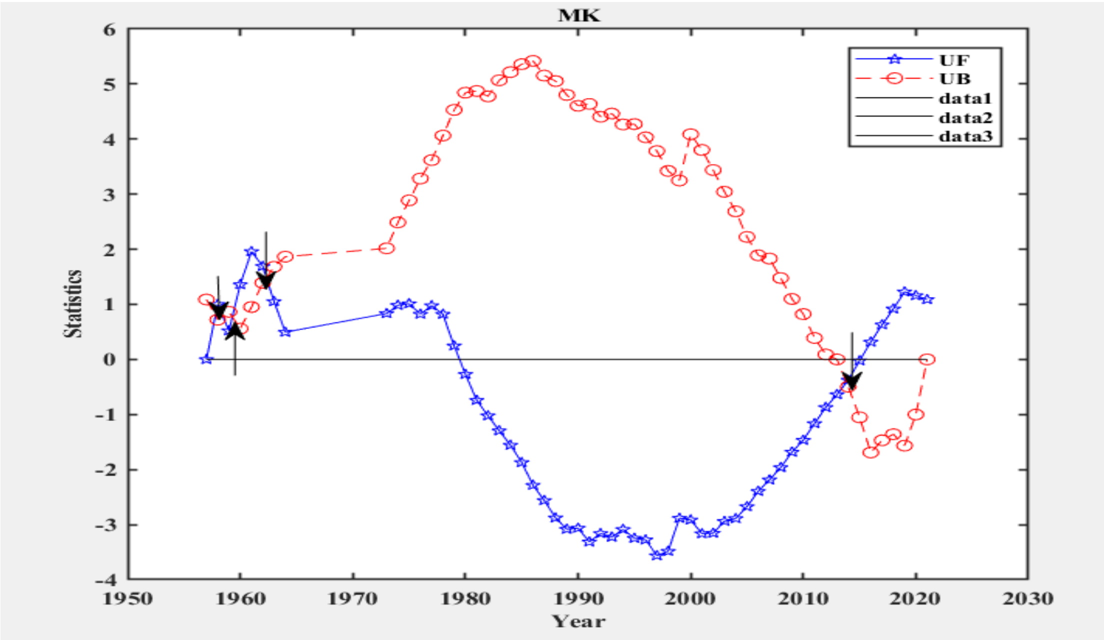

Figure 4.2 STATION for 57083099999 site MK statistics

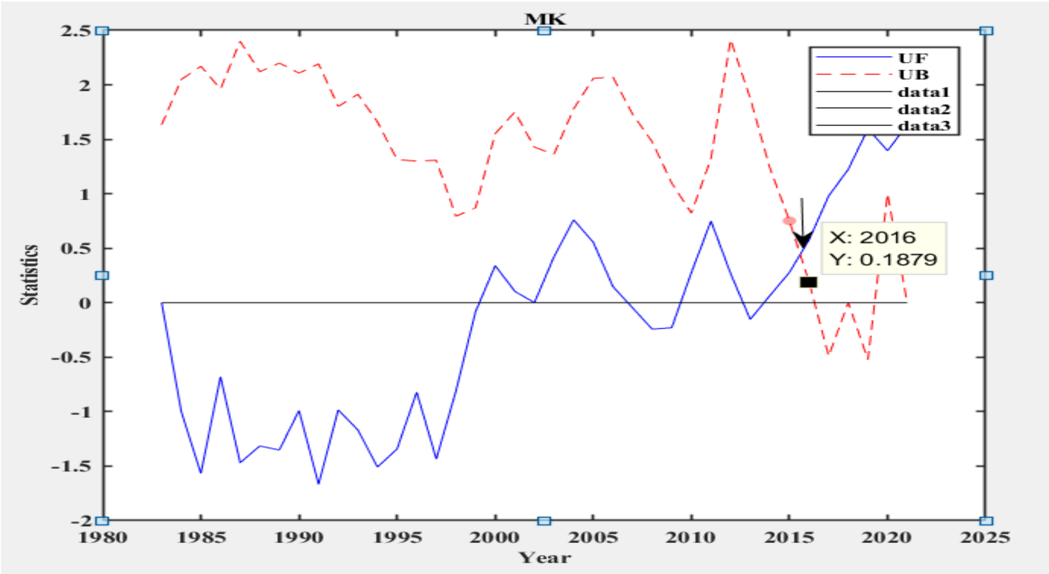

Figure 4.3 STATION for 57071099999 site MK statistics

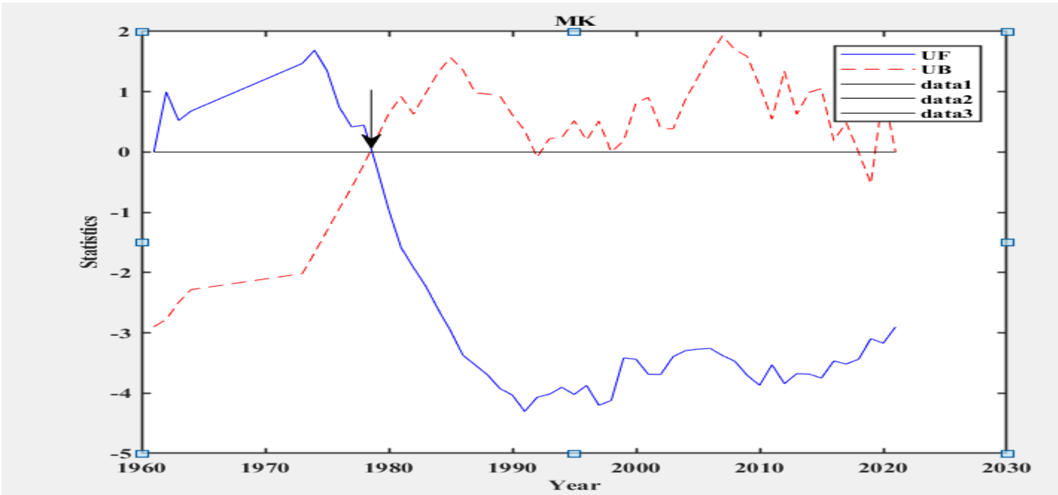

Figure 4.4 STATION for 53975099999 site MK statistics

From Fig. 4.2, Fig. 4.3 and Fig. 4.4, it can be concluded that the trend of UF_k trend line of the annual rainfall MK test at site 57083099999 clearly shows that the trend of annual precipitation in Zhengzhou: it rose in the early 1950s, then fell, and showed a downward trend in the early 1960s, and showed a downward trend after the 1970s until the early 21st century, and showed a rising downward trend in the early 2000s. The trend of UF_k The sudden change of annual precipitation can be seen from the intersection of UB_k and UF_k and the years of sudden change of precipitation are 1958, 1959, 1960, 1963, and 2013. From the UF_k trend line of the annual precipitation MK test at site 57071099999, it can be clearly seen that the precipitation at site 57071099999 has been fluctuating up and down after the 1980s, and in the mid-1990s, it was on an upward fluctuation trend, and the intersection of UB_k and UF_k shows that the sudden change of annual precipitation occurred in 2015. The trend line of UF_k of the annual precipitation MK test at site 53975099999 clearly shows that the trend from the 1960s to the 1980s is up, but after the 1980s is a downward trend, from the intersection of UB_k and UF_k can be seen in the sudden change of annual precipitation phenomenon, the sudden change of precipitation year is 1979[1,3].

