Problem Chosen	2021 ShuWei Cup Summary Sheet	Team Control Number
B		202111087164

Summary

Under the background of global warming, various extreme weather and disaster events are increasing, especially extreme rainfall disasters, which bring serious impacts to both human life and social economy. In this paper, we quantitatively analyze the extreme precipitation weather in major cities in China, apply correlation analysis to derive the correlation degree between rainfall and other meteorological indicators, analyze the historical precipitation of different typical cities, and then construct an urban prediction model for potential extreme precipitation events.

For problem 1, the annexed data are firstly pre-processed and corrected, and then the Pearson correlation coefficients between rainfall and other meteorological indicators are calculated for two correspondences. Meanwhile, considering the small data samples and the existence of data errors, gray correlation analysis can be used to filter out the years 2003, 2004, 2016, and 2018 with high precipitation by M-K mutation test and cluster analysis based on the characteristics of total annual precipitation and the number of rain days in a year. Based on the main influencing factors, it was found that the sudden variability of precipitation in 2021 was obvious in July.

For question 2, the precipitation data of Beijing, Taiyuan, Chengdu, Guiyang, Shenzhen and other cities since recent years were obtained by consulting meteorological websites and data mining for data analysis. The trend of monthly precipitation in the above cities was generally stable, but it started to increase significantly in May 2020, with extreme weather such as heavy rainfall, reaching the highest peak of precipitation in recent years.

For question 3, in order to analyze the cities where extreme rainfall may occur in the future, monthly precipitation prediction analysis using seasonal time series models and neural network models, as well as annual precipitation analysis using gray prediction models, were performed based on the city data collected in the previous question. Combining the above prediction results, it can be obtained that: extreme precipitation will occur in Shenzhen in the 10th month of the future; extreme precipitation will occur in Chengdu in the 10th month, which may exceed 3000mm per month.

For question 4, factor analysis was used to correlate the rainstorm characteristics of Zhengzhou in July 2021 and Taiyuan in October 2021, and it was found that the precipitation in Zhengzhou in July and Taiyuan in October is the same in that the precipitation is correlated with the atmospheric pressure Po at the weather station level. Comparing the affected data of the two places, it was found that the heavy rainfall in Shanxi and Zhengzhou had the characteristics of long duration, high accumulated rainfall and outstanding extremity. Agriculture in Shanxi is greatly affected by extreme rainfall, while the transportation industry in Zhengzhou suffers serious losses due to extreme rainfall.

For Question 5, through the analysis of the above question for the characteristics of heavy rainfall in Zhengzhou and Shanxi, the long-term planning of future cities under extreme precipitation conditions should focus on the emergency response capability and effectiveness of infrastructure construction for various possible sudden extreme precipitation, and establish a sound long-term effect response to extreme precipitation warning mechanism.

Keywords: Extreme Rainfall Weather; Correlation Analysis; ARIMA Model; M-K Salient Test; Factor Analysis

一、 Restatement of issues

Many parts of Henan were severely affected by the historically rare heavy rainfall in July 2021. Among them, Zhengzhou suffered an extraordinarily heavy rainfall, which is the scale of a once-in-a-millennium event. All of the single-day precipitation on July 20 broke the 60-year record since Zhengzhou established a weather station in 1951, reaching the extreme value of hourly rainfall on land in China. The extreme rainfall brought huge losses and disasters to Zhengzhou, which has a population of tens of millions - subways were flooded, stations leaked, streets turned into rivers, and flooding and secondary disasters from the rainstorm have posed serious threats to the lives, safety and property of local people.

On October 6, 2021, Shanxi ushered in a very heavy rainfall, Taiyuan City due to heavy rainfall part of the river directly broke the embankment, flooding directly submerged the village, during the flood, Shanxi around a total of 54,947 people transferred, Shanxi Province has launched the provincial natural disaster relief level III emergency response. In recent years, China's Shaanxi, Hubei and other areas have suffered extreme rainfall weather, to the local people's lives, property security caused a great threat.

In the context of global warming, various extreme weather and disaster events are increasing, especially extreme rainfall disasters, which have serious impacts on human life and socio-economics. According to relevant data and statistics by the end of this century, precipitation is expected to increase by about 10 per cent, and the probability of extreme precipitation will increase significantly. Moreover, the natural environment in China varies greatly, and the comprehensive influence of the complex geographical environment, precipitation in different cities has different characteristics. Therefore, the establishment of prediction models for extreme precipitation in cities is of great significance for the prevention of floods in advance.

Question 1: Conduct a correlation analysis of previous annual precipitation characteristics in Zhengzhou, screen for years with extreme weather such as heavy rainfall, and also conduct a quantitative analysis for the very heavy rainfall in 2021.

Question 2: Collect historical precipitation data from other cities and analyze the trends in precipitation in different cities.

Question 3: Using the precipitation data collected for different cities, apply different methods to predict the cities that will experience extreme rainfall in the future and analyze the results of the predictions.

Question 4: Compare the characteristics of the July 2021 storm in Zhengzhou and the October 2021 storm in Shanxi and analyze the difference in damage caused by the storm to Zhengzhou and Taiyuan.

Question 5: Long-term construction planning ideas for future city construction under extreme rainfall conditions in selected typical cities in China.

二、 **Analysis of the problem**

2. 1 **Analysis of question I**

Before the analysis related to the annual characteristics of precipitation in Zhengzhou area, the annexed data were first analyzed and processed; the data from observation station 1 were from December 1957 to November 2021, the data from observation station 2 were from July 1983 to November 2021, and the data from observation station 3 were from October 1961 to November 2021. From the perspective of comprehensiveness of monitoring, the data from the three observatories around Zhengzhou were combined, and the more complete precipitation data were analyzed from 1984. At the same time, the precipitation data of each year are more than 360 items, and the missing data are less, and considering that the weather data have the variability of different seasons in different years, so the missing data can be added without correction.

After processing the data, the annual changes of precipitation characteristics in Zhengzhou area were correlated using the data from 1984-2021. Firstly, the scatter plots between different meteorological indicators and precipitation were plotted, and the Pearson correlation coefficients corresponding to the two were calculated, and significance tests were conducted to determine whether they were correlated and the strength of the correlation. At the same time, considering the small data samples and the existence of data errors, we can use gray correlation analysis, and then determine the correlation size between different meteorological indicators such as dew point temperature, air temperature, visibility and precipitation. Combining the above two correlation analyses, the correlation between different meteorological indicators and precipitation is derived.

In order to effectively count the characteristics of precipitation changes in different years, we selected the total annual precipitation, the average annual precipitation, the number and proportion of rain days in the year, and the average annual dew point for analysis, based on which we can classify the precipitation in Zhengzhou in different years by the cluster analysis method, so as to group the years with excessive precipitation as well as abnormal precipitation into one category and arrive at the higher precipitation years required by the topic. Also, in order to derive the trends of precipitation and precipitation frequency in Zhengzhou, linear regression was performed to observe future precipitation trends. The precipitation in Zhengzhou in 2021 was quantified based on the main influencing factors.

