Problem Chosen	2021 ShuWei Cup Summary Sheet	Team Control Number
B		202111077005

Quantitative analysis of extreme rainfall

Summary

2021 is an extraordinary year. After COVID-19 swept through China and was effectively treated, extreme rainfall and extreme flooding events have again played out in tragic scenes, extreme rainfall events occur that humans cannot control, only by taking more precautions can we reduce the loss as much as possible.

For question one, divide the problem into two parts. **Part I:** Using **nonlinear regression model**, the relationship between precipitation at three stations and years was plotted, the years with higher precipitation were observed as follows: **1973**, **1999**, **2016**, **2017**, **2019** . **Part II**: First, use A to determine the weight of the index. Secondly, B is used for feature estimation. Finally, C is used to analyze the maximum correlation. Then the quantitative analysis of 2021 Zhengzhou flood event is obtained.

For question two, to solve problem two, after searching a large number of website resources(Web sites are listed in **Appendix I**) , found the average annual precipitation of Beijing, Guangzhou, Wuhan and Xi'an in the past 15 years. To simply predict the rainfall trend of the four cities in recent years, the **trend prediction model** is established in the second question. The **curve fitting** principle is used to draw images to observe and analyze the future precipitation trend of the four cities.

For question three, this paper uses the predictive method to test to solve the third problem. Error analysis of three forecasting methods is obtained. The average error of **Time series**, **ELM** and **K-means-Fuzzy Markov** is **0.16958**, **0.12644** and **0.05608** respectively. After comparison, **K-means-Fuzzy Markov** is the best prediction scheme.

For question four, the **evaluation-analogy model** is designed. Set evaluation levels and evaluation rules, after making **polygon scoring graph**, the conclusion is drawn: The precipitation of Zhengzhou storm event is much larger than that of Xi'an. The affected area is also far beyond Xi'an. However, the heavy rain in Xi'an lasts for a long time, the loss of cultural relics and energy is huge.

For question five, the main cause of urban flood disaster in China is the use of **mixed sewage system**. Rainwater and sewage do not separate transfer. However, if all cities were to switch to **distributary sewer systems**, it would cause huge costs. Therefore, this paper proposes to reform and improve the sewer system. In the sewer, the sewage **diversion pump** is established. At ground level, more culverts are built for drainage, in order to respond to emergency extreme rainfall events.

Keywords: **Nonlinear regression** **K-means-Fuzzy Markov** **ELM** **Time series** **Distributary sewer systems**

Content

Content....2
1. Introduction....1
1.1 Background....1
1.2 Work....1
2. Problem analysis....2
2.1 Data analysis....2
2.1.1 Information processing....2
2.1.2 Data preprocessing....3
2.2 Analysis of question one....3
2.3 Analysis of question two....4
2.4 Analysis of question three....5
2.5 Analysis of question four....6
2.6 Analysis of question five....6
3. Symbol and Assumptions....6
3.1 Symbol Description....6
3.2 Fundamental assumptions....6
4. Model....7
4.1 Question one....7
4.1.1 Model preparation....7
4.1.2 Design and Implementation....7
4.1.3 Result....10
4.2 Question two....11
4.2.1 Model preparation....11
4.2.2 Design and Implementation....12
4.2.3 Result....12
4.3 Question three....13
4.3.1 Model preparation....13
4.3.2 Design and result....16
4.4 Question four....16
4.4.1 Model preparation....16
4.4.2 Design and Implementation....17
4.4.3 Result....18
4.5 Question five....19
4.5.1 Model preparation....19
4.5.2 Result....19
5. Strengths and Weakness....20
5.1 Strengths....20
5.2 Weakness....20
Proposals for improvements to the sewer system....21
References....22
Appendix....23
Appendix I....23
Appendix I....23

1. Introduction

1.1 Background

Over the past half century, under the background of global climate warming, global extreme rainfall has shown an increasing trend in most regions, with temperature rising and glacier melting, aggravating the occurrence of extreme precipitation events. Extreme rainstorm is characterized by heavy precipitation and strong impact force. Due to the modernization of cities, the development of industry and agriculture, heavy rain is easy to collapse mountains, dam collapse, increase the flow of rivers and lakes. Therefore, it is determined that the probability of natural disasters such as debris flow and flood caused by heavy rain is very large. Once natural disasters occur, human society will also be greatly affected. The chain reaction of extreme rainstorms is shown in Figure 1.1.

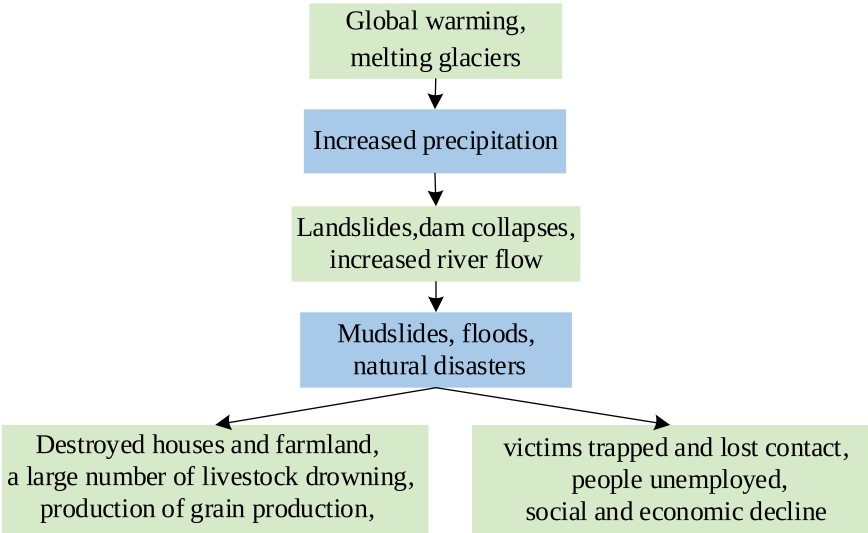


Figure1.1 Knock-on effects of extreme rainstorms

Since extreme rainstorm is difficult to control, in order to ensure the normal development of human society and reduce losses and casualties, precipitation prediction models are established to predict the probability of extreme rainfall in the next few years or even decades. Therefore, prevention in advance and timely preparation of emergency measures can be done at present, which is also an important way to help prevent great losses.

1.2 Work

Read the material, consider the background and information given by the problem, and build a mathematical model to solve the following problems:

(1) Firstly, this paper will analyze the correlation between annual precipitation and annual variation in Zhengzhou, select years with higher precipitation. Secondly, a quantitative analysis of Zhengzhou flood events in 2021 is made.

(2) Collect and sort out the precipitation data of more cities in China for many years, and analyze the precipitation trend of each city. Question two requires a clear explanation of data sources and methods of acquisition.

(3) Question 3 according to the precipitation data of each city in Question 2, different methods are used to predict and analyze the cities with extreme rainstorm phenomenon in the future.

(4) This paper will compare the characteristics of heavy rain in Zhengzhou in July 2021 with those in Shanxi in October 2021, and compare the difference of loss caused.

(5) In this paper, the long-term construction planning of future cities under extreme precipitation conditions is proposed, and several typical cities in China are analyzed in depth.

2. Problem analysis

2.1 Data analysis

2.1.1 Information processing

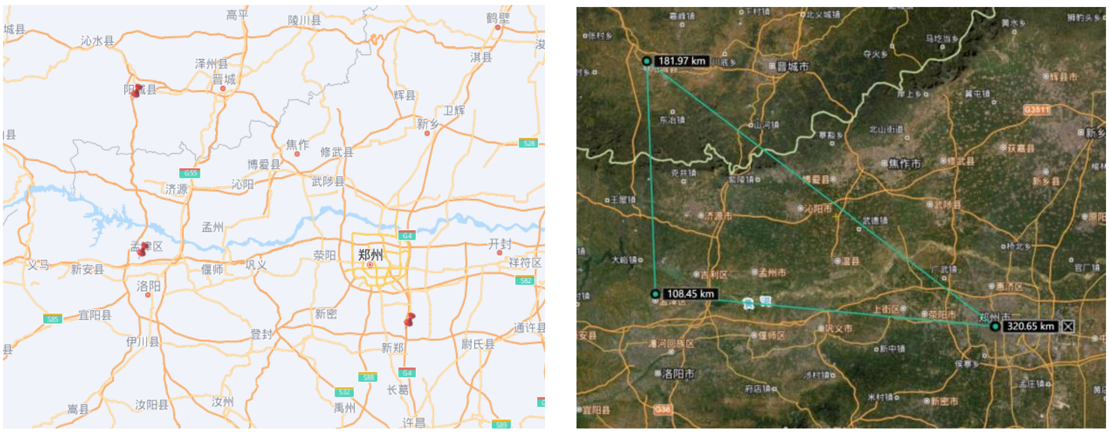
The problem is given in Annex 1. Annex 1 provides the daily precipitation observation data of three weather stations near Zhengzhou in the past 70 years. According to the Website [Listing 1.1], the three weather stations are shown in Table 1 respectively.