In summary, the years with sudden changes in annual precipitation in Zhengzhou are 1958, 1959, 1960, 1963, 1979, 2013, 2015, and 2021, i.e., with a trend of heavy precipitation in these years.

4.4.2 Solutions for wavelet models

The wavelet cycle run image can be calculated by MATLAB programming as shown below.

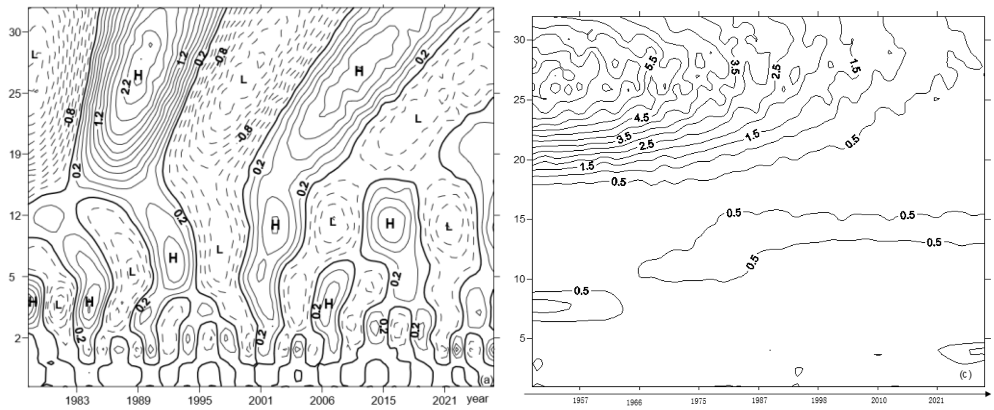

Figure 4.5 Wavelet time-frequency distribution
(left: site 57083099999, right: site 57071099999)

A comprehensive analysis shows that the annual precipitation in Zhengzhou is influenced by the fluctuations of 1-year, 5-year and 19-year scales, and the 5-year time scale accounts for the major part of the energy and is the main time scale affecting the future precipitation in the region, which mainly occurs from the 1960s to the end of 1970s; the main cycle of the annual precipitation time series in Zhengzhou is 12 years, and the oscillation is centered on the 12-year cycle in 1979 The center of the oscillation is in 1979. The annual precipitation in the whole time series of Zhengzhou region shows a cyclic alternating feature of rising and falling[1,2,3].

The final years with higher annual precipitation in Zhengzhou are 1958, 1959, 1960, 1963, 1979, 2013, 2015, and 2021, and the annual precipitation in the whole time series of Zhengzhou area shows a cyclic alternating characteristic of rising and falling.

4.5 Quantitative analysis of flooding in Zhengzhou

We selected 2021 data, with the day as the period unit, for visualization image as shown below, and then performed ANOVA on all 2021 data, and applied Excel's pivot function to produce the results shown in Figure 4.6 below.

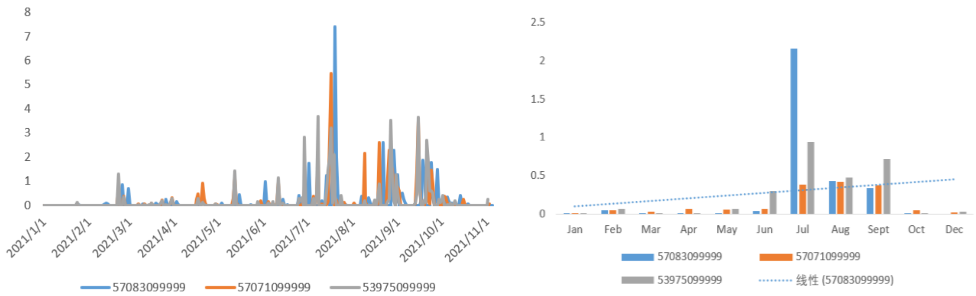

Figure 4.6 Precipitation data (left) and ANOVA (right) for 2021 in Zhengzhou

From Figure 4.5, it can be concluded that the variance-processed data for

Zhengzhou in 2021 can visualize the huge amount of precipitation in July 2021, which became larger and led to the flooding. Also observing the ANOVA image, it is found that the precipitation in July 2021 suddenly increases and then rapidly decreases.

5 Modeling and Answering the Question 2 Problem

The monthly precipitation data from 2009 to 2020 for four cities, namely Changsha, Guangzhou, Nanning, and Haikou, were selected from http://data.cma.cn/analysis/yearbooks.html as our research object to study and analyze the relationship between precipitation and practice changes in these four cities, i.e., to study their sudden change points and the law of cycle changes. The visualization of the selected data is shown in the following figure5.1[9].

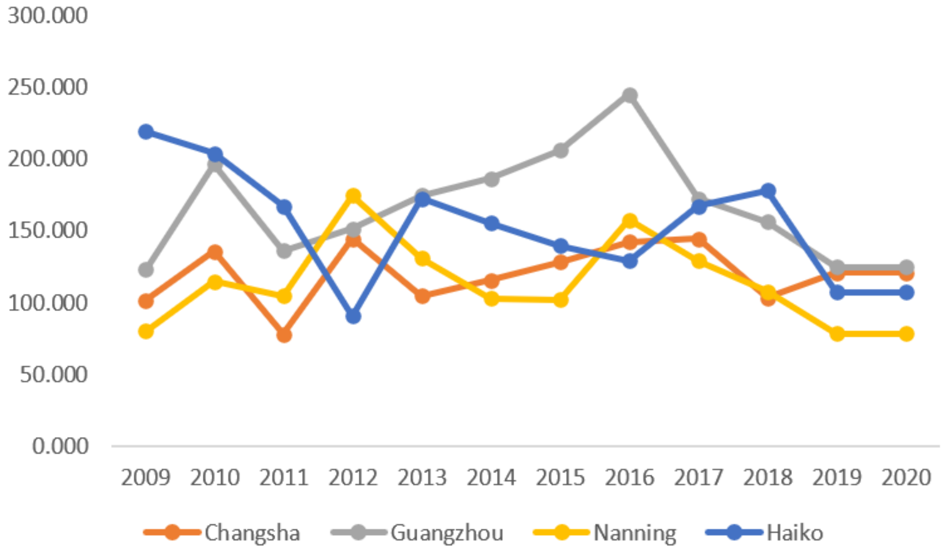

Figure 5.1 Visualization images of precipitation data for the four selected cities

The data from these four cities were substituted into the MK model of question 1 for analysis, and the results are shown in the following figure[2].