2. 2 **Analysis of issue three**

Based on the data collected in Problem 2, we forecast and analyze future precipitation for the cities of Zhengzhou, Taiyuan, and Beijing. The data collected and compiled are time series data, and based on the analysis of time series variables, a forecasting model is developed to extend the time trend outward so as to predict the future market development trends and determine the variable forecast values.

The data collected from different cities are organized into monthly precipitation and annual precipitation, and in order to analyze cities with potential extreme rainfall in the future, based on the city data collected in the previous question, we use seasonal time series model and neural network model for monthly precipitation prediction analysis, as well as grey prediction model for annual precipitation analysis. And the effects of different precipitation prediction models are compared to finally build a city prediction model for potential extreme precipitation events, and to give an opinion for the long-term construction planning of future cities.

2. 3 Analysis of issue four

Whether the rainstorm characteristics of Zhengzhou in July 2021 are the same as those of Shanxi in October 2021, that is, to study whether precipitation in Zhengzhou is correlated with precipitation in Shanxi. We selected Taiyuan city as the research object in Shanxi province and Zhengzhou city as the research object in Henan province, and counted out the precipitation change-related indicators, and the data were counted in months, and factor analysis was used to condense the many variables already in the original data into a few factors, reflecting the characteristics of rainfall and the remaining indicators in the two cities.

In order to further explore the difference between the losses caused by heavy rainfall in Zhengzhou, Henan Province and Taiyuan, Shanxi Province, we selected the economic indicators of the two provinces, combined with the physical geographic characteristics of Zhengzhou, Taiyuan, based on the comparison of the conditions under which rainfall occurs to derive an analysis of the impact of extreme weather rainfall conditions on the economic level, combined with the relevant industries affected by extreme rainfall to explore the difference between the losses in Zhengzhou and Shanxi.

2. 4 Analysis of issue five

Based on the results of the above four questions, a typical city in China is selected for its long-term construction planning advice for future urban construction under extreme rainfall conditions, in order to improve the early warning response capability of cities to deal with potential extreme rainfall.

2. 5 Analysis of Question Five

Based on the results of the above four questions, the typical cities in China are selected for their long-term construction planning advice for future urban construction under extreme rainfall conditions, in order to improve the early warning response capability of cities to deal with potential Based on the above four questions, the typical cities in China are selected for their long-term construction planning advice for future urban construction under extreme rainfall conditions, in order to improve the early warning response capability of cities to deal with potential

三、 Notation And Model Assumptions

3. 1 Model assumptions

1. Considering only the core factors in rainfall, without considering the effects of secondary factors.
2. Exclusion of small probability events that occur during rainfall.
3. Excluding the influence of various types of topography and other factors on precipitation and considering only the meteorological indicators given in the question.
4. Exclusion of errors in the detection of meteorological indicators at different weather stations.

5. Ignoring the effect of outliers on the validity of the data.
6. A significant increase in the probability of extreme precipitation is assumed.
7.

3. 2 Description of symbols

symbolic	instructions	unit
DEWP	Average dew point	degrees centigrade
PRCP	precipitation (meteorology)	
MAX	highest temperature	degrees Fahrenheit
MIN	minimum temperature	degrees Fahrenheit
SLP	Average sea level pressure	millipax
TEMP	average temperature	degrees Fahrenheit
WDSP	Average wind speed	m/s
VISIB	Average visibility	kilometres

四、 Model building and solving

4. 1 Modeling and solving of problem one

4. 1. 1 Modeling

Question 1 is for us to analyze the data in the annexes and analyze the Pearson correlation coefficients between rainfall and other meteorological indicators in two counterparts and perform significance tests to determine whether they are correlated and the strength of the correlation. At the same time, considering the small data sample and the existence of data errors, grey correlation analysis can be used, which in turn leads to the main effects of dew point, air temperature and wind speed on the variation of precipitation. The years with higher precipitation are filtered out by M-K mutation test, cluster analysis based on the total annual precipitation, the number of days of rain in the year and other characteristics.

A correlation is a non-deterministic relationship, and the correlation coefficient is a quantity that studies the degree of linear correlation between variables. Due to the different objects of study, the correlation coefficient is defined in several ways as follows.

Pearson's correlation coefficient: also known as the correlation coefficient or linear correlation coefficient, generally denoted by the letter r, is used to measure the linear relationship between two variables with the following formula

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

where the value of the correlation coefficient r is between -1 and +1, i.e. $-1 \le r \le +1$. It can generally be classified in three levels: $|r| < 0.4$ for low linear correlation; $0.4 \le |r| < 0.7$ for significant correlation; and $0.7 \le |r| < 1$ for high linear correlation.

To further investigate the correlation between precipitation and other meteorological indicators, we used grey correlation analysis, first we manually processed the data in the annexes, and the meteorological indicators in the annexes were DEWP, FRSHTT, GUST, MAX, MIN, MXSPDSLP, SNDP, ISTP, TEMP, VISIB, WDSP twelve influential factors, which are mean dew point, occurrence of unusual hazards, gust wind speed, maximum temperature, minimum temperature, maximum sustained wind speed, precipitation, mean sea level pressure, snow depth, mean station pressure, mean temperature, mean visibility, and mean wind speed.

The formula for calculating the grey correlation coefficient is as follows.

$$\xi_i(k)=\frac{\min\limits_s \min\limits_t |x_0(t)-x_s(t)|+\rho \max\limits_s \max\limits_t |x_0(t)-x_s(t)|}{|x_0(k)-x_i(k)|+\rho \max\limits_s \max\limits_t |x_0(t)-x_s(t)|}$$

The above equation is defined as the grey coefficient, i.e. the correlation coefficient of each influencing factor is.

$$y(x_0(t),x_i(t))=\frac{a+\rho b}{|x_0(t)-x_i(t)|+\rho b}$$

where a is the bipolar minimum difference, b is the bipolar maximum difference, and ρ is the resolution factor.

$$a=\min\limits_i \min\limits_t |x_0(t)-x_i(t)|$$

$$a=\max\limits_i \max\limits_t |x_0(t)-x_i(t)|$$

Define the gray correlation $y(x_0,\ x_i)$, i.e., find the mean value of each column of the matrix for which the correlation coefficients are obtained.

$$y(x_{0,}x_i)=\frac{1}{n}\sum_{k=1}^{n} y(x_0(t),x_i(t))$$

The influence of the correlation of different meteorological indicators on precipitation can be derived by comparing the grey correlation.

4. 1. 2 **Solving the model**

The Pearson correlation coefficients were derived from the SPSS analysis of precipitation and other meteorological indicators in the annex, which yielded the following results.

Table 1Correlation data of precipitation PRC and dew point DEWP

	PRCP	DEWP

PRCP	Pearson Correlation	1	.069 **
	Sig. (bobtail)		.000
	Number of cases	19447	19447
DEWP	Pearson Correlation	.069 **	1
	Sig. (bobtail)	.000	
	Number of cases	19447	19447

**. Significant correlation at the 0.01 level (two-tailed).

By analyzing the data in Table 1, we can conclude that the Pearson correlation coefficient is close to 1 and the two-tailed is at the level of 0.01, we can conclude that the precipitation PRC is significantly correlated with the dew point DEWP.