Table 1 Station code

Station	Station code
Yangcheng	57083099999
Mengjin	57071099999
Zhengzhou	53975099999

Among them, from top to bottom represents the data in Table 1, Table 2 and Table 3 respectively.

The specific locations of the three weather stations and Zhengzhou urban area are accurately positioned in Google Map [Listing 1.2] and WheatA [Listing 1.3], as shown in Figure 2.1.



a. Geographic location of 3 stations　　　　b. Distance between 3 stations

Figure2.1 The geographical position

In Annex I, the expressions and definitions of everyday elements related to precipitation are referred to the website [Listing 1.4], as shown in Table 2.

Table2 Symbol definition and description of precipitation elements

Symbol	Description	Symbol	Description
DEWP	Average dew point	PRCP	Precipitation
FRSHTT	Bad weather	MXSPD	Maximum sustained wind speed
MAX	Highest temperature	WDSP	Average wind speed
MIN	Lowest temperature	SNDP	Snow depth
TEMP	Average temperature	VISIB	Average visibility
SLP	Mean sea level pressure	STP	Average site pressure

2.1.2 Data preprocessing

When processing the data given in Annex I, we found that many data were ‘too large’. This paper regarded such ‘too large’ data as unmeasured data, and ignored some data processing in quantitative analysis for the convenience of solving problems. Table 3 shows the data that should be taken into account and the data that should not be taken into account temporarily.

Table 3 Description of Weather Symbols

Important data to consider	Data not considered for the time being
PRCP	FRSHTT
DEWP	GUST
MAX	SLP
MIN	SNDP
TEMP	STP
WDSP	VISIB

Among them, snow depth is also one of the important characteristics of precipitation, but it is not considered in the flood world, but in the case of extreme precipitation, snow depth is also an important indicator that needs to be taken into key consideration.

2.2 Analysis of question one

This paper divides the question one in the title into two parts:

Part I: The correlation between precipitation characteristics and annual transformation characteristics in Zhengzhou was analyzed, and the years with higher precipitation were screened out.

This paper will answer the question according to the following process, the general process is shown in Figure 2.2.

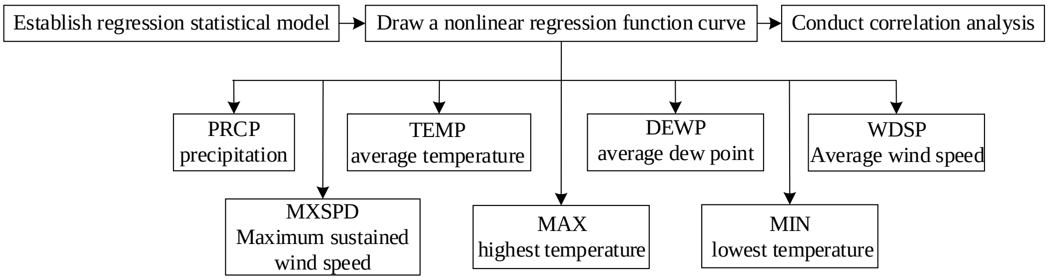


Figure 2.2 Regression model flow diagram

Step1 Establish a nonlinear regression model.

Step2 The nonlinear regression function images of each precipitation feature with year as variable are drawn (Precipitation characteristics of nonlinear function images to be drawn are shown in Figure 2.2.

Step3 The correlation between each of precipitation characteristics and annual transformation characteristics was analyzed.

Part II: The flood events in Zhengzhou in 2021 were analyzed quantitatively.

This part will establish **the quantitative analysis model of precipitation events**. The model building process is shown in Figure 2.3.

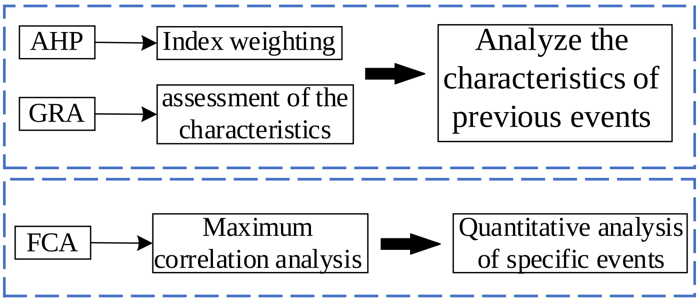


Figure 2.3 Quantitative analysis model

Step1 Using **Analytic Hierarchy Process (AHP)** and **Grey Relational Analysis (GRA)** method to analyze the maximum correlation of precipitation characteristics in Zhengzhou in previous years.

Step2 Based on the data presented in previous years, the **Fuzzy Comprehensive Assessment (FCA)** is used to make a specific quantitative analysis of the flood events in Zhengzhou in 2021.

2.3 Analysis of question two

In the second question, this paper first found the national precipitation distribution map on the website [Listing 1.5] after consulting a large number of data.

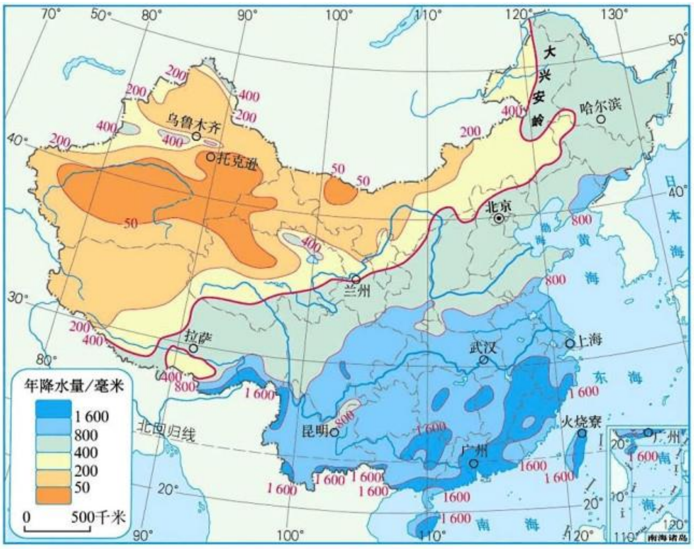


Figure 2.4 National precipitation distribution map

By observing the national precipitation distribution map, this paper intends to select Beijing, Xi’an, Guangzhou and Wuhan to collect precipitation data.

In order to analyze the rainfall trend of four cities, **the trend prediction model** was established, and the linear fitting and regression prediction were made according to the rainfall data of each city. Then get the rainfall trend of each city.

2.4 Analysis of question three

After consulting a large number of data, this paper intends to establish a prediction method test model, using three prediction methods for prediction, by predicting the precipitation in Beijing in the past few years, comparing the predicted value with the actual value to observe which method is the best, and then using this method to predict the future precipitation of the four cities.

The prediction method test model flow is shown in Figure 2.5.

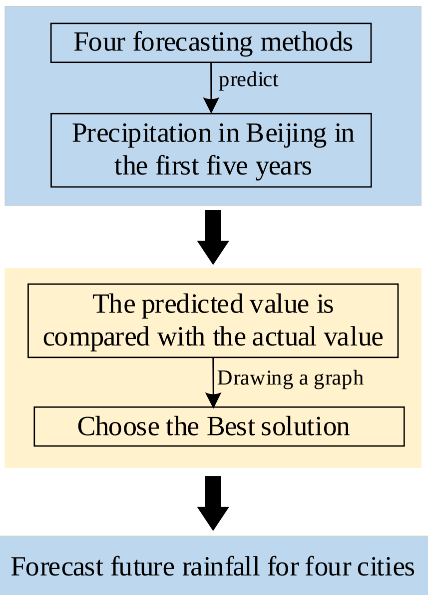


Figure 2.5 Prediction method test model

The three precipitation prediction methods to be used are listed below:
(1) K-means-Fuzzy Markov.
(2) Time series prediction model.
(3) Extreme learning machine prediction model.