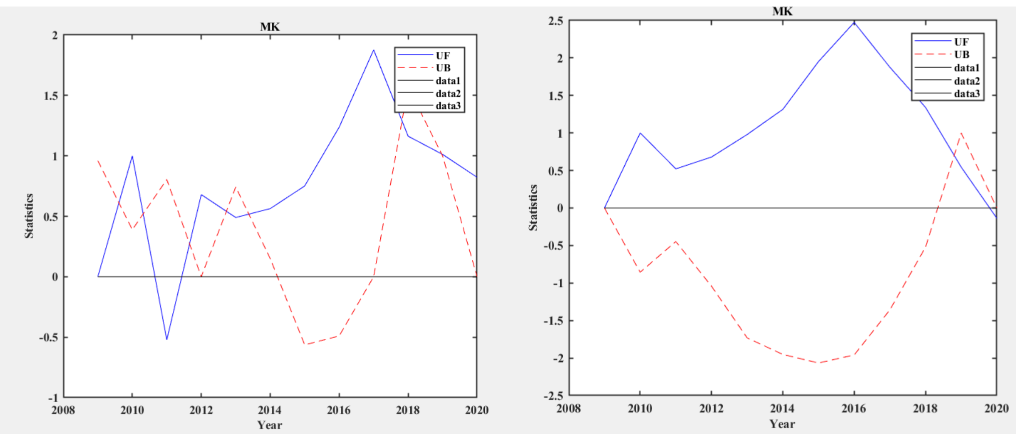

Figure 5.2 MK analysis of precipitation in Changsha (left) and MK analysis of precipitation in Guangzhou (right)

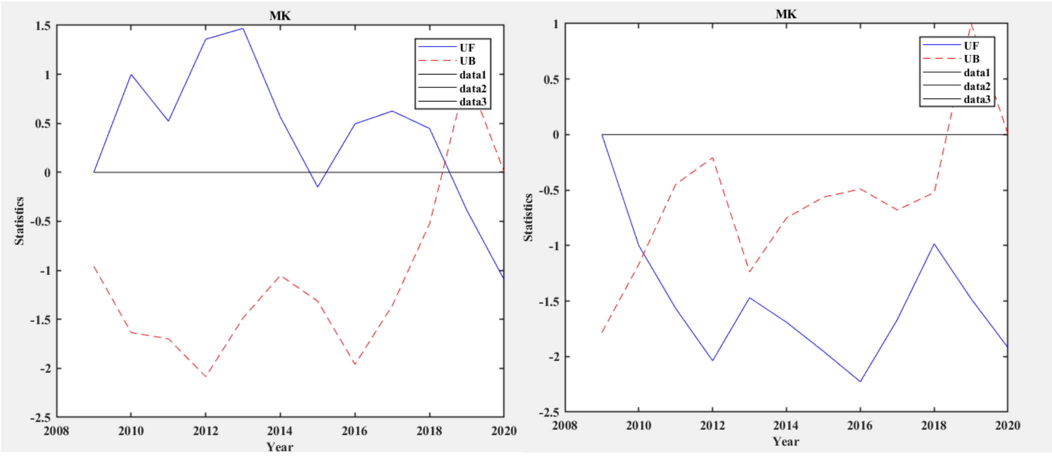

Figure 5.3 MK analysis of precipitation in Nanning (left) and Haikou (right)

From Fig. 5.1, Fig. 5.2 and Fig. 5.3, we can get that Changsha City has a decreasing trend during 2009-2011, and then from 2011, the precipitation has an increasing trend, and the sudden change of annual precipitation can be seen from the intersection of UB_k and UF_k, so Changsha City has four sudden change points of precipitation, which are in 2009, 2010, 2013 and 2017. The period of heavy precipitation is about 4 years.

Observing the MK trend line we can conclude that Guangzhou city has an increasing trend of precipitation from 2009 to 2016, and then after 2016, it has a slow decreasing trend. Observing the intersection of UB_k and UF_k, we can conclude that Guangzhou city has 2 sudden precipitation change points, in 2009 and 2019, and the period of heavy precipitation is about 10 years.

Observing the MK curve of precipitation in Nanning, we can get that the precipitation in Nanning has a slow upward trend from 2009 to 2013, and a slow downward trend after 2013. Observing the intersection of UB_k and UF_k, we find that Nanning has 1 abrupt change point in 2018. Observing its trend, we can find that the period of heavy precipitation is about 5 years.

Observing the MK curve of precipitation in Haikou city, we find that the precipitation in Haikou city decreases with a slow trend, and checking the intersection of UB_k and UF_k, we find that Haikou city has a sudden change point, in 2010. Observing its MK trend and precipitation trend, we can find that the period of heavy precipitation in Haikou City is about 7 years.

6 Question 3 Modeling and Answering

According to question 2, it can be concluded that there are more sudden change points of precipitation in Changsha, which indicates that Changsha has more extreme rainfall situations, so we choose to study the data of Changsha in this question. Since only the precipitation for each month of the 12 years was collected, the sample data is relatively small and also in the time domain condition, so we apply the quadratic exponential smoothing prediction model and the gray prediction model for prediction, and then compare the error rate of the test data and choose the model with a smaller error rate for the extreme rainfall situation prediction. Using monthly precipitation as the metric, 80% of the data is applied as the data set, and the remaining 20% of the data is used as the test set to compare the accuracy of their predictions[4].

6.1 Quadratic exponential smoothing prediction model

6.1.1 Establishment of the secondary exponential smoothing forecasting model

Let the time series be $y_1, y_2, ..., y_t$, α is a weighting factor, $0 < \alpha < 1$, The formula for primary exponential smoothing is equation (6.1).

$$S_t^{(1)} = \alpha y_t + (1+\alpha) S_{t-1}^{(1)} = S_{t-1}^{(1)} + \alpha (y_t - S_{t-1}^{(1)}) \quad (6.1)$$

Where: y_t is the observation series and $S_t^{(1)}$ is the primary smoothing index in period t.

The recursive formula for the moving average is equation (6.2).

$$M_t^{(1)} = M_{t-1}^{(1)} + \frac{y_t - y_{t-N}}{N} \quad (6.2)$$

where: $M_t^{(1)}$ is a smoothed exponential shift term in period t[4].