Table 2 Data table of correlation between precipitation PRC and temperature TEMP

		PRCP	TEMP
PRCP	Pearson Correlation	1	.029**
	Sig. (bobtail)		.000
	Number of cases	19447	19447
TEMP	Pearson Correlation	.029**	1
	Sig. (bobtail)	.000	
	Number of cases	19447	19447

**. Significant correlation at the 0.01 level (two-tailed).

Analysis of the data in Table 2 shows that the Pearson correlation coefficient is close to 1 and the two-tailed is at the level of 0.01, we can conclude that the precipitation PRC is significantly correlated with the temperature TEMP.

Table 3 Correlation data table between precipitation PRC and gust wind speed WDSP

		PRCP	WDSP
PRCP	Pearson Correlation	1	.022**
	Sig. (bobtail)		.002
	Number of cases	19447	19447
WDSP	Pearson Correlation	.022**	1
	Sig. (bobtail)	.002	

Number of cases	19447	19447

**. Significant correlation at the 0.01 level (two-tailed).

By analyzing the data in Table 3, we can conclude that the Pearson correlation coefficient is close to 1 and the two-tailed is at the level of 0.01, we can conclude that the precipitation PRC is significantly correlated with the wind speed WDSP.

The magnitude of Pearson's correlation coefficient led to the conclusion that dew point, air temperature and wind speed play a major role in precipitation variation and passed the significance test.

According to the grey correlation analysis can be obtained according to the grey correlation of each indicator in the sub-series are: 09892, 0.9894, 0.9888, 0.9889, 0.9890, 0.9892, 0.9888, 0.9888, 0.9888, 0.9889, 0.9887, 0.9894, comparing the magnitude of the scores of different indicators can be the following table.

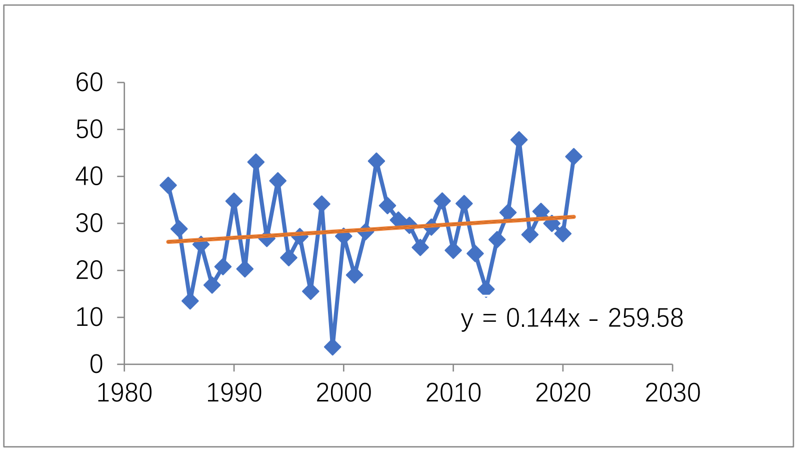
Table 4 Table of grey correlation between precipitation and other meteorological indicators

targets	connotation	gray correlation
DEWP	Average dew point	0.9892
FRSHTT	Occurrence of unusual disasters	0.9894
GUST	gust wind speed	0.9888
MAX	highest temperature	0.9889
MIN	minimum temperature	0.9890
MXSPD	Maximum sustained wind speed	0.9892
SLP	precipitation (meteorology)	0.9888
SNDP	snow depth	0.9888
STP	Average station pressure	0.9888
TEMP	average temperature	0.9889
VISIB	Average visibility	0.9887
WDSP	Average wind speed	0.9894

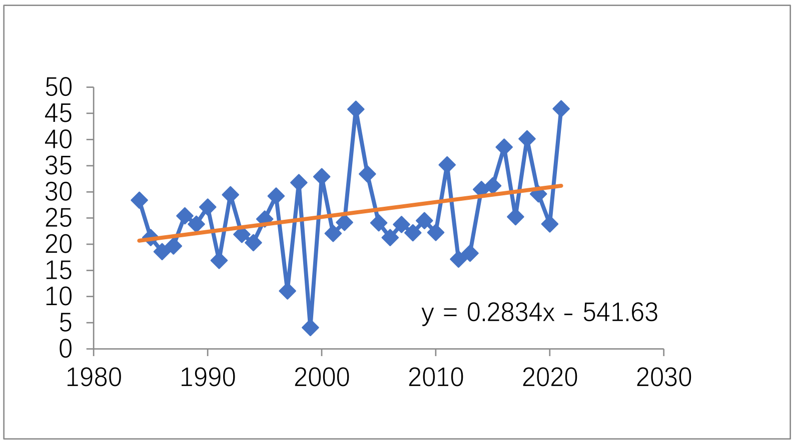
The ranking of the above data table of grey correlation between precipitation and other meteorological indicators shows that FRSHTT occurrence of anomalous hazard, WDSP average wind speed, and DEWP average dew point have the greatest influence on precipitation with grey correlation of 0.9894 , 0.9894 , and 0.9892 respectively.

To observe the trend of precipitation at different meteorological monitoring points in

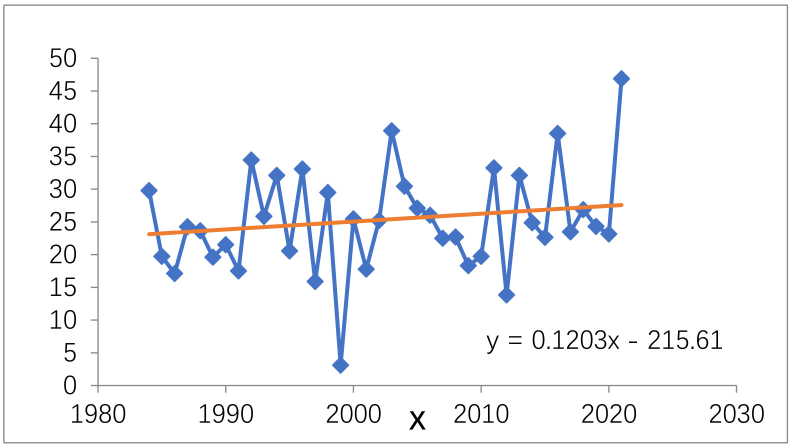
Zhengzhou, we performed a linear fit to the precipitation data collected at meteorological stations 1, 2 and 3 from 1984 to 2021, respectively, as shown in the following figure.