2.5 Analysis of question four

The fourth question is to compare the characteristics and loss of heavy rain between Zhengzhou and Shanxi. In question 4, **the evaluation-analogy model** of evaluation characteristics and loss is constructed. Firstly, the characteristics and losses of heavy rain in Zhengzhou and Shanxi are listed one by one. By assuming evaluation grades, polygonal scoring maps are made for display and specific analysis.

The following known conditions are given:
(1) 2021/7/18-2021/7/21, the average cumulative precipitation in Zhengzhou was 449mm.
(2) 2021/7/20, in one hour, Zhengzhou station received 201.9mm of rain, exceeding the maximum rainfall per hour on land in China.
(3) The average annual precipitation of Zhengzhou is 640.8mm.

2.6 Analysis of question five

The key point of the third problem is to propose a long-term underground pipeline construction plan suitable for extreme precipitation conditions. As China's terrain is complex and diverse, it often needs to be analyzed in specific cases. In this paper, the underground pipeline planning and design is carried out in Guangzhou area which has a lot of rainfall.

Three aspects should be paid attention to in underground pipeline design:
(1) Satisfy the life and activities of the community.
(2) Create a comfortable and safe water environment.
(3) Good for the development of the region.

In order to cope with the extreme rainfall that may occur in the future, the size of the underground pipe must be able to handle a large volume of rainwater without hindering the normal activities of the citizens.

3. Symbol and Assumptions

3.1 Symbol Description

Symbol	Description
w_j	Weight
μ_{ij}	Correlation
ε_{ij}	Correlation coefficient
Δ	Matrix element

3.2 Fundamental assumptions

In order to make the model more scientific and reasonable, this paper puts forward the following hypotheses:

(1) The dependent variable is regarded as a random variable, that is, when the independent variable takes a fixed value in the definition domain, the observed value of the dependent variable is random and obeys a certain probability distribution.

(2) When an independent variable takes a fixed value in its domain, the function of the independent variable takes the average of all possible observations of the dependent variable.

(3) The random errors are normally distributed $N\left(0,\sigma^2\right)$, independent of each other. Also, variance is a constant independent of independent and dependent variables.

(4) It is assumed that the year with average annual precipitation of 2000mm is recorded as the year with high precipitation.

4. Model

4.1 Question one

4.1.1 Model preparation

Part I: Nonlinear regression model

$$y=\mu+\varepsilon, i=1,2,...,\mathrm{n} \tag{4.1}$$

$$\mu=\overline{y}=f\left(x,\beta_1,...,\beta_k\right) \tag{4.2}$$

Where, x is the independent variable, y is the dependent variable, ε is the random error, μ is the function of x, $\beta_1,...,\beta_k$ is the regression coefficient(k undetermined parameters contained in μ).

Part II: Quantitative analysis model of precipitation events

In this paper, fuzzy comprehensive evaluation is used to reduce the fuzziness and randomness of factors affecting precipitation, **GRA** is used to consider the transmissibility of the interaction between factors and improve the evaluation efficiency, and **AHP** is used to show the correlation factors of precipitation characteristics.

4.1.2 Design and Implementation

(1) Analytic Hierarchy Process

The flow chart of hierarchy analysis is shown in 4.1.

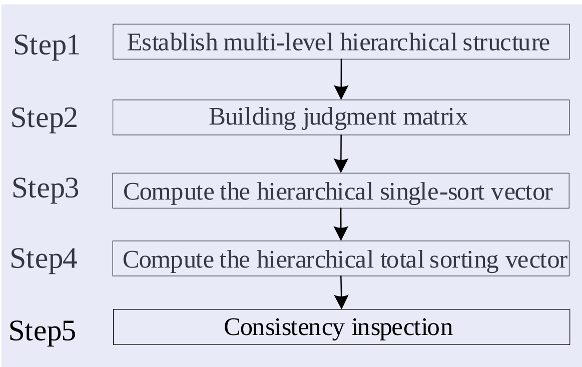


Figure 4.1 Flow chart of AHP

AHP method is used to determine the weighting of indicators as follows:

Step1 Establish a multi-level hierarchical structure, as shown in Figure 4.2. Firstly, the indicators are divided into the top layer (target layer), middle layer (criterion layer) and bottom layer (scheme layer) from top to bottom.

Step2 Establish the judgment matrix. The comparative judgment matrix is established to obtain the relative importance of the indicators at the next level through the comparison between indicators, and the experts score the indicators according to the literature.

Step3 Compute the hierarchical single-sort vector.

Step4 Compute the hierarchical total sorting vector.

Step5 Consistency inspection.

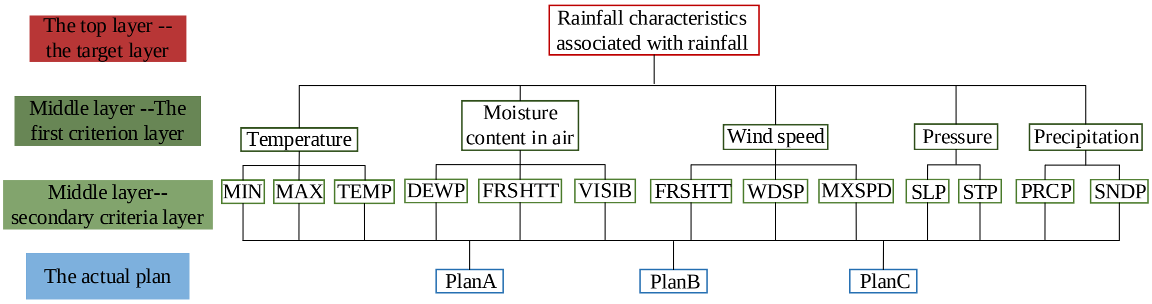


Figure 4.2 Hierarchical structure of analytic hierarchy Process

(2) Grey Relational Analysis (GRA)

The flow chart of GRA is shown in Figure 4.3.

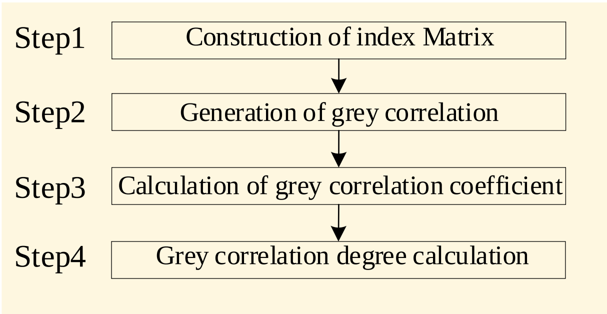


Figure 4.3 Flow chart of GRA

The GRA method was **used for feature evaluation**, complete the following steps

Step1 Construct the indicator matrix. The evaluation indicator matrix is shown as follows:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \tag{4.3}$$

Where, n represents the number of schemes, m represents the number of indicators, and x_{ij} represents the value of the j indicator of the i scheme.

Step2 Generation of grey correlation. Positively, indexes are divided into positive type (the higher the evaluation value is, the higher the corresponding index rainfall), negative type (the smaller the evaluation value is, the higher the corresponding index rainfall), and moderate type (the moderate evaluation value is, the higher the corresponding index rainfall). All indicators can be covered in the interval [0,1] through normalization.

The grading treatment formula of positive indicators is as follows:

$$\mu_{ij}^{+}\left(x_{ij}\right)=\frac{x_{ij}-\min_i\left\{x_{ij}\right\}}{\max_i\left\{x_{ij}\right\}-\min_i\left\{x_{ij}\right\}} \tag{4.4}$$

The grading treatment formula of negative indicators is as follows:

$$\mu_{ij}^{-}\left(x_{ij}\right)=\frac{\max_i\left\{x_{ij}\right\}-x_{ij}}{\max_i\left\{x_{ij}\right\}-\min_i\left\{x_{ij}\right\}} \tag{4.5}$$

Among them, μ_{ij} is suitable for evaluating the similarity degree of each feature, and the larger μ_{ij} is, the rainfall is more influenced by this rainfall feature.

Step3 Grey correlation coefficient calculation. The difference matrix element Δ is calculated as:

$$\Delta_{ij}=\left|\mu_{0j}-\mu_{ij}\right| \tag{4.6}$$

Find $\Delta_{\min}$ and $\Delta_{\max}$ in Δ.Thus, the grey correlation coefficient between μ_{ij} and the preset reference value μ_{0j} can be calculated by equation (4.7).

$$\varepsilon_{ij}=\frac{\Delta_{\min}+\rho\Delta_{\max}}{\Delta_{\mathrm{ij}}+\rho\Delta_{\max}} \tag{4.7}$$

Among them, the resolution coefficient used to adjust the range of grey correlation coefficient ρ is usually 0.5.