Using $M_{t-1}^{(1)}$ as the best estimate of y_{t-N}, we have equation (6.3).

$$M_t^{(1)} = M_{t-1}^{(1)} + \frac{y_t - M_{t-1}^{(1)}}{N} = \frac{y_t}{N} + (1 - \frac{1}{N}) M_{t-1}^{(1)} \quad (6.3)$$

Letting $\alpha = \frac{1}{N}$ and replacing $M_t^{(1)}$ with S_t, we obtain equation (6.4).

$$S_t^{(1)} = \alpha y_t + (1+\alpha) S_{t-1}^{(1)} \quad (6.4)$$

Further understanding the essence of exponential smoothing, then expanding equation (6.1) in sequence, we have equation (6.5).

$$S_t^{(1)} = \alpha y_t + (1+\alpha)[\alpha y_{t-1} + (1-\alpha) S_{t-2}^{(1)} = \alpha \sum_{i=0}^{\infty} (1-\alpha)^j y_{t-j} \quad (6.5)$$

Where: $S_t^{(1)}$ is the weighted average of all historical data with a weighting factor score of $\alpha, \alpha(1-\alpha), \alpha(1-\alpha)^2, ...;$. Then, it clearly follows that equation (6.6)[5].

$$\sum_{i=0}^{\infty} (1-\alpha)^j = \frac{\alpha}{1-(1-\alpha)} = 1 \quad (6.6)$$

The primary exponential smoothing model is equation (6.7).

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha) \hat{y}_t \quad (6.7)$$

where: the meaning of the equation is to use the exponential smoothed value of period t as the forecast value of period 1+t.

The one-time exponential smoothing method overcomes the shortcomings of the moving average method. However, when there is a linear trend in the movements of the time series, forecasting with primary exponential smoothing still has significant lagging bias. Therefore, it must also be corrected. The correction method is the same as the trend moving average method, i.e., a second exponential smoothing is done to model the linear trend using the law of lagged deviation. This is the quadratic exponential smoothing method. The formula is equation (6.8)[4,6].

$$S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha) S_{t-1}^2 \qquad (6.8)$$

Where: $S_t^{(2)}$ is the quadratic exponential smoothing value, when the time series $\{y_t\}$, starting from a period with a linear trend, similar to the trend moving average method, can be predicted by a linear trend model, as in equation (6.9) and equation (6.10).

$$\hat{y}_{t+T} = a_t + b_t T, T = 1, 2, \ldots \qquad (6.9)$$

$$s.t \begin{cases} a_t = 2S_t^{(1)} - S_t^{(2)} \\ b_t = \dfrac{\alpha}{1-\alpha}(S_t^{(1)} - S_t^{(2)}) \end{cases} \qquad (6.10)$$

Where: t is the number of current periods; T is the number of periods from t to the forecast period; a_t is the intercept; and b_t is the slope, both of which are also known as smoothing coefficients.

6.1.2 Solution of the model

This can be derived by MATLAB programming and importing the data for calculation, which leads to Figure (6.1).

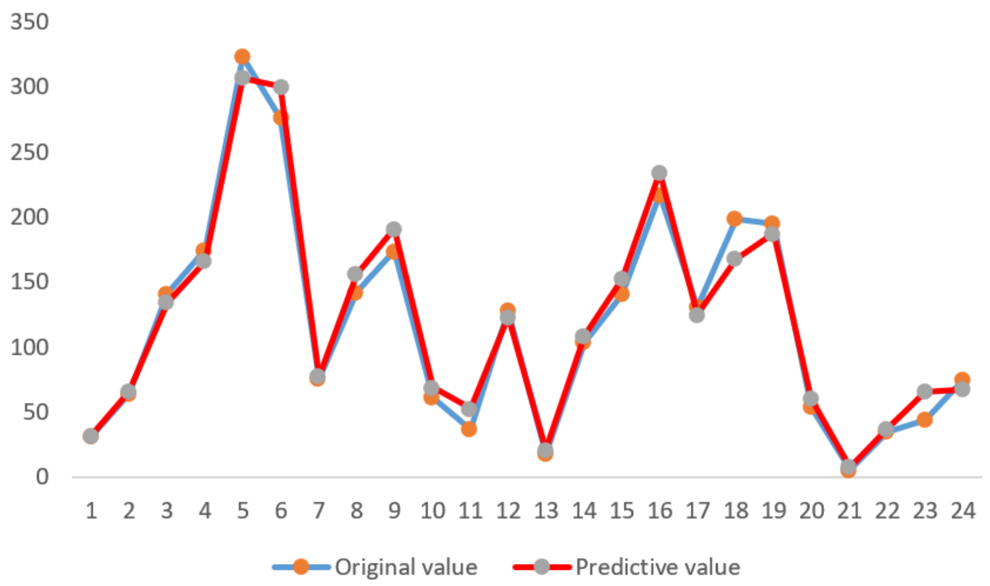

Figure 6.1 Exponential smoothing prediction comparison image

From Figure 6.4, it can be found that the accuracy of the model's prediction is higher when the predicted and tested values fit better.

6.1.3 Evaluation of errors

In this paper, the mean square error (MSE) and standard deviation (SD) are chosen as evaluation indexes to judge the performance of each prediction model.The formulae for calculating MSE and SD are shown in Equation (6.11), Equation (6.12).

$$MSE = \frac{1}{n}\sum_{i-1}^{n}(f(x_i) - y_i)^2 \tag{6.11}$$

$$SD = \sqrt{\frac{1}{n}\sum_{i-1}^{n}(f(x_i) - \mu_i)^2} \tag{6.12}$$

Where n is the number of data samples, $f(x_i)$ is the model prediction value, y_i is the ideal optimal value (i.e., reference value), and μ_i is the mean value of the model prediction value. the smaller the value of MSE, the smaller the error between the model prediction value and the ideal optimal value, and the higher the prediction accuracy. the smaller the value of SD, the less the model prediction value deviates from the prediction mean, and the higher the prediction stability.

Then after calculation, the MSE value of the quadratic exponential smoothing prediction model can be obtained as 0.128 and the value of SD as 0.357.

6.2 Gray prediction model

Gray prediction refers to the use of the GM(1, 1) model [4] to estimate and predict the development of the behavioral characteristics of the system, as well as to estimate and calculate the moment of occurrence of anomalies in the behavioral characteristics, and to study the future time distribution of the events occurring in a specific time zone. In essence, these works treat "stochastic processes" as "gray processes" and "random variables" as "gray variables", and are mainly based on gray systems. These works essentially treat "stochastic processes" as "gray processes" and "random variables" as "gray variables" and deal with them mainly with the GM(1, 1) model of gray system theory.