Linear Regression Plot of Zhengzhou Weather Station 1, 1984-2021

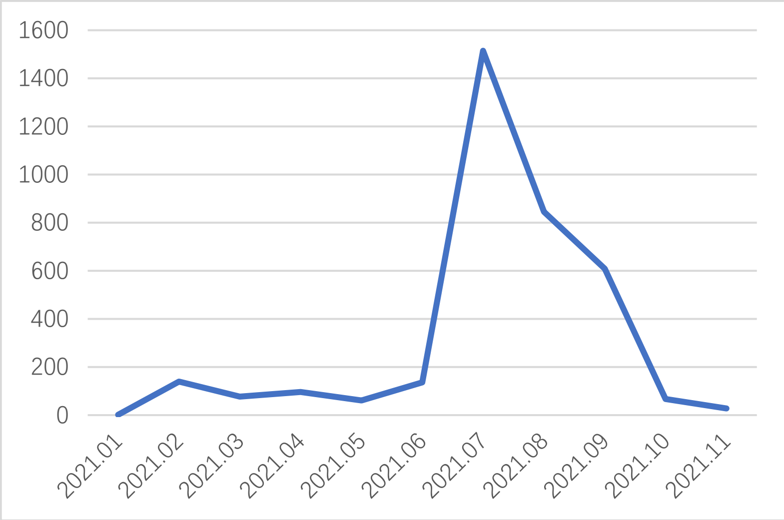


Zhengzhou Weather Station 2 Linear Regression Plot, 1984-2021



Zhengzhou Weather Station 3 Linear Regression Plot, 1984-2021

Observing the trend of annual precipitation changes in three meteorological stations in Zhengzhou, we find that the precipitation from 1984 to 2021 all show an upward trend. The linear regression equation of station 1 is: $y = 0.144x - 259.58$; the linear regression equation of station 2 is: $y = 0.2834x - 541.63$; the linear regression equation of station 3 is: $y = 0.1203x - 215.61$. The regression coefficients of all three regression equations are greater than zero, so the annual precipitation of Zhengzhou shows an upward trend.



By plotting the month-by-month precipitation folding statistics for 2021 in Zhengzhou, we find that the peak precipitation in 2021 was reached in July 2021, and the precipitation in August 2021 declined, but was still higher than the precipitation in other months. the single day precipitation on July 20 had both broken the 60-year historical record since the establishment of the weather station in Zhengzhou in 1951. At the same time, the heavy rainfall in Zhengzhou lasted for many days. In the previous rainfall history of Zhengzhou, the heaviest rainfall was at 3pm on July 3, 1986, and the cumulative six-hour precipitation for that time was as high as 201mm.

4. 2 Problem 2 modeling and solving

4. 2. 1 Solving the model

Trend analysis of rainfall change is an important aspect of climate change research, but the trend of change is not the same between different seasons, we mainly analyze from the highest peak of rainfall per year, in order to be able to fully analyze the precipitation situation of different cities, should be selected from different cities with different precipitation and different topography in China.

By using Excel to draw a line statistical graph of the data we collected, we found that from recent years, the total monthly rainfall in Chengdu and Shenzhen has similar characteristics of change, from January 2015 to April 2020, basically no major fluctuations, but from May 2020, the rainfall increases sharply, there are strong rainfall and other extreme weather, and the monthly rainfall peak becomes an increasing trend.

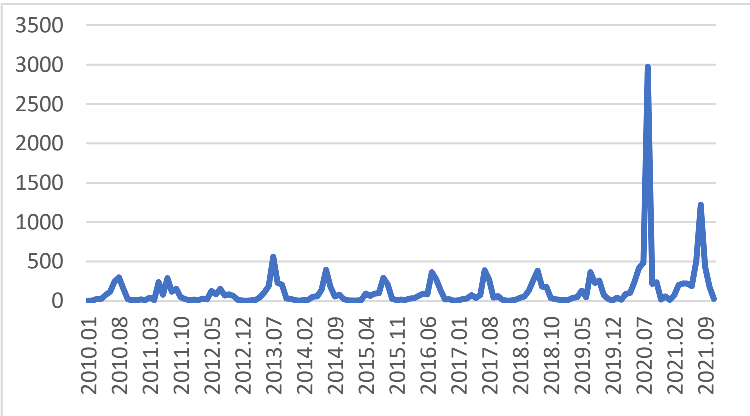


Figure 1Chengdu precipitation trend from 2010-2021

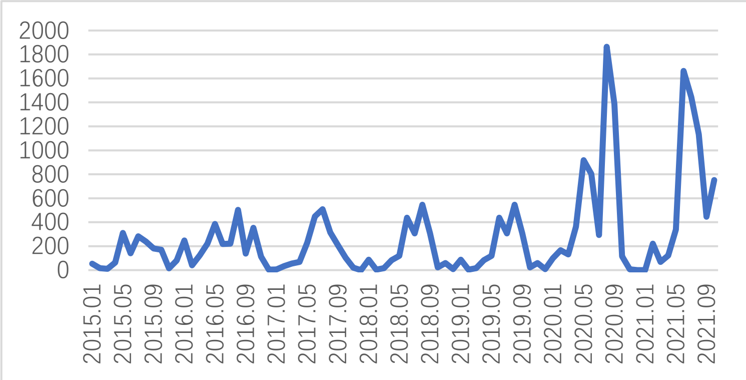


Figure 2Trend of precipitation in Shenzhen from 2015 to 2021

Monthly rainfall in Taiyuan is basically stable without major fluctuations from January 2010 to April 2020, but heavy precipitation occurs in July 2020, reaching the highest peak in recent years, and subsequent rainfall decreases, but still shows an upward trend compared to previous years.

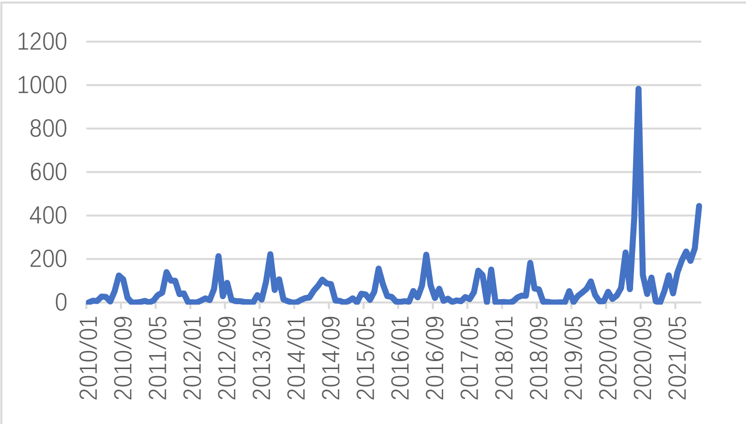


Figure 3Trend of precipitation in Taiyuan from 2010-2021

Compared to these previous cities, Beijing receives relatively little rainfall, with relatively stable rainfall from January 2010 to January 2020, but again, starting in May 2020, rainfall increases sharply to a near decade-long peak in September 2021.