Step4 Grey correlation degree calculation.

$$r_i=\sum_{j=1}^{m}w_j\varepsilon_{ij} \tag{4.8}$$

Among them, ε_{ij} represents the grey correlation coefficient, and w_j represents the total sorting weight. The precipitation features are sorted according to the level of correlation degree r_i, from which we can see which rainfall feature has the largest correlation degree with rainfall.

(3) Fuzzy Comprehensive Assessment (FCA)

The flow chart of **FCA** is shown in Figure 4.4.

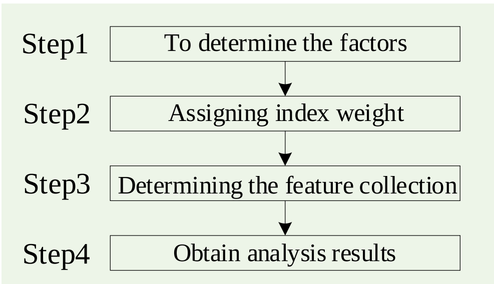


Figure 4.4 Flow chart of FCA method

The steps of **maximum correlation analysis** using FCA method are as follows:

Step1 To determine the factors.

Step2 Assigning index weight. Using AHP method, the weight of each index can be obtained.

$$W=\left[w_1,w_2,...,w_n\right] \tag{4.9}$$

Where, w_i represents each feature weight, $\sum_{i=1}^{n} w_i = 1$.

Step3 Determine the analysis set, build the evaluation set based on actual requirements, there are four grades: irrelevant, membership degree, moderate, and membership. Qualitative transformation of grade into hundred points, so that the evaluation results are easy to distinguish: [0, 40] indicates irrelevant, [40,60] indicates membership degree,[60,80] indicates moderate, and [80,100] indicates membership. The median value of the grade interval was taken and the correlation evaluation set was quantified as follows:

$$V = [30, 50, 70, 90] \tag{4.10}$$

Step4 Obtain analysis results.

$$E = W \circ R \tag{4.11}$$

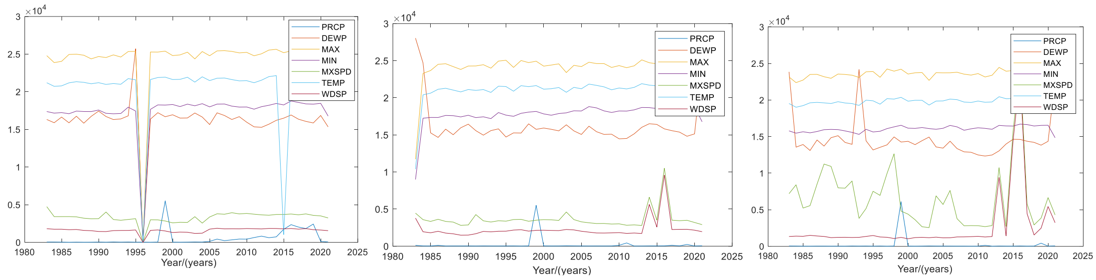
Where, represents the feature weight obtained by analytic hierarchy process, and represents the correlation degree obtained by grey correlation method.

The analysis results are obtained, and formula (4.12) is as follows:

$$Z = EV^T \tag{4.12}$$

4.1.3 Result

After processing the data in Appendix 1, the nonlinear regression diagrams of precipitation characteristics of the three stations about annual transformation are obtained.

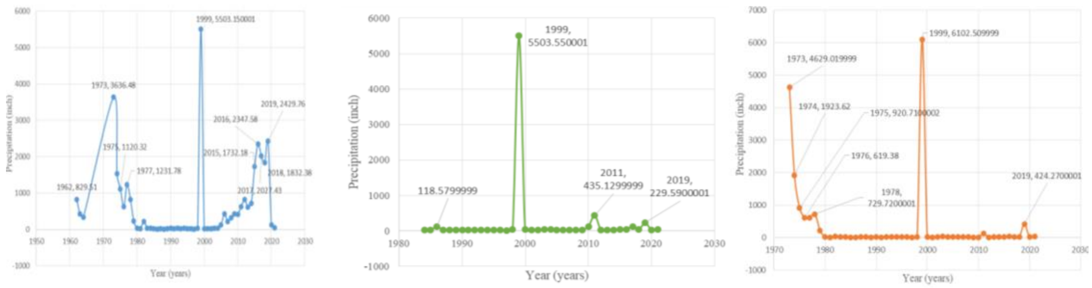


a.Yangcheng b. Mengjing c. Zhengzhou

Figure. 4.5 Nonlinear regression diagram of rainfall characteristics

Yangcheng: The indexes in Yangcheng area showed an extreme downward trend in 1996. After this extreme, the rainfall from 1998 to 1999 showed an obvious upward trend. In 2015, the average temperature showed an extreme decline, and precipitation showed an increasing trend in the following five years. **Mengjin:** From 1998 to 1999, the rainfall in Mengjin area showed an obvious upward trend, but there was little correlation with other indexes. **Zhengzhou:** the average dew point and maximum sustained wind speed increased sharply in Zhengzhou, and there was an obvious upward trend from 1998 to 1999. In 2015, the mean dew point and maximum sustained wind speed changed extremely again, and the rainfall also showed an upward trend after that, but it was not obvious.

Precipitation of the three stations in previous years is shown in Figure 4.6.



a. Yangcheng b. Mengjin c. Zhengzhou

Figure. 4.6 Annual precipitation at the three stations

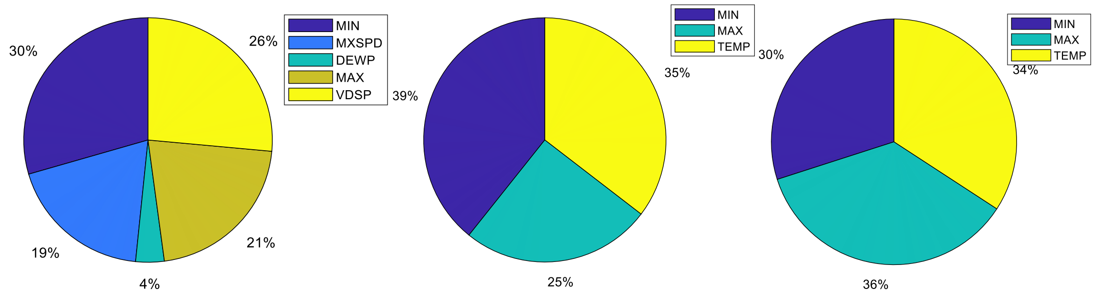
In this paper, it is assumed that the year with average annual precipitation of 2000mm is recorded as the year with high precipitation.

According to Figure 4.6, the years in Table 4 can be selected as the years with high precipitation.

Table 4 Years with higher rainfall

Station	Year(years)	Rainfall(mm)	Station	Year(years)	Rainfall(mm)
Yangcheng	1973	3636.48mm	Yangcheng	2019	2429.76mm
Yangcheng	1999	5503.15mm	Mengjin	1999	5503.55mm
Yangcheng	2016	2347.58mm	Zhengzhou	1973	4629.02mm
Yangcheng	2017	2027.43mm	Zhengzhou	1999	6102.51mm

Based on the calculation of quantitative analysis model, the pie chart of correlation between precipitation of three stations in previous years and rainfall indexes is obtained, as shown in Figure 4.7.



a. Yangcheng b. Mengjin c. Zhengzhou

Figure 4.7 Pie chart of precipitation characteristics at three stations.

According to the correlation degree of the three graphs in Figure 4.7, it is found that the cause of Zhengzhou flood event is temperature、visibility、average dew point、Maximum sustained wind speed。Among them, it has the greatest correlation with temperature. According to the conclusion in Figure 4.6, before the flood event in Zhengzhou in 2021, both the average dew point and the average wind speed in Zhengzhou show an extreme upward trend. Therefore, in addition to temperature changes, the judgment of flood events in Zhengzhou is greatly related to the extreme rising trend of average dew point and average wind speed in previous years.

4.2 Question two

4.2.1 Model preparation

The precipitation data

Refer to the precipitation data of Beijing [Listing 1.6], Xi’an [Listing 1.7], Guangzhou [Listing 1.8] and Wuhan [Listing 1.9]. The data sources are as follows:

(1) The precipitation data of Beijing comes from the official website of Beijing Municipal Bureau of Statistics, and the precipitation data from 1978 to 2020 are selected.