6.2.1 G(1,1) rank ratio tests

First of all, in order to ensure the feasibility of the modeling method, it is necessary to do the necessary test processing of the known data columns [5]. Twenty sample data were randomly selected and the reference data were set to $x^{(0)} = \left(x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(20)\right)$ calculate the series of cascade equation (6.13)[7].

$$\lambda(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)}, k = 2, 3, \cdots, 20 \triangleright \tag{6. 13}$$

If all the level ratios $\lambda(k)$ fall within the tolerable coverage

$\Theta = \left(e^{-\frac{2}{21}}, e^{\frac{1}{11}} \right) =$ （-0.09524,0.09091）, then the series $x^{(0)}$ can be used as data for model $GM(1,1)$ for gray prediction. Otherwise, the necessary transformations need to be done on the series $x^{(0)}$ to make it fall within the tolerable coverage. That is, take an appropriate constant c and make a translation transformation as in equation (6.14).

$$y^{(0)}(k) = x^{(0)}(k) + c, k = 1, 2, \cdots, 20 \qquad (6.14)$$

Then the ratio of the series $y^{(0)} = \left(y^{(0)}(1), y^{(0)}(2), \cdots, y^{(0)}(n) \right)$ such that the series equation (6.15).

$$\lambda_y(k) = \frac{y^{(0)}(k-1)}{y^{(0)}(k)} \in \Theta, k = 2, 3, \cdots, \mathrm{n} \qquad (6.15)$$

By using the known data and using matlab to calculate $\lambda_y(k) \in \Theta$, then a gray prediction model can be built.

6.2.2 Establishment of G(1,1) model

Building the model $GM(1,1)$ equation (6.16), the predicted values can be obtained[8].

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-ak} + \frac{b}{a}, k = 0, 1, \cdots, n-1, \cdots \qquad (6.16)$$

And

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k), k = 1, 2, \cdots, n-1, \cdots 。 \qquad (6.17)$$

6.2.3 Solution of G(1,1) model

The results that can be derived from MATLAB calculations with the training comparison fitting accuracy are shown in the following figure6.2.

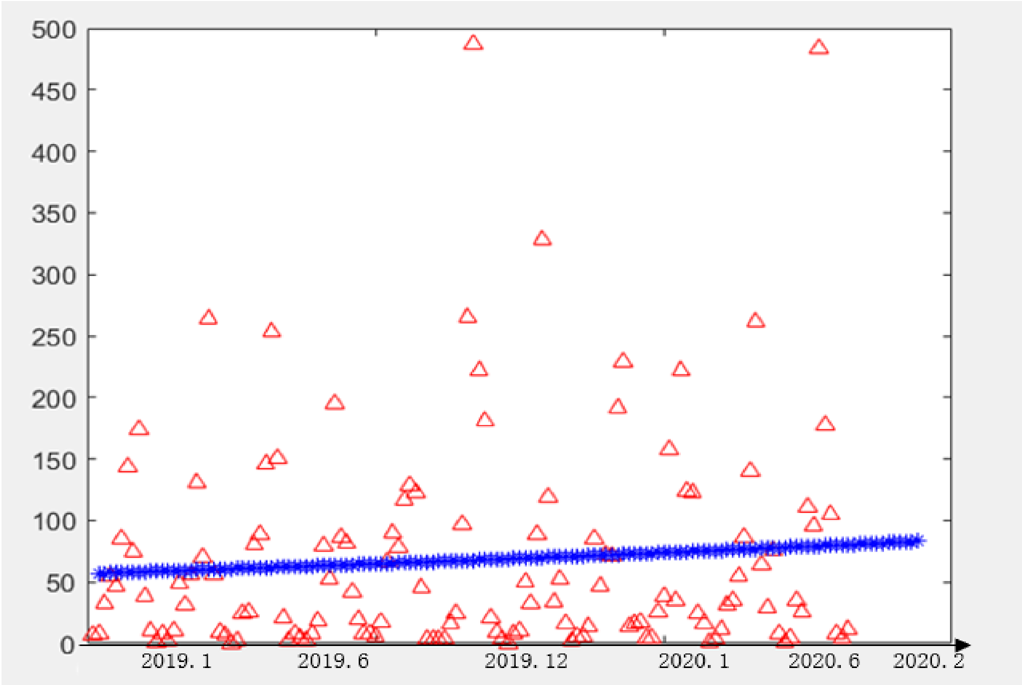

Figure 6.2 GM(1,1) calculation diagram

From Figure 6.2, it is known that the fit of GM(1,1) is low, and thus the prediction accuracy of the model can be found to be low.

6.2.4 Evaluation of model errors

In order to achieve uniformity and facilitate comparative analysis of the models, mean square error (MSE) and standard deviation (SD) continue to be chosen as evaluation indicators for judging the performance of each prediction model. the formulae for calculating MSE and SD are shown in Equation (6.11), Equation (6.12).

Then it can be calculated that the MSE value of the GM(1,1) model is 0.892 and the SD value is 0.479.

6.3 Comparison of models

In summary, the test value and error value of the two models are compared as shown in Table 6.1 below

Table 6.1 Comparison of model errors

Model	MSE	SD
GM(1,1)	0.892	0.479
Quadratic Exponential Smoothing Forecast Model	0.128	0.357

Then there is Table 6.1. The accuracy of the quadratic exponential smoothing prediction model is better than that of the GM(1,1) model, and its prediction stability is also better than that of the GM(1,1) model. At the same time, by observing the image, it can be found that the degree of fit is relatively high. Then the monthly precipitation data of Changsha from 2009 to 2020 is substituted into the quadratic exponential smoothing prediction model to predict the data for the next two years, that is, the total of 2021 and 2022 24 months of data, and then substituting the data into the MK model of question 1, analyze its mutation points.

6.4 Forecast and analysis of extreme rainfall

Substituting 144 data of 12 years into the quadratic exponential smoothing prediction model for prediction, the results obtained are shown in the following table6.2.

Table 6.2 Predicted values of extreme rainfall

Year	Predictive value	Year	Predictive value
2021.1	38.13	2022.1	20.73
2021.2	78.47	2022.2	130.26
2021.3	161.01	2022.3	182.61
2021.4	199.10	2022.4	234.36
2021.5	307.03	2022.5	149.85
2021.6	300.31	2022.6	167.70
2021.7	77.18	2022.7	224.57
2021.8	156.09	2022.8	72.15
2021.9	229.15	2022.9	9.04
2021.10	68.90	2022.10	37.15
2021.11	62.88	2022.11	78.84
2021.12	147.49	2022.12	67.10

From Table 6.2, the predicted values for 2021 and 2022 can be obtained, then we substitute it into the MK model, and the results are shown in the following figure:.