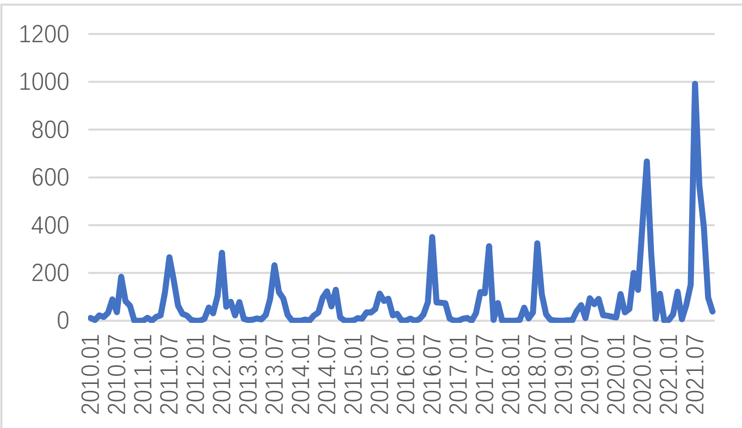


Fig. 4Precipitation trends in Beijing, 2010-2021

From the above analysis most cities started to receive heavy precipitation from 2020, and one of the main reasons for this phenomenon is: good water vapor conditions. There are four main types of rainfall in China: frontal rain, convective rain, topographic rain, and typhoon rain. However, all types of rainfall need to have three conditions: sufficient water vapor, suitable power lifting conditions and the presence of a large number of condensation nuclei in the air. For example, the recent plum rains that have been coiled around Jiangnan and Jianghuai are sustained rainfall brought about by sufficient water vapor. There are two typical "water vapor channels" that cause heavy and continuous precipitation in China during the flood season, with the Pacific and Indian Oceans being strong "suppliers".

China is a monsoon climate country, the winter is mainly controlled by the north wind, like the Siberian cold air in the north wind "urge", frequently "cooling" China; and in the summer, the wind direction is greatly reversed, from blowing north wind to blowing south wind, transporting water vapor from the ocean to the land. In summer, the wind direction is reversed from north to south, transporting water vapor from the ocean to the land. So, with the "east wind" of summer winds, both the Pacific Ocean and the Indian Ocean can continuously deliver to China's land the important "raw material" needed for precipitation: water vapor.

4. 3 **Problem 3 modeling and solving**

4. 3. 1 **Modeling**

Model building is the abstraction of the original problem into an expression in mathematical language, which must be based on the previous problem analysis and model assumptions. Because of the time constraints of the competition, most of the time we use models that have already been built by others. This part must be closely related to the question asked in the question and the model, and never apply the model arbitrarily. We can also improve or optimize some aspect of the existing model, or build a different model to solve the same problem, which is the innovation and highlight of the paper.

1） $ARMA(p,q)$The parameter moments of the model are estimated in three steps.

i) φ_1, $\varphi_2, \cdots$, φ_p Estimates of

$$\begin{pmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \\ \vdots \\ \hat{\varphi}_p \end{pmatrix} = \begin{pmatrix} \hat{\gamma}_q & \hat{\gamma}_{q-1} & \cdots & \hat{\gamma}_{q-p+1} \\ \hat{\gamma}_{q+1} & \hat{\gamma}_q & \cdots & \hat{\gamma}_{q-p+2} \\ \vdots & \vdots & & \vdots \\ \hat{\gamma}_{q+p-1} & \hat{\gamma}_{q+p-2} & \cdots & \hat{\gamma}_q \end{pmatrix}^{-1} \begin{pmatrix} \hat{\gamma}_{q+1} \\ \hat{\gamma}_{q+2} \\ \vdots \\ \hat{\gamma}_{q+p} \end{pmatrix}$$

(ii) Let $Y_t = X_t - \hat{\varphi}_1 X_{t-1} - \cdots - \hat{\varphi}_p X_{t-p}$, then Y_t the moment estimate of the self-covariance function of is :

$$\hat{\gamma}_k^{(Y)} = \sum_{i=0}^{p}\sum_{j=0}^{p} \hat{\varphi}_i \hat{\varphi}_j \hat{\gamma}_{k+j-i}, \hat{\varphi}_0 = -1$$

iii) Consider the Y_t approximation as $MA(q)$ a sequence, and use the parameter estimation method of 2) for the $MA(q)$ sequence.

(4) Model testing

The type and order of the model were determined by the correlation analysis method and the AIC criterion, and the parameters in the model were determined by the method of moment estimation, thus creating an ARMA model to fit a truly random series. However, how good or bad this fit is should be tested mainly by the results of practical applications, but also by mathematical methods.

The following describes the residual autocorrelation test for model fitting, i.e., the white noise test, which for the ARMA model should be performed step by step from $ARMA(1,1)$, $ARMA(2,1)$, $ARMA(1,2)$. $ARMA(2,2)\cdots$ In turn, the parameter estimates are derived as follows.

In general, for the ARMA model

$$u_t = X_t - \sum_{i=1}^{p} \hat{\varphi}_i X_{t-i} + \sum_{j=1}^{q} \hat{\theta}_j u_{t-j}$$

Taking the initial value $u_0, u_{-1}, \cdots, u_{1-q}$ $X_0, X_{-1}, \cdots, X_{1-p}$ of , one can recursively obtain the residual estimate $\hat{u}_1, \hat{u}_2, \cdots, \hat{u}_N$

The following hypotheses are tested.

$H_0 : \hat{u}_1, \hat{u}_2, \cdots, \hat{u}_N$ is the sample from the white noise

$$\hat{\gamma}_j^{(u)} = \frac{1}{N} \sum_{t=1}^{N-j} \hat{u}_{t+j} \hat{u}_t \qquad j = 0,1,\cdots,k$$

$$\hat{\rho}_j^{(u)} = \frac{\hat{\gamma}_j^{(u)}}{\hat{\gamma}_0^{(u)}} \qquad j = 1,\cdots,k$$

$$Q_k = \sum_{j=1}^{k} \left(\sqrt{N}\, \hat{\rho}_j^{(u)} \right)^2 = N \sum_{j=1}^{k} \left(\hat{\rho}_j^{(u)} \right)^2$$ which k is taken $\frac{N}{10}$ around.

When H_0 holds, Q_k obeys χ^2 a distribution with degrees k of freedom of For a given level of significance α

$Q_k > \chi_k^2(\alpha)$ (b) If the model is rejected H_0, the modeling needs to be reconsidered.

$Q_k < \chi_k^2(\alpha)$, then the fit is good and the model test passes.

Also a grey prediction model can be performed for the GM(1.1) model, as follows.

Let $Z^{(1)}$ be $X^{(1)}$ the sequence of Means of Immediate Neighbors (MEAN) generation.

$$Z^{(1)}(k) = \frac{x^{(1)}(k) + x^{(1)}(k\text{-}1)}{2}$$

Then the differential equation of GM(1.1) can be modeled as

$$x^{(0)}(k) + az^{(1)}(k) = \mathrm{b}$$

where $x^{(0)}(k)$ is called the gray derivative, *a is called* the development coefficient, is $z^{(1)}(k)$ called the whitening background value, and b is called the amount of gray action. Let $\hat{\alpha} = (a,b)^{\mathrm{T}}$, then the least squares estimated parameters of the grey differential equation satisfy the following equation.

$$\hat{\alpha} = (B^T \mathrm{B})^{-1} B^T Y_n$$

among others

$$B=\begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \dots & \dots \\ -z^{(1)}(n) & 1 \end{bmatrix} \quad , \quad Y_n=\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \dots \\ x^{(0)}(n) \end{bmatrix}$$

Call $\frac{dx^{(1)}}{dt}+ax^{(1)}=b$ the vernacular equation for the gray differential $x^{(0)}(k)+az^{(1)}(k)=\mathrm{b}$ equation, also called the shadow equation. Call Y the data vector, B the data matrix, and $\hat{\alpha}$ is the parameter vector. Then.