(2) The precipitation data of Xi’an comes from the official website of Xi’an Statistics Bureau, and the precipitation data from 2005 to 2019 are selected.

(3) The precipitation data of Guangzhou comes from the official website of Guangzhou Municipal Bureau of Statistics, and the precipitation data from 2005 to 2019 are selected.

(4) The precipitation data of Wuhan is from the official website of Wuhan Statistics Bureau, and the precipitation data from 2005 to 2019 are selected.

(The above official sites are listed in Appendix 1.)

Trend prediction model

The establishment of trend prediction model is shown in Figure 4.8.

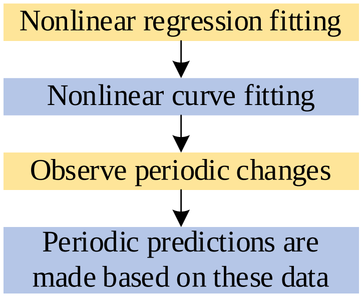


Figure 4.8 Trend prediction process.

According to the precipitation data, the nonlinear curve fitting is conducted to observe whether there is periodicity and analyze the future precipitation trend of each city according to periodicity.

4.2.2 Design and Implementation

Basic principles of curve fitting:

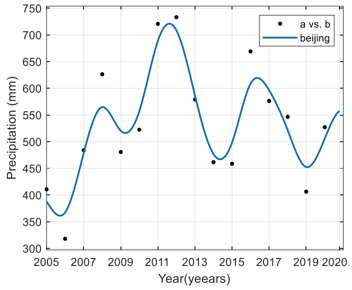
$$Q=\sum_{i=1}^{N}\delta_i^2 \tag{4.13}$$

$$\delta_i=\phi(x_i)-y_i, i=1,2,\dots,\mathrm{n} \tag{4.14}$$

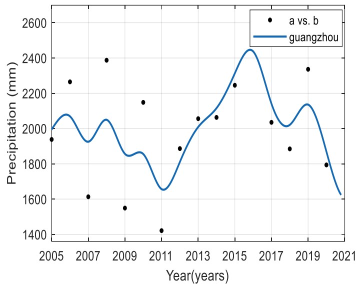
Where, x represents the independent variable, y represents the dependent variable, and there are n nodes (x_i, y_i) in total. Formula (4.13) is required to achieve the minimum, that is, the common principle of least square method.

4.2.3 Result

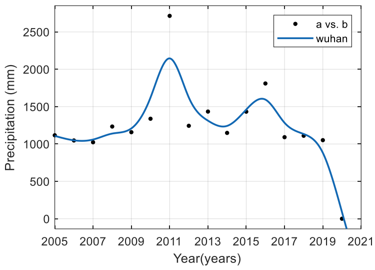
Curve fitting images of the four cities are shown in Figure 4.9.



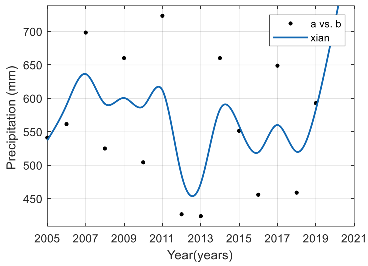
a. Beijing



b. Guangzhou



c. Wuhan



d. Xi’an

Figure 4.9 Curve fitting diagram of four cities

Beijing: According to Figure a in Figure 4.9, the average extreme rainfall occurs once every four years in Beijing, so the rainfall in 2021 in Beijing will be larger than that in 2020, but it will not exceed the rainfall in 2016 in terms of the general trend. **Guangzhou**: According to Figure b in Figure 4.9, the rainfall in Guangzhou will rise in 2021 but will not exceed the rainfall in 2019.**Wuhan:** According to figure c in Figure 4.9, the rainfall in Wuhan will increase in 2021 compared with that of the previous year, but it will not exceed that of 2019 in general. **Xi’an**: According to Figure d in Figure 4.9, the rainfall in Beijing will decrease in 2021, but it will not be lower than that in 2018.

4.3 Question three

4.3.1 Model preparation

(1) K-means- Fuzzy Markov model [4]

The model flow is shown in Figure 4.10.

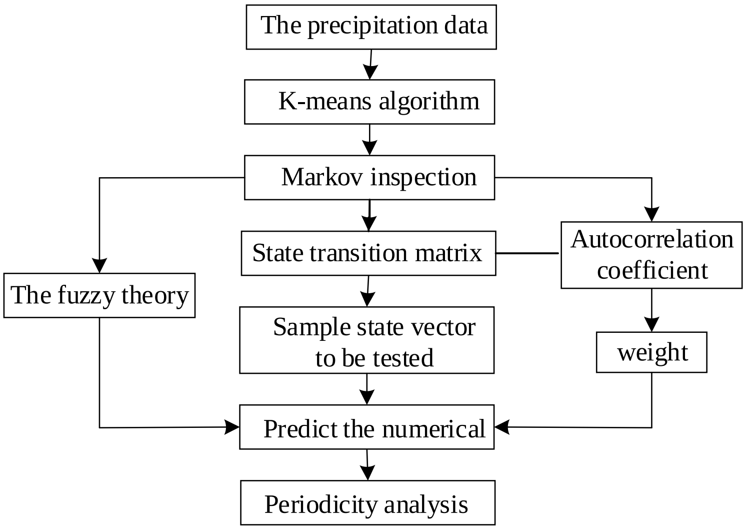


Figure 4.10 K-means- Fuzzy Markov model

The steps of Markov model are as follows.

Step1 K-means clustering analysis of precipitation data.

Step2 State transition probability matrix. According to the clustering result of step 1, the precipitation sequence is divided into several states, and the transition probability matrix of each state is obtained.

Step3 Mahogany test. Mahalanobis test is carried out to determine whether precipitation events have Markov properties.

Step4 Autocorrelation coefficient and weight. The normalized autocorrelation coefficients of each step are taken as the weights of Markov chain.

Step5 The precipitation state vector takes the past precipitation state as the initial state, extracts the row vector of the state transition probability matrix of each order, and takes the state corresponding to the maximum probability as the precipitation state at this stage.

Step6 The Fuzzy theory. Both the influence of maximum probability and other influencing factors of time lag probability should be considered.

Step7 Periodicity analysis. Markov chain has ergodicity, and its distribution is stationary when it reaches stationary state, and periodic analysis can be carried out at this time.

(2) Time series prediction model [1]

The flow diagram of the model is shown in Figure 4.11.

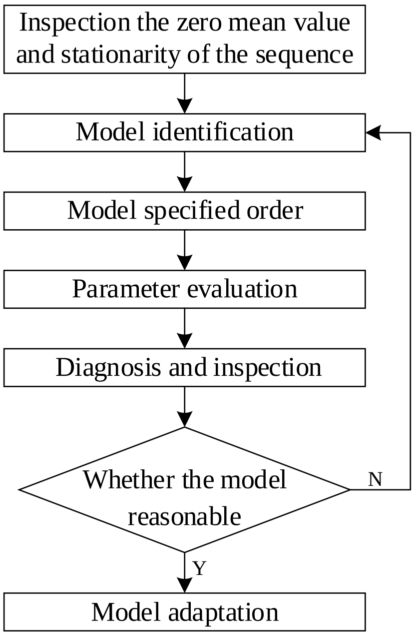


Figure 4.11 Model flow chart

Step1 Sampling. Time series dynamic data are extracted from the obtained data.

Step2 Drawing. According to the dynamic data for correlation graph, correlation analysis, find autocorrelation function.

Step3 Fitting. Identification of appropriate random model, curve fitting, urgent use of random model to fit the observation data of time series.

(3) Extreme learning machine [2]

The model flow is shown in Figure 4.12.

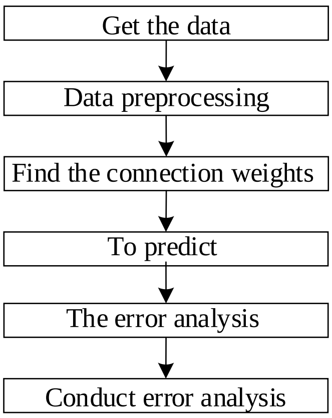


Figure 4.12 Model flow

Step1 Get the data.

Step2 Data preprocessing. In general, minimax normalization is adopted to eliminate data dimension and data level influence.

Step3 Find the connection weights. Train THE ELM model and work out the link weights of hidden layer and output layer.

Step4 Predictions. The trained model is predicted using the input data from the test.

Step5 Error analysis. Analyze the error between the predicted value and the actual value.