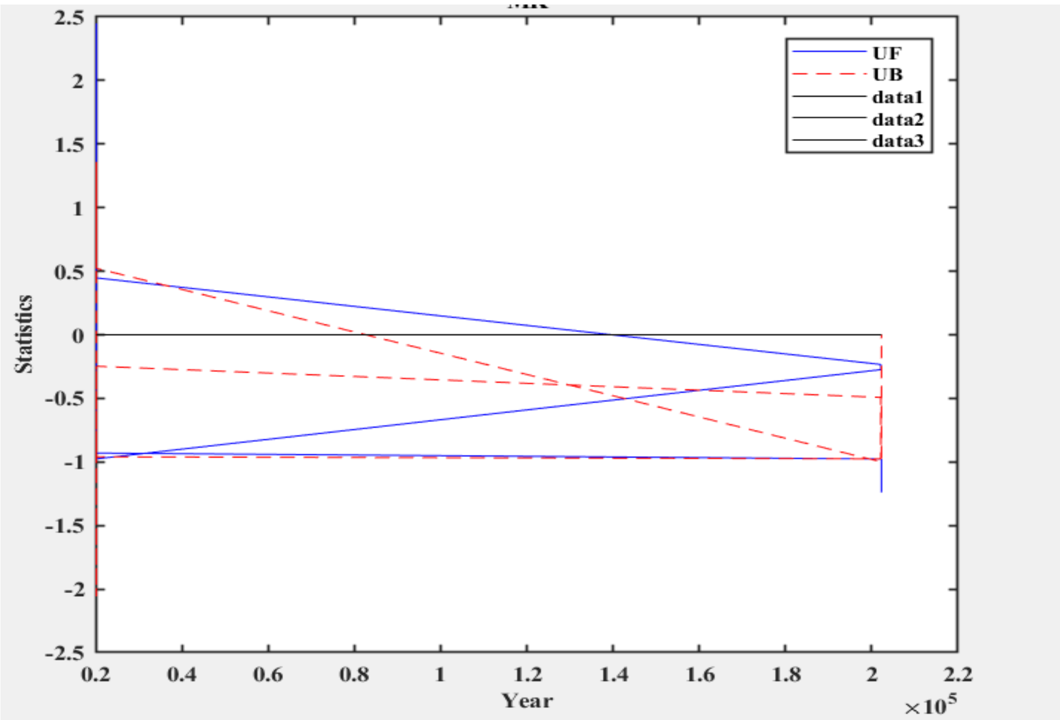

Figure 6.3 MK analysis of precipitation forecasts for Changsha

From Figure 6.3, it can be concluded that there are two sudden changes in Changsha in the forecast, which are May 2021 and April 2022. That is, Changsha will experience heavy rainfall in May 2021 and April 2022.

7 Analysis and answers to question 4

7.1 Cause analysis of Zhengzhou flood

From Question 1, it can be concluded that 2021 is the year of sudden change in precipitation in Zhengzhou. From Figure 4.5, it can be concluded that the data of Zhengzhou in 2021 after the variance processing can visually see the huge precipitation in July 2021. After the precipitation increased, Zhengzhou City was located on the edge of the Yellow River. The height of the Yellow River bed was higher than that of the plains near Zhengzhou, and it was impossible to rely on natural flow for drainage. Relying on electric water pumps for pumping and drainage, Zhengzhou's drainage pipe network owes too much. Caused the flood. Analyzing the data of the remaining indicators, the results that can be obtained are shown in the following table7.1.

Table 7.1 Related index data

time	TEMP	STP	WDSP
Apr	72.24	656.35	5.51
May	60.81	105.09	5.84
Jun	73.13	865.84	4.77
Jul	82.75	991.14	4.75
Aug	78.51	798.93	5.30
Sept	82.71	990.23	4.38
Oct	79.07	897.89	4.57
Dec	73.59	499.01	4.24

From Table 7.1, TEMP in July is 82.75, STP is 991.14, and WDSP is 5.30. It can be concluded that the temperature in Zhengzhou in July is higher, the air pressure at the site is higher, and the wind speed is lower. So to sum up, Zhengzhou's July floods are related to the sudden change of heavy precipitation, but also related to Zhengzhou's geographical location. The temperature is higher, the pressure is higher, and the cumulonimbus cloud is formed more quickly, and the wind speed is lower. , The cumulonimbus cloud cannot be blown away, resulting in concentrated and large rainfall. In addition, the drainage system is relatively weak, so a super flood is formed.

7.2 Shanxi flood analysis

For the situation of the October floods in Shanxi, the following reasons can be derived from the literature search information.

First, the atmospheric circulation situation is stable. The abnormally strong western Pacific subtropical high pressure was firstly extended westward and northward and then maintained steadily in the Yellow and Huaihua areas, which formed a stable east-high and west-low circulation situation in Shanxi with the low value system of the westerly wind belt, which was conducive to the appearance of prolonged precipitation weather in Shanxi.

Second, the water vapor conditions are abundant. The southerly airflow and low-level southwestern rapids from the west side of the subtropical high pressure will transport water vapor from the South China Sea and the Bay of Bengal northward through the southwestern region to the south-central region of Shanxi, providing an abundant source of water vapor for the continuous precipitation in Shanxi.

Finally, low-level lifting conditions are maintained for a long time. Under the stable weather situation, the low-level shear convergence system is maintained for a long time and precipitation echoes repeatedly pass through central Shanxi, superimposed on the complex topography of the Luliang Mountains and Taihang Mountains in Shanxi on the precipitation increase effect of the easterly airflow, leading to the appearance of extreme heavy precipitation in central Shanxi and the northern part of Linfen.

In order to form a comparison between the perspectives of Shanxi and Zhengzhou, the precipitation data of Taiyuan City in Shanxi from 2016 to 2020 were sought as a comparative analysis with Zhengzhou, the results of which are shown in Figure 7.1 below.

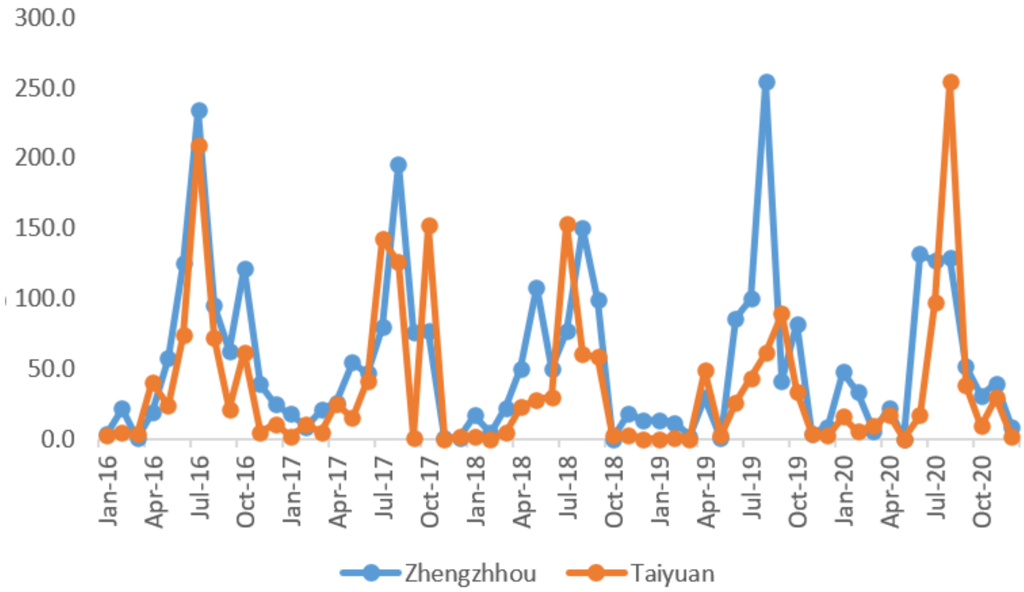

Figure 7.1 Comparison of precipitation data between Taiyuan City and Zhengzhou City