1.The solution to $\frac{dx^{(1)}}{dt}+ax^{(1)}=b$ the whitening equation is also known as the time response function.

$$\hat{x}^{(1)}(t)=(x^{(1)}(0)-\frac{b}{a})e^{-at}+\frac{b}{a}$$

2.The time response series of the grey differential equation $x^{(0)}(k)+az^{(1)}(k)=\mathrm{b}$ GM(1.1) is

$$\hat{x}^{(1)}(k+1)=\left[x^{(1)}(0)-\frac{b}{a}\right]e^{-at}+\frac{b}{a}, \quad k=1,2,\dots,n$$

3. Take $x^{(1)}(0)=x^{(0)}(1)$, then

$$\hat{x}^{(1)}(k+1)=\left[x^{(0)}(1)-\frac{b}{a}\right]e^{-at}+\frac{b}{a}, \quad k=1,2,\dots,n$$

4. The values are reduced to obtain the prediction equation.

$$\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k)$$

Autocorrelation tests for white noise

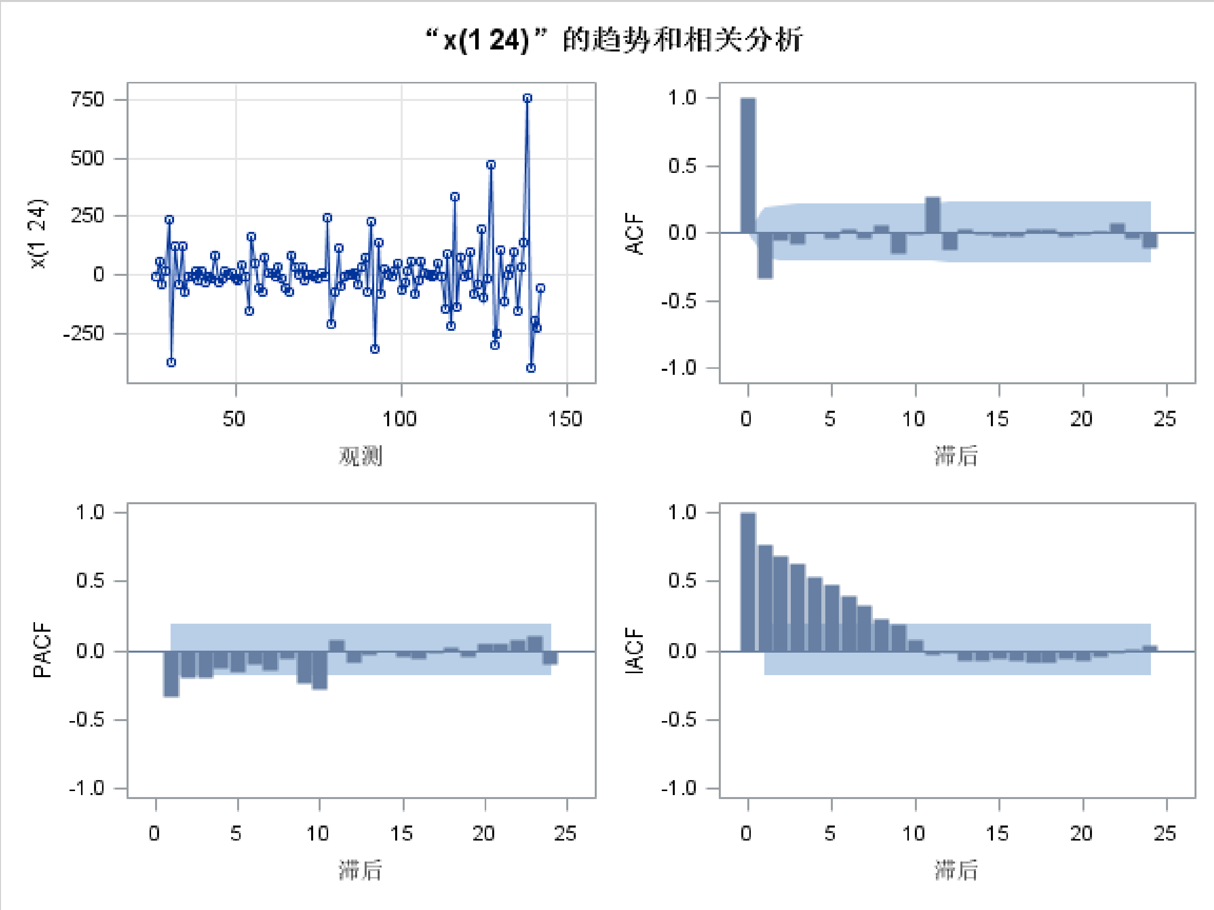
to lagging	**chi-square (math.)**	**(number of) degrees of freedom (physics and statistics)**	**Pr>cardinality**	**autocorrelation**					
6	15.30	6	0.0180	-0.340	-0.055	-0.076	0.005	-0.047	0.030
12	29.87	12	0.0029	-0.042	0.059	-0.150	-0.013	0.263	-0.121

18	30.39	18	0.0338	0.032	-0.007	-0.027	-0.025	0.024	0.027
24	33.39	24	0.0961	-0.028	-0.015	0.019	0.072	-0.039	-0.110

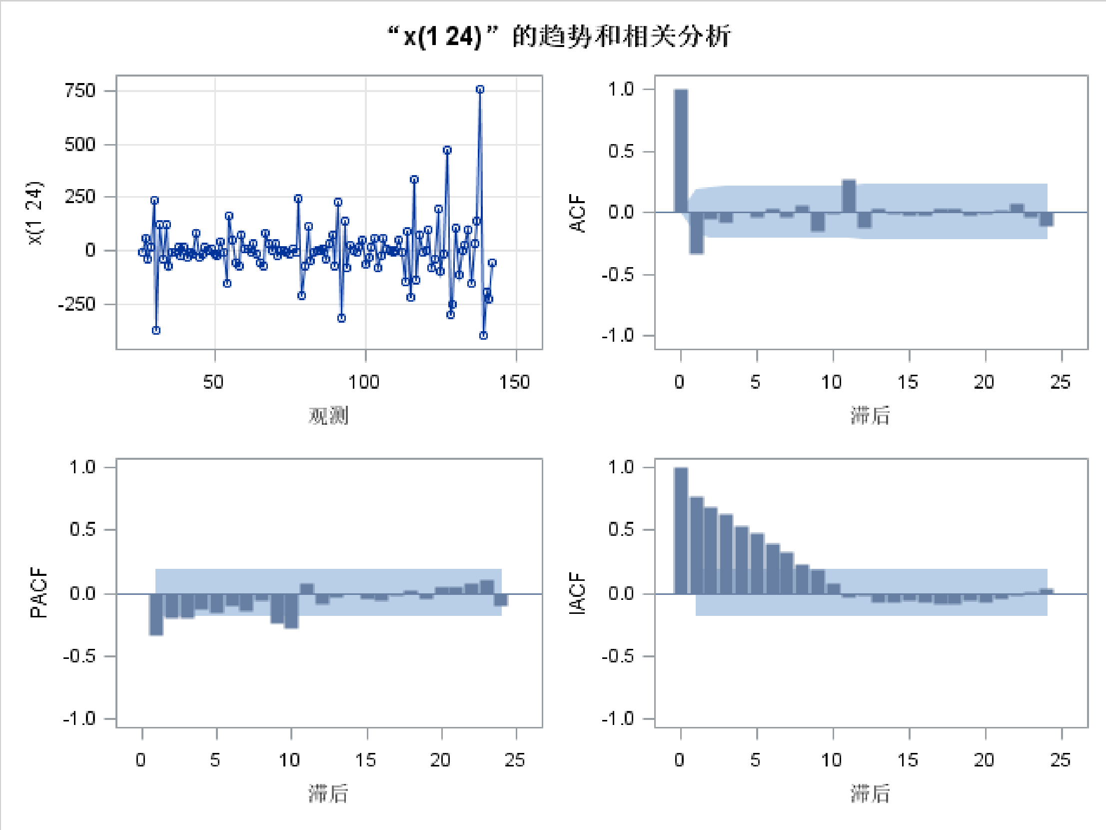
After doing the 1st order 24-step difference, the data series are smooth and non-white noise series, so there is some correlation between the data, and we decided to first try to use the simple seasonal model in the ARIMA model for prediction