Step6 Drawing. After error analysis, draw a diagram for display.

4.3.2 Design and result

The predicted values obtained by each prediction method are shown in Table 5.

Table 5 Comparison between predicted value and actual value

	The actual value	Time series	ELM	K-means-Fuzzy Markov
2016	699.1	499.8	559.3	623.3
2017	576.2	497.7	497.7	556.7
2018	546.5	495.7	521.7	523.7
2019	406.3	493.7	493.6	450.2
2020	527.1	491.7	500.2	511.2

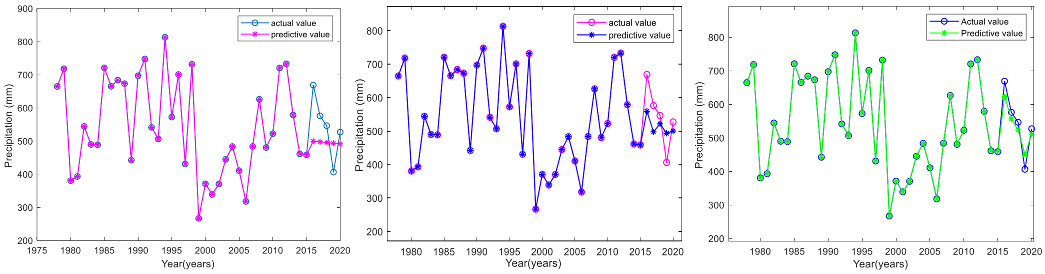
The relative error values of the predicted results obtained by error analysis of the three methods are shown in Table 6.

Table 6 Relative errors of predicted values

	Time series	ELM	K-means-Fuzzy Markov
2016	0.3388	0.1962	0.0734
2017	0.1576	0.1576	0.0350
2018	0.1025	0.0475	0.0436
2019	0.1769	0.1768	0.0974
2020	0.0721	0.0541	0.0310

As can be seen from the above table, the average error of Time series is **0.6958**.The mean error of ELM was **0.12644**.The average error of K-means-Fuzzy Markov is **0.05608**.It can be concluded that the data predicted by **K-means-Fuzzy Markov** is the most accurate.

In order to more intuitively represent the predicted value, matlab2019B version was used to draw the image for observation.



a. Time series b. ELM c. K-means-Fuzzy Markov

Figure. 13 Comparison of predicted and actual values

It can be seen from FIG. 13. a that time series prediction is almost inaccurate. It can be seen from Figure 13. b that there are two accurate values predicted by extreme learning machine. It can be seen from Figure 13. c that the **K-means-Fuzzy Markov** value is almost close to the actual value.

4.4 Question four

4.4.1 Model preparation

The following information was obtained by checking the website[Listing 1.10] information.

Table 7 Comparison table of rainfall characteristics between Zhengzhou and Xi’an.

Serial number	Characteristics	Zhengzhou	Xi'an
1	precipitation	449mm	302.2mm
2	range	39(pcs)	18(pcs)
3	Extreme point	51(pcs)	63(pcs)
4	Time of heavy rainfall	2(days)	4(days)
5	time of duration	6(days)	6(days)

Table 8 Comparison of losses between Zhengzhou and Xi’an

Serial number	damage	Zhengzhou	Xi’an
6	The number of the affected	128(ten thousand)	175.71(ten thousand)
7	The affected area	89	76
8	Crop loss	44209.73(hektare)	238000
9	Destroy the house	17015	19500
10	amount of loss	655(Hundred million)	50.29
11	Damage of cultural relics	400	1783
12	Shutdown project	910	1035
13	Number of closed scenic spots	50	166
14	Anomalous values	5	13
15	Transfer of personnel	395989	54947

The rating criteria

The evaluation scale is divided into five levels, and the two values are compared in pairs:

(1) Similar values, both rated third.

(2) There is a big difference in values, and it is rated as level two and level three.

(3) There's a big difference between two and three.

4.4.2 Design and Implementation

The steps of evaluation-analog model establishment are as follows.

Step1 The input features.

The characteristic indexes are expressed as

$$c_{ij}\left(i=1,2\right)\left(j=1,2,...,15\right) \tag{4.16}$$

Among them， $i=1$ representative Zhengzhou， $i=2$ representative Xi'an， j representative 15 indicators for comparison between the two cities.

Step2 Check whether the indicator is positive. If the indicator meets the requirement, it is a positive indicator; if the indicator does not meet the requirement, it is a negative indicator.

Step3 Index rating. Rating according to rating criteria. Among them, the larger the positive index value, the higher the series. The larger the negative index value, the lower the series.

Step4 Draw analogies. Output the rating and draw polygon analogies according to the rating.

Step5 Loss judgment. According to the above steps, the larger the area of the analog diagram is, the smaller the area of the analog diagram is, the larger the loss is.

4.4.3 Result

The evaluation results of 15 indicators in the two places are shown in Tableb7.

Table 9 Evaluation results

Indicators	Zhengzhou	Xi'an	Indicators	Zhengzhou	Xi'an
1	2	4	9	4	2
2	1	5	10	1	5
3	4	2	11	5	1
4	5	1	12	4	2
5	3	3	13	5	1
6	4	2	14	4	2
7	2	4	15	5	1
8	5	1			

According to the evaluation results, an analogy diagram can be drawn, as shown in Figure 4.14.

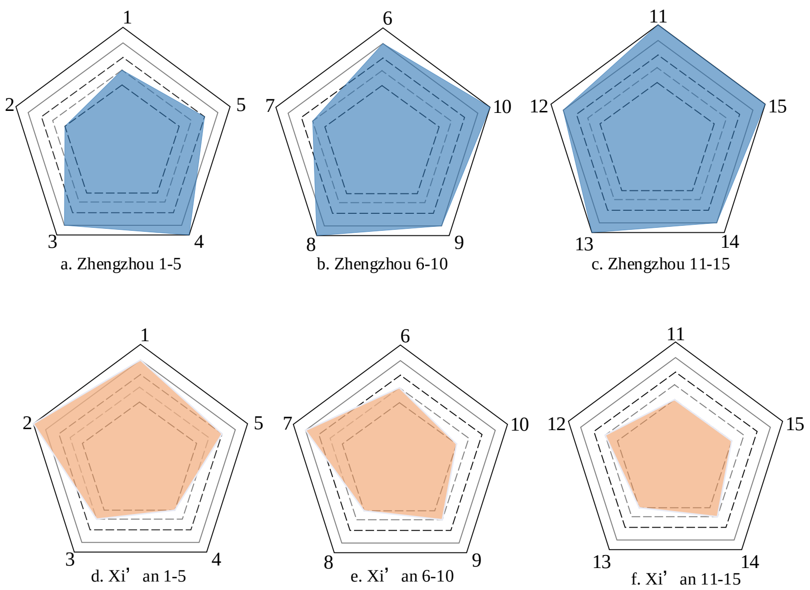


Figure4.14 Analogy of the figure

Analysis of rainstorm characteristics in two places:

Zhengzhou is much larger in terms of rainfall and rainfall range, but the number of stations that reach the extreme historical rainfall is not as many as Xi;an, and the duration of heavy rainfall is not as long as Xi'an, and the duration of rainfall in both places is roughly similar.

Difference analysis of loss between two places:

Zhengzhou is much better than Xi’an in terms of the damage area and direct losses. As can be seen from the analogical figure, the amount of crop loss, cultural relic damage, scenic spot closure and the gap between normal rainfall and extreme rainfall in Xi’an are much higher than that in Zhengzhou. The number of displaced people is much smaller than Zhengzhou, which is also responsible for the large number of victims. Xi’an is rich in cultural relics and has a lot of ancient architectural complexes. Therefore, there is a large amount of cultural relics damage and a large number of scenic spots are destroyed, leading to the closure of entrances, which is the biggest difference between Xi’an and Zhengzhou in the loss of objects.

4.5 Question five

4.5.1 Model preparation

Relevant policy items concerning urban pipeline planning and design are shown in Table 8.

Table 10 Related policy items

Tasks and Problems	Policy direction	Related projects
The use of water	Efficient use of water	Reclaimed water
Clean water	Sewer perfection	Sewer modification
Flood control	Comprehensive utilization of rainwater	Strategies for utilizing rainwater
Water in times of disaster	Fire Water assurance	River water, reclaimed water utilization

Most Chinese cities have combined sewer systems, under which sewage and rainwater are not separated. In the case of extreme rainfall, it is easy to flood. In order to ensure the normal drainage system, it is worthwhile to transform the combined sewer system into a split sewer system.