From Figure 7.1, it can be seen that the trends of precipitation in Taiyuan and Zhengzhou are roughly the same, with a significant increase in rainfall in Taiyuan from August to December each year, but relatively less rainfall on average throughout the year.

From this, it can be obtained that heavy rainfall in Shanxi in October has a strong probability of occurrence, i.e., there is an interval of sudden change points of precipitation. At the same time, the geographical conditions of Shanxi are complex, with more mountains and more thermal refluxes, while passing through the Yellow River basin, and the water content is particularly high when strong rainfall occurs.

In summary, the rainstorm characteristics of Zhengzhou in July and Shanxi in October are roughly the same, then the hazards for both are analyzed by the following.

7.3 Hazard Analysis

● **Impact of heavy rainfall in October in Shanxi.**

For one, the accumulated rainfall of heavy rainfall in Shanxi. Drainage system pressure is relatively large, while the ancient city of Shanxi has more cultural relics, extensive precipitation does not get drainage, the river in Shanxi Province, the river surge, geological disasters, traffic, cultural relics and other aspects have suffered different degrees of loss.

Second, the rainstorm lasted for a long time. The heavy rainfall caused a total of 76 counties (cities, districts) in 11 cities in Shanxi 1,757,100 people affected, 15 people died as a result of the disaster, 3 people missing, crop disaster area of 3,576,900 mu, 19,500 collapsed houses, 18,200 seriously damaged, direct economic losses of at least 5.029 billion yuan.

Third, the precipitation extremes are outstanding. A total of 59 national meteorological observation stations in the province daily precipitation exceeded the historical extremes of the same period since the establishment of the station, 63 national meteorological observation stations process cumulative precipitation exceeded the historical extremes of the same period.

Fourth, strong precipitation in Shanxi accompanied by cooling, causing cold waves, bringing inconvenience to rescue.

● **Impact of heavy rainfall in Zhengzhou in July.**

1. The harm is particularly great. According to incomplete statistical news, Zhengzhou City, continued heavy rainfall weather, the city's people's lives and property security caused huge losses, production and life brought serious impact. As of 12:00 on July 23, according to preliminary statistics, 395,989 people were relocated, 44,209.73 hectares of crops were affected, direct economic losses of 65.5 billion yuan, flooding and secondary disasters caused by heavy rainfall have led to the death of 51 people.

2. The scope and intensity are particularly strong. Almost all areas of Zhengzhou City were subjected to the great plan of the mega-level rainstorm, and most public facilities were damaged to varying degrees, while almost all railroad operations were halted, including high-speed railroad traffic before and after the disaster, and the Zhengzhou air port was halted, with losses caused by the stagnation of domestic consumption.

3. High-intensity rainfall, the temperature did not drop significantly, but locally increased.

● **Differences between.**

The scale and hazard of the July rainstorm in Zhengzhou was larger than the October rainstorm in Shanxi, while the high temperature during the rainstorm in Zhengzhou caused difficulties in high temperature rescue. It is also located in the low-lying area of the Yellow River basin, with more plains and large areas of standing water, resulting in huge losses in agriculture, tourism and transportation.

8 Analysis and Answers for Question 5

In the case of cities such as Zhengzhou, the damage caused in the event of a sudden rainstorm will be very large. Observing the precipitation in Zhengzhou, it can be concluded that the normal annual precipitation in Zhengzhou is relatively low, but the chance of strong rainfall is high, so when the city receives a lot of rainfall, it drains more than the water while also storing the right amount of water for use when there is less rainfall. Then it is recommended to build a sponge city, which is a city like a sponge with good adaptive capacity in the face of external environmental changes related to water. Specifically, the city can absorb, filter and save water when it encounters rainfall, preventing large amounts of standing water from appearing in the city; and when the city faces a water shortage problem, it can release this stored water to meet the city's normal water demand.

A sponge city does not mean rebuilding a new water supply and drainage system by overturning the existing one. Rather, it is to make use of the potential of the city itself to reduce and share the burden of the existing drainage system, while making the existing drainage system function as much as possible.

Sponge city drainage facilities, or sponge city in the "sponge body" is mainly the city's rivers, lakes, pools and other natural water systems, as well as parks, green spaces, permeable roads and other municipal facilities can drain. After the rainfall, rainwater through these facilities infiltration, storage, purification, reuse, and the last remaining part through the pipe network, pumping stations outside the drainage. This not only relieves the drainage pressure of the city's existing drainage system, but also enhances the efficiency of urban drainage.

The concept of sponge city has a very important guiding significance to the current urban construction in China, it can not only enhance the ability of the city to withstand heavy rainfall, but also regulate the ecological environment of the city and improve the livability of the city. The planning and construction of the sponge city is a comprehensive process, not led by a department or in someone to complete. It requires the coordination of multi-sectoral work. In the process of specific construction, must be from the actual situation in the region, scientific planning, fully consider all aspects of the situation, integrated planning, proper arrangements.

To sum up, we believe that the long-term construction planning of cities under extreme precipitation conditions can start with the construction of sponge cities, strengthening the existing drainage system to function, and at the same time, giving full play to the city's own potential in this area as far as possible, so as to reduce and share the burden of the existing drainage system. At the same time, the construction of vegetation and other "sponge bodies" can be strengthened to enhance the natural water storage and drainage capacity of the city, so that the city under extreme precipitation conditions will minimize economic losses and at the same time increase some economic benefits and promote the healthy development of the city.

9 Evaluation of the models

9.1 Advantages of the model

1. The MK analysis model is powerful, does not require the sample to follow a certain distribution, part of the missing data will not affect the results, is not disturbed

by a few outliers, is highly applicable, and is very effective for precipitation analysis, which can not only test the trend of the time series, but also test whether the time series has undergone sudden changes. Together with wavelet transform, the trend of precipitation in Zhengzhou can be perfectly derived.