Minimum Information Criterion

Lags	**MA 0**	**MA 1**	**MA 2**	**MA 3**	**MA 4**	**MA 5**
AR 0	9.747663	9.537035	9.538991	9.546053	9.580629	9.61256
AR 1	9.690874	9.536309	9.575201	9.586304	9.618093	9.650455
AR 2	9.690359	9.572615	9.595609	9.626905	9.653784	9.677521
AR 3	9.688936	9.611415	9.636076	9.645007	9.683245	9.715178
AR 4	9.676017	9.649806	9.675989	9.681059	9.721656	9.752609
AR 5	9.675632	9.678885	9.712052	9.718521	9.759127	9.792225

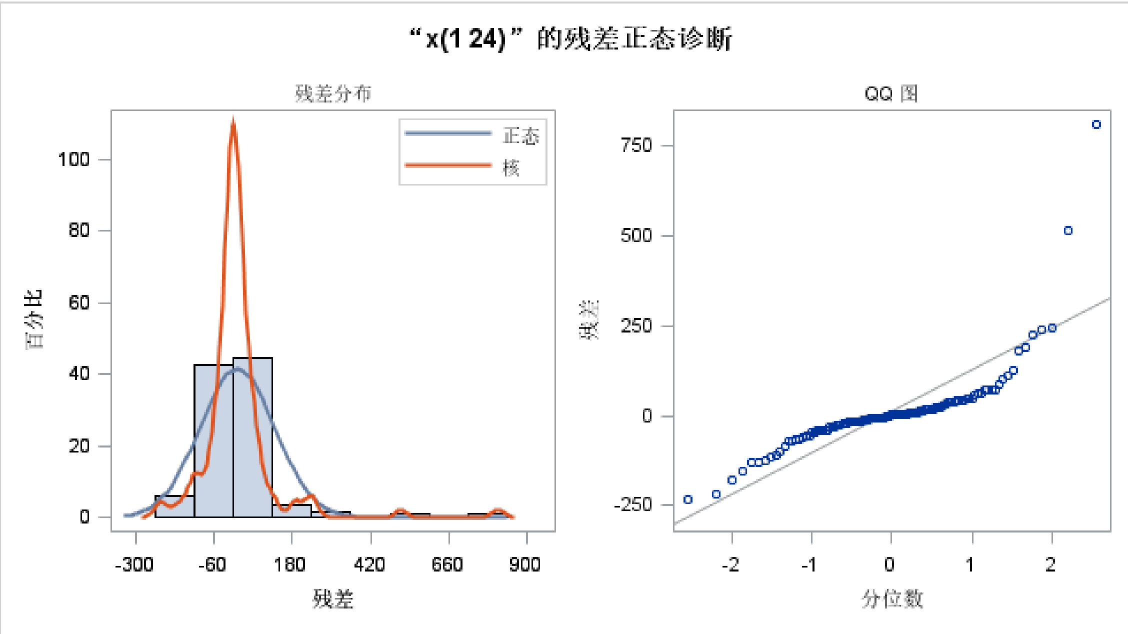


parameters	reckon	standardized inaccuracies	t-value	about the same as Pr > |t|	hysteresis
MA1,1	0.92839	0.05358	17.33	<.0001	1
AR1,1	0.29794	0.11525	2.59	0.0110	1

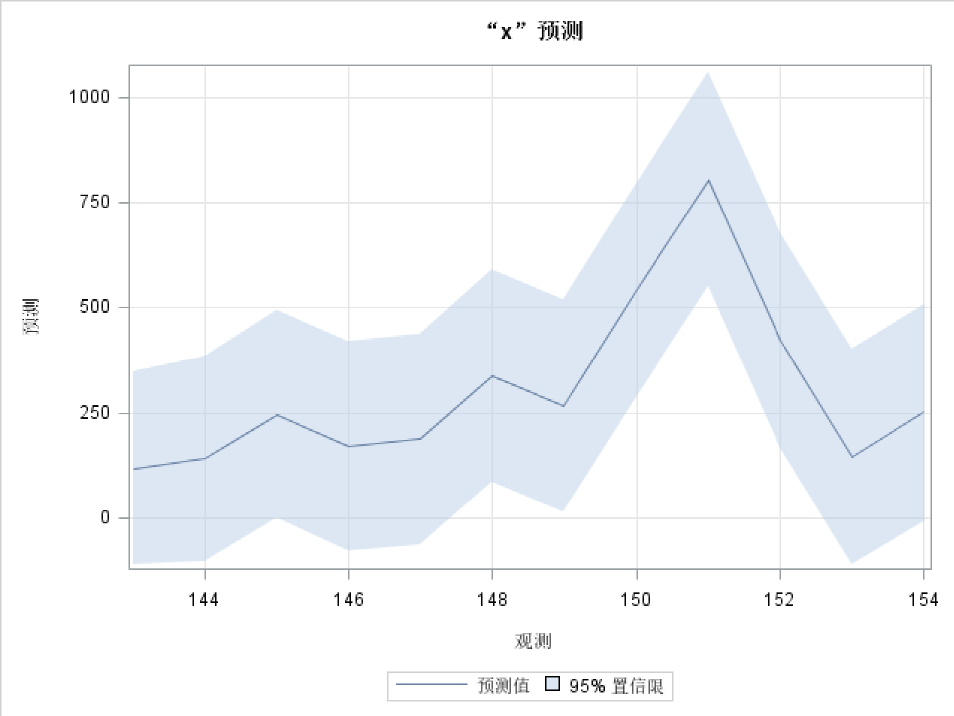


From the above table, it is clear that the model is optimal for p=1 and q=1. The optimal model is ARME(1,1) and the parameter estimation is performed next.