In a split sewer system, rainwater and sewage do not mix. Rainwater is discharged directly into the river or ocean, while sewage is sent through the sewer via self-gravity channels to sewage treatment plants.

4.5.2 Result

The architectural circulation system [5] in the city is shown in Figure 4.15.

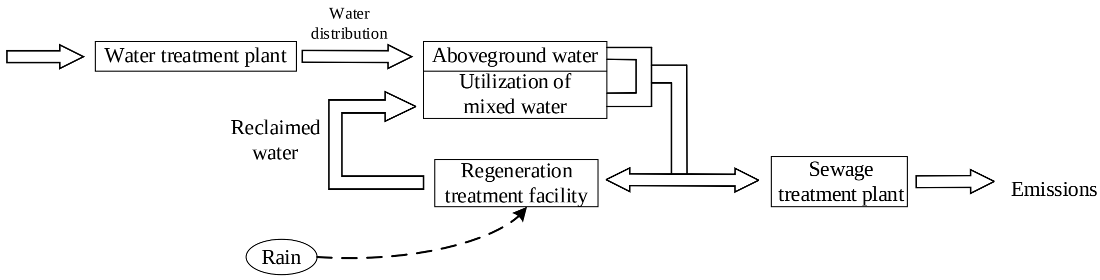


Figure 4.15 Building circulation system

This paper proposes the following four suggestions for the long-term planning of future cities facing extreme rainfall events:

(1) Because the sewerage system of our country has been formed initially when the city was established, the diversion sewerage system has many advantages, but if

it is changed completely, it will consume too much manpower and material resources. Therefore, it is the best method to reform and improve the original combined sewer system.

(2) Appropriate amount of rainwater will be collected and purified, used as the guarantee of domestic water and fire water, when the storage reaches the maximum, all into rivers and oceans.

(3) Because of the complex topography of our country, the rainfall situation is not unified. According to the specific situation, surface filtration channels should be opened in cities with heavy rainfall and low-lying areas, in areas with low rainfall, underground channels can be opened to filter above-ground water if conditions permit, and small flood discharge can be carried out when extreme rainfall occurs.

(4) Regular screening and clearing of passages is necessary.

(The remaining details are set out in the proposal letter.)

5. Strengths and Weakness

5.1 Strengths

(1) The establishment of rating rules makes the model more reasonable and quantifies the materials collected online, which makes the model have better reference value.

(2) The model has strong operability, and different reference schemes are put forward for different situations, which expands the application scope of the model to a certain extent.

(3) The model is simple and easy to use, and is suitable for quantitative analysis and prediction of precipitation events in various regions.

(4) The algorithm can be extended to the fuzzy Markov chain model [8] which is easy to cluster analysis.

5.2 Weakness

(1) The analytic hierarchy process has disadvantages. The importance indexes obtained by subjective judgment and experience are not objective and accurate, and there are not enough high-level experts, so it is difficult to determine the weight of indexes by expert evaluation method.

(2) The information obtained on the website is somewhat deceptive, some of the information is inaccurate, and the data is very rare.

Proposals for improvements to the sewer system

Sewerage system Management:

After COVID-19 swept through China and was effectively treated, extreme rainfall and extreme flooding events have again played out in tragic scenes, extreme rainfall events occur that humans cannot control, only by taking more precautions can we reduce the loss as much as possible.

Two major rainfall events in a row were enough to get everyone's attention, after investigating the flood events in Zhengzhou, we found that extreme rainfall events are closely related to temperature. Rising global temperatures, melting glaciers and rising sea levels have all been blamed for extreme rainfall in recent years.

In order to prevent the occurrence of extreme rainfall events as much as possible, here are three suggestions for improving the sewer system:

Suggestion 1:

With regard to historical rainfall, more ground drainage outlets should be added in areas with high historical rainfall, especially in low-lying areas. Prevent extreme rainfall phenomenon when drainage is not timely, the surface of the flood phenomenon.

Suggestion 2:

On the basis of the combined sewer system, the original pipe is used as waste water and sewage discharge pipe. The widened section serves as a storm drain. Thus, the combined sewer system is improved and realized.

Suggestion 3:

Rainwater drains in two places. First of all, rain water purification stations can be built in fire stations, car washes and other areas that need a lot of water. After the purification of rain water, it can be used as a source of water for fire protection. When the water purification station storage tank is full, in the selection of discharge to rivers and lakes.

Suggestion 4:

The drainage channel and outlet must be checked regularly to see whether there is foreign body blockage. Blocking the drain with foreign matter and discharging solids into the drain are strictly prohibited.

Suggestion 5:

Conduct comprehensive publicity, with emphasis on how to open manhole covers in emergency situations.

The above suggestions are not comprehensive, but hope to provide help to the sewer system management department in the improvement of the sewer system.

Wishing you every success,

Competition team

2021/11/14

References

[1] Minghui Wu, Guangjie Zhang ,Canghong Jin. Time series prediction model based on multi-modal information fusion [J/OL]. Computing applications:1-8[2021-11-14].h ttp://kns.ki.net/kcms/de tail/51.13 07.TP.20 211 0 14.1357.0 1 4.html.

[2] Yaojian Lu, Hexiang Liu, Meng Wang. Risk entropy-extreme learning machine prediction model for Typhoon disaster in South China [J]. disaster,2 019,3 4(04):216-221.

[3] Yue Wang ,Jinmin He, WeiPin Liu. Based on grey correlation analysis and fuzzy comprehensive evaluation .GNSSS.poofing effectiveness evaluation [J]. Electronic journals,2020, 48(12):2 352 -2359.

[4] Ya Dong, Dujin Lin, Yujie Hu. Precipitation prediction based on k-means Fuzzy Markov model [J]. Modern electronic technology,2021,44(19):85-89.

[5] Yong Chen. Sustainable use of Tokyo's sewerage system [J]. Urban Planning abroad,2003(03):50-54.

[6] Guo Lei, Hua Zhang, Qubin Yuan. Study on quantitative analysis method of regional environment-economy coordinated development [J]. Sichuan environment,2003(05):67-72.

[7] Yubo Wang, Xiujuan Liang, Yu Qiao ,Liang Wang,Haiyan Xu ,Wei Chen. Precipitation prediction based on Superimposed Markov chain and BP neural network model [J]. China's rural water and hydropower,2014(09):80-82+86.

[8] Fan Song ,Xiaohua Yang ,Feifei Wu,Tong Liu . Application of fuzzy Markov chain in rainfall prediction based on cluster analysis [J]. Water saving irrigation,2018(10):3 3-36+41

Appendix

Appendix I

[1] https://tieba.baidu.com/p/7205861089
[2] https://bajiu.cn/ditu/?id=20523
[3] http://www.wheata.cn/
[4] http://data.tpdc .ac.cn/zh-hans/data/24b ee059-df68-4c01-ac 93-6b5 70e36aa d 3/?q=&ivksa=10 243 20u
[5] https://wenku.baidu.com/view/5a7ea160dd88d0d232d46a8f.html
[6] http://nj.tjj.beijing.gov.cn/nj/main/2021-tjnj/zk/indexce.htm
[7] http://tjj.gz.gov.cn/zyxz/tjnjdzzy/
[8] http://tjj.xa.gov.cn/tjsj/tjxx/1.html
[9] http://tjj.hubei.gov.cn/tjsj/sjkscx/tjnj/gsztj/whs/2019 11/P02019110465335679548 0.pdf
[10] tianqi.com

Appendix I

The code for problem number one

```
clc
clear
load station1;
load station2;
load station3;
A= station1;
B = station2;
C = station3;
b = B(:,1);
a = B(:,2);
plot(b,a,'-ob','LineWidth',1);
hold on;
xlabel('Year(years)');    ylabel('Precipitation (mm) ');
clear all
close all
clc
a = [0.177/0.591 0.212/0.591 0.202/0.591];
    pie(a);
    legend('MIN','MAX','TEMP');
    im_hatch = applyhatch_plusC(gcf,'|-+.\/','rgbcmy',[],100,1);
    imwrite(im_hatch,'im_hatch.tiff','tiff')
```

The code for problem number two

```
load beijing;
load guangzhou;
```

```
load wuhan;
load xian;
load Beijingce;
load Beijingcet;
load BeijingM;
D = beijing;
E = guangzhou;
F = wuhan;
G = xian;
F = BeijingM;
b = D(:,1);
a = D(:,2);
plot(b,a,'-ob','LineWidth',1);
hold on;
d = F(:,1);
c = F(:,2);
plot(d,c,'-*g','LineWidth',1);
hold on;
xlabel('Year(years)');    ylabel('Precipitation (mm) ');
```