2. The quadratic exponential smoothing prediction model can predict the future precipitation using limited historical data. While historical index data can be affected by complex factors, the model can adapt itself to changes in the data, while the model is more accurate and predicts more accurately, which is more in line with the requirements of the topic.

9.2 Disadvantages of the model

1. the MK analysis model process has no error comparison and lacks convincing power.

2. the quadratic exponential smoothing prediction model is not able to accurately consider the impact of influencing factors in the actual environment, i.e., it is not possible to perform multi-factor intervention.

References

[1] Zhao SQ, Fan W, Yu NY. Analysis of annual precipitation in Dongzhilu based on wavelet and MK tests [J]. Journal of Hebei Engineering University (Natural Science Edition). 1673-9469(2020)01-0084-07.
[2]Zuo Hongchao, Lv Shihua,Hu Yinqiao. Trend analysis of temperature and precipitation in China in the last 50 years[J]. Highland Meteorology, 2004(02):238-244.
[3]Cao Xiaomin,Liu Zhihong,Yang Zhaoming,et al. Spatial and temporal variability of rainfall in the central Loess Plateau[J]. Journal of Sichuan Normal University: Natural Science Edition, 2011,34(05):724728.
[4] FU G, CHEN S,LIU C, et al. Hydro-Climatic Trends of the Yellow River Basin for the Last 50 Years[J]. Climat-ic Change, 2004, 65(12):149-178.
[5] Xu Zongxue, Zhang Nan. Analysis of precipitation trends in the Yellow River basin over the last 50 years[J]. Geography Research, 2006(01):27-34.
[6]Zhang Jun. Algorithm design and analysis [M]. Tsinghua University Press. 2011.8. P100-P215
[7] Deng Julong. Gray prediction and decision making [M]. Huazhong University of Technology Press, 1986.
[8]Yang H. L.,Liu J. X.,Zheng B. Improvement and application of the gray prediction GM(1,1) model [J]. The practice and understanding of mathematics,2011,41:39-46.
[9] Data reference: http://data.cma.cn/analysis/yearbooks.html

Appendix

```
%% MannKendall .m%%
function [ UF,UB ] = MannKendall( x,y,p )

% x±íÊ¾Ê±¼äÈç1982-2015

% y±íÊ¾¶ÔÓ¦Ê±¼äµÄ½á¹û

% p±íÊ¾ÏÔÖøË®Æ½

N = length(y);

UF = SMK(y)

yy = reshape(y,1,length(y));

yy = fliplr(yy);

UB = -fliplr(SMK(yy))

zp(1:length(UF)) = norminv(p/2);

%return

h = figure;

set(h,'position',[100 100 600 500]) %100
100Í¼Ïñ×óÏÂÂµã±íÊ¾ÔÚÆÁÄ»µÄÎ»ÖÃ£¬400ºÍ300·Ö±ð±íÊ¾Í¼ÏñµÄ³¤ºÍ¸ß

if isempty(x)

plot(UF,'b');

hold on

plot(UB,'r--');

hold on

legend('UF','UB')

plot(abs(zp),'k');
```

```
hold on

plot(-abs(zp),'k');

hold on

plot(zeros(1,N),'k');

else

plot(x,UF,'b');

hold on

plot(x,UB,'r--');

hold on

legend('UF','UB')

plot(x,abs(zp),'k');

hold on

plot(x,-abs(zp),'k');

hold on

plot(x,zeros(1,N),'k');

end
xlswrite('G:\matlab2017\bin\KingModel\MK.xlsx',UF,'Sheet2','A1');
xlswrite('G:\matlab2017\bin\KingModel\MK.xlsx',UB,'Sheet2','A2');
set(gca,'linewidth',1,'fontsize',10,'fontname','Times New
Roman','FontWeight','bold'); %ÉèÖÃ×ÖÌå¡¢Ïß¿í¡¢¼Ó´Ö

xlabel('Year','Fontname', 'Times New Roman','FontSize',10,'FontWeight','bold')

ylabel('Statistics','Fontname', 'Times New Roman','FontSize',10,'FontWeight','bold')

title('MK','Fontname', 'Times New Roman','FontSize',10,'FontWeight','bold')
%%SMK.m
function U = SMK( Y )
```

```
N = length(Y);

s = zeros(1,N);

U(1) = 0;

for k=2:N

r = 0;

s(k) = 0;

for j=1:k-1

if Y(k)>Y(j)

r = r+1;

end

s(k) = s(k-1)+r;

end

E = k*(k-1)/4;

VAR = k*(k-1)*(2*k+5)/72;

U(k) = (s(k)-E)/sqrt(VAR);

end
```

```
%% Quadratic exponential smoothing forecasting model%%
function [Y,S1,S2,a,b] = expsmooth2(Yt,alpha,t)
%Yt:原时间序列；alpha:平滑系数；t:预测时长
%Y:预测值；S1/S2:一次/二次指数平滑值；a/b:预测公式参数
n=length(Yt);
%计算一次指数平滑值
S1(1)=Yt(1);
for i=2:n
    S1(i)=alpha*Yt(i)+(1-alpha)*S1(i-1);
end
```

```
%计算二次指数平滑值
S2(1)=S1(1);
for i=2:n
    S2(i)=alpha*S1(i)+(1-alpha)*S2(i-1);
end
%计算参数 a 和 b
for i=1:n
    a(i)=2*S1(i)-S2(i);
    b(i)=alpha/(1-alpha)*(S1(i)-S2(i));
end
%计算预测值 Y
for i=1:t
    Y(i)=a(n)+b(n)*i;
end
%绘图
plot(1:n,Yt,(n+1):(n+t),Y,'*');
end
```

```
%%GM(1,1)
y=input('请输入数据');
n=length(y);
yy=ones(n,1);
yy(1)=y(1);
for i=2:n
    yy(i)=yy(i-1)+y(i)
end
B=ones(n-1,2);
for i=1:(n-1)
    B(i,1)=-(yy(i)+yy(i+1))/2;
    B(i,2)=1;
end
BT=B';
for j=1:(n-1)
    YN(j)=y(j+1);
end
YN=YN';
A=inv(BT*B)*BT*YN;
a=A(1);
u=A(2);
t=u/a;
t_test=input('输入需要预测的个数');
i=1:t_test+n;
```

```
yys(i+1)=(y(1)-t).*exp(-a.*i)+t;
yys(1)=y(1);
for j=n+t_test:-1:2
    ys(j)=yys(j)-yys(j-1);
end
x=1:n;
xs=2:n+t_test;
yn=ys(2:n+t_test);
plot(x,y,'^r',xs,yn,'*-b');
det=0;
for i=2:n
    det=det+abs(yn(i)-y(i));
end
det=det/(n-1);
disp(['百分绝对误差为：',num2str(det),'%']);
    disp(['预测值为：',num2str(ys(n+1:n+t_test))]);
```