The code for problem number three

```
clear;clc
year =[1978:1:2015]';
x =
[664.8,718.4,380.7,393.2,544.4,489.9,488.8,721.0,665.3,683.9,673.3,442.2,697.3,747.9,541.5,506
.7,813.2,572.5,700.9,430.9,731.7,266.9,371.1,338.9,370.4,444.9,483.5,410.7,318.0,483.9,626.3,48
0.6,522.5,720.6,733.2,578.9,461.5,458.6]';
figure(1);
plot(year,x,'o-'); grid on;
xlabel('year');    ylabel('Precipitation (mm) ');
ERROR = 0;
if sum(x<0) > 0
    disp(' There can't be negative numbers in the time series of grey prediction ')
    ERROR = 1;
end
m = length(x);
disp(strcat('lengthen',num2str(m)))
if m<=3
    disp('cannot go ')
    ERROR = 1;
end
if m>10
    disp('use ARIMA ')
end
```

```
if size(x,1) == 1
        x = x';
end
if size(year,1) == 1
        year = year';
end
if ERROR == 0
        x1 = cumsum(x);
        figure(2)
        plot(year(2:end),rho,'o-',[year(2),year(end)],[0.5,0.5],'-'); grid on;
        text(year(end-1)+0.2,0.55,' Critical path ')
        set(gca,'xtick',year(2:1:end))
        xlabel('year');    ylabel(' The smoothness of the raw data ');
        Judge = input('input 1 or 2 ');
        if Judge == 0
                ERROR = 1;
        end
        disp('------------------------------------------------------------')
end
if ERROR == 0
        if    m > 4
                disp(' Since the number of periods in the original data is greater than 4, we
can divide the data group into training group and experimental group ')
                if m > 7
                        test_num = 3;
                else
                        test_num = 2;
                end
                train_x = x(1:end-test_num);
                disp('test data: ')
                disp(mat2str(train_x'))
                test_x =    x(end-test_num+1:end);
                r1 = gm11(train_x, test_num);
                r2 = new_gm11(train_x, test_num);
                r3 = metabolism_gm11(train_x, test_num)
                test_year = year(end-test_num+1:end);
                figure(3)
plot(test_year,test_x,'o-',test_year,r1,'*-',test_year,r2,'+-',test_year,r3,'x-'); grid on;
                set(gca,'xtick',year(end-test_num+1): 1 :year(end))
                legend(' Group of real data ', 'traditional GM (1, 1) prediction results',' new
information GM (1, 1) prediction results', 'metabolism GM (1, 1) prediction results ')
                xlabel('year');    ylabel('Precipitation (mm) ');
                S1 = sum((test_x-r1).^2);
                S2 = sum((test_x-r2).^2);
```

```
        S3 = sum((test_x-r3).^2);
        if S1<S2
            if S1<S3
                choose = 1;
            else
                choose = 3;
            end
        elseif S2<S3
            choose = 2;
        else
            choose = 3;
        end
        Model = {' Traditional GM(1,1) model ',' new information GM(1,1) model ',' metabolic GM(1,1) model ' '};
        predict_num = input(' Please enter the number of periods you want to predict further down：  ');
        [r, x_hat, relative_residuals, eta] = gm11(x, predict_num);
        if choose == 2
            r = new_gm11(x, predict_num);
        end
        if choose == 3
            r = metabolism_gm11(x, predict_num);
        end
    disp(' The fitting results of the original data：  ')
        for i = 1:n
            disp(strcat(num2str(year(i)), '  ：   ',num2str(x_hat(i))))
        end
        disp(strcat(' Predict the result of the ',num2str(predict_num),' phase：  '))
        for i = 1:predict_num
            disp(strcat(num2str(year(end)+i), '  ：   ',num2str(r(i))))
        end
    else
        predict_num = input(' Please enter the number of periods you want to predict further down：  ');
        [r1, x_hat, relative_residuals, eta] = gm11(x, predict_num);
        r2 = new_gm11(x, predict_num);
        r3 = metabolism_gm11(x, predict_num);
        r = (r1+r2+r3)/3;
        for i = 1:n
            disp(strcat(num2str(year(i)), '  ：   ',num2str(x_hat(i))))
        end
        for i = 1:predict_num
            disp(strcat(num2str(year(end)+i), '  ：   ',num2str(r1(i))))
        end
```

```
        disp(strcat(' The new information GM(1,1) is the result of the subsequent prediction ',num2str(predict_num),' phase：'))
        for i = 1:predict_num
            disp(strcat(num2str(year(end)+i), ' ：  ',num2str(r2(i))))
        end
        disp(strcat(' Metabolism GM(1,1) after the prediction ',num2str(predict_num),' phase of the obtained results：'))
        for i = 1:predict_num
            disp(strcat(num2str(year(end)+i), ' ：  ',num2str(r3(i))))
        end
        disp(strcat(' Three methods are used to average the results obtained by the future prediction ',num2str(predict_num),' period：'))
        for i = 1:predict_num
            disp(strcat(num2str(year(end)+i), ' ：  ',num2str(r(i))))
        end
    end
    figure(4)
    subplot(2,1,1)
    plot(year(2:end), relative_residuals,'*-'); grid on;
    legend(' The relative residual '); xlabel('year');
    set(gca,'xtick',year(2:1:end))
    subplot(2,1,2)
    plot(year(2:end), eta,'o-'); grid on;
    legend(' Level than the deviation '); xlabel('year');
    set(gca,'xtick',year(2:1:end))
    disp(' ')
    average_relative_residuals = mean(relative_residuals);
    disp(strcat(' The average relative residual is ',num2str(average_relative_residuals)))
    if average_relative_residuals<0.1
        disp('good')
    elseif average_relative_residuals<0.2
        disp('just so so ')
    else
        disp('bad')
    end
    average_eta = mean(eta);
    disp(strcat(' The average stage ratio deviation is ',num2str(average_eta)))
    if average_eta<0.1
    end
    figure(5)
    plot(year,x,'-o',  year,x_hat,'-*m',  year(end)+1:year(end)+predict_num,r,'-*b' );
    grid on;
    hold on;
    plot([year(end),year(end)+1],[x(end),r(1)],'-*b')
```

```
    legend(' Raw data ',' fit data ',' forecast data ' ')
    set(gca,'xtick',[year(1):1:year(end)+predict_num])
    xlabel('Year(years)');    ylabel('Precipitation (mm) ');
end
```

ELM

```
clear all
clc
load('Beijing.mat');
load('Beijingc.mat');
input=Beijing;
output=Beijingc';
P =input(1:39,:)';
T =output(1:39);
P_t=input(39:end,:)';
T_t=output(39:end);
[PT,inputps] = mapminmax(P,-1,1);
Pt = mapminmax('apply',P_t,inputps);
[Tn,outputps] = mapminmax(T_t,-1,1);
Tn_t = mapminmax('apply',T_t,outputps);
Tic
[IW,B,LW,TF,TYPE] = elmtrain(PT,Tn,20,'sig',0);

Tn_s = elmpredict(Pt,IW,B,LW,TF,TYPE);
Tm_sim=elmpredict(PT,IW,B,LW,TF,TYPE);
T_s = mapminmax('reverse',Tn_s,outputps);
T_s2 = mapminmax('reverse',Tm_sim,outputps);

toc
result = [T_test' T_s'];
E = mse(T_s - T_t)
N = length(T_t);
sum((T_t).^2)-(sum(T_t))^2))
R1=sqrt(sum((T_s2-T).^2)/size(T,2));
disp([' Root mean square error of extreme learning machine to training set：',num2str(R1,'%1.4f')])
R2=sqrt(sum((T_s-T_t).^2)/size(T_t,2));
disp([' Root mean square error of extreme learning machine to training set：',num2str(R2,'%1.4f')])
figure(1)
subplot(211)
plot(1:length(T),T,'r*')
hold on
plot(1:length(T_s2),T_s2,'b:o')
xlabel(' Training set sample number ')
```

```
ylabel(' Training set output ')
title('ELMTraining set output')
legend(' Expected output ',' training set predicts output ')
subplot(212)
plot(1:length(T_t),T_t,'r*')
hold on
plot(1:length(T_s),T_s,'b:o')
xlabel(' Test set sample number ')
ylabel(' Test set output ')
title('ELMTest set output')
legend(' Expected output ',' predicted output ')
```