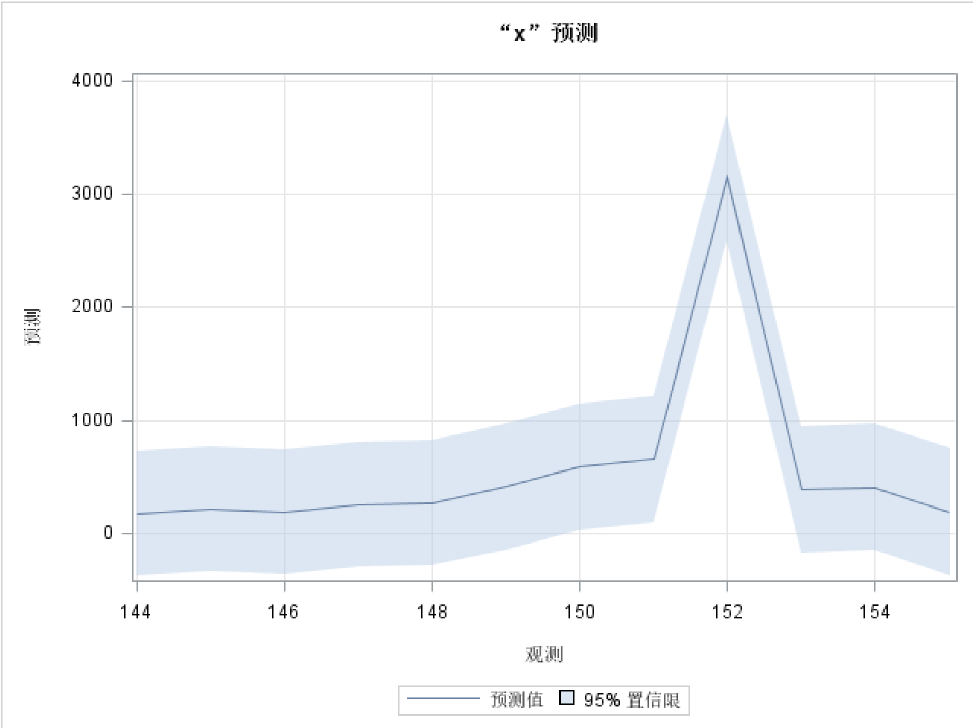
The parameters all passed the test, followed by a significance test of the model, i.e., a white noise on the residuals



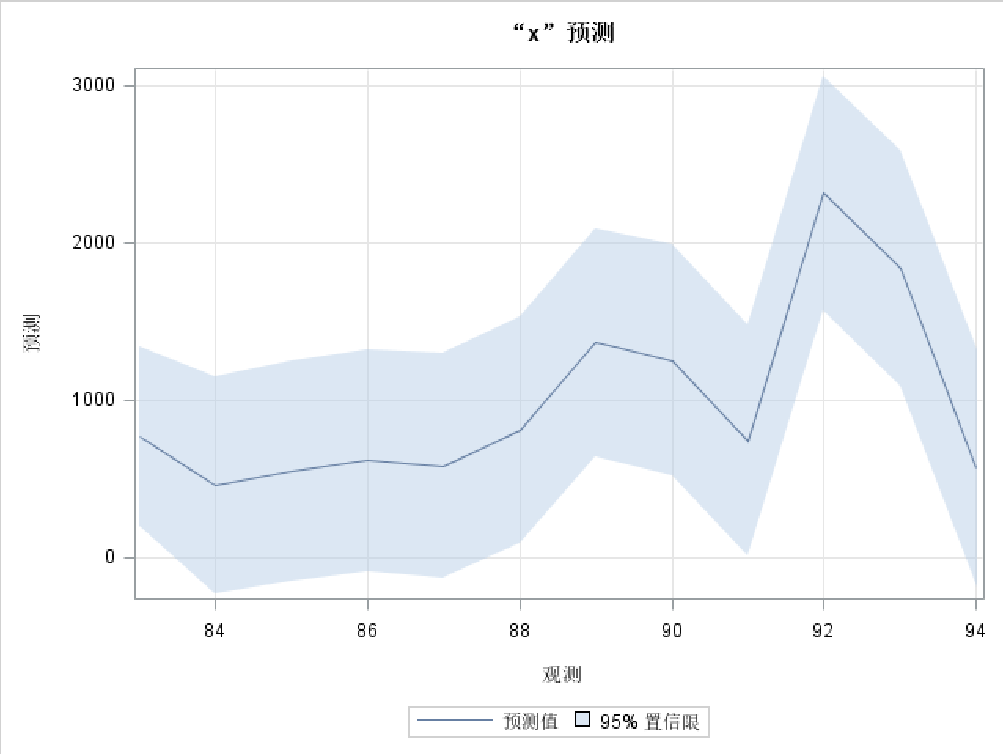
The residual white noise test passed and the model was good, followed by twelve months of precipitation predictions.



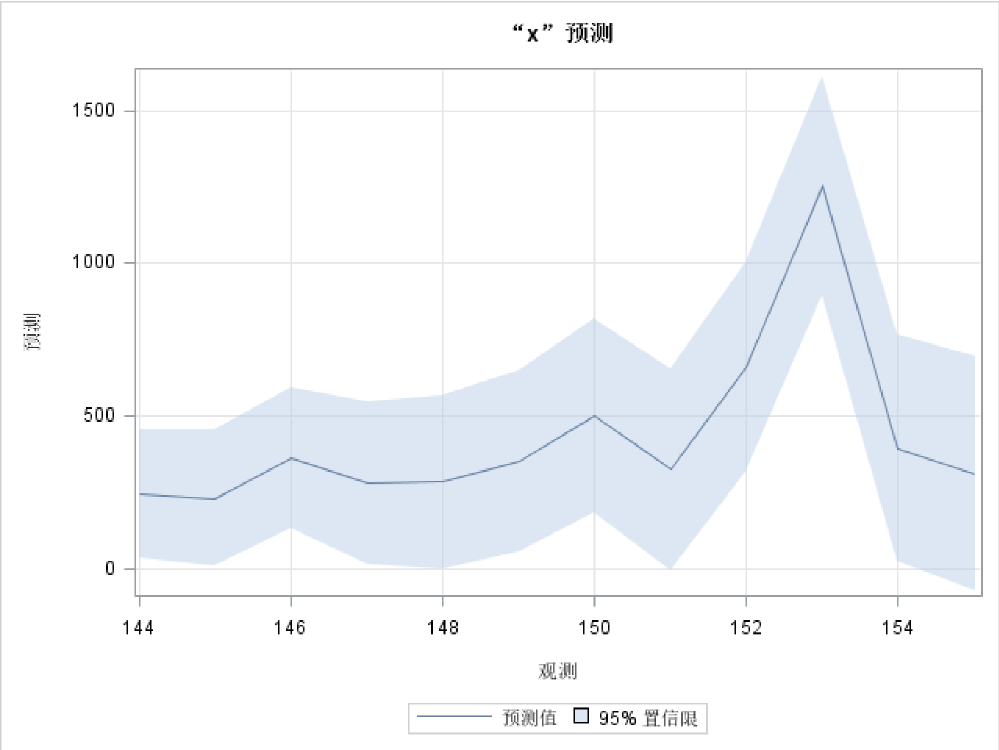
It is expected that Beijing may experience monthly precipitation in excess of your 750mm in month 9, with all predicted values falling within the 95% confidence interval, which is a good forecast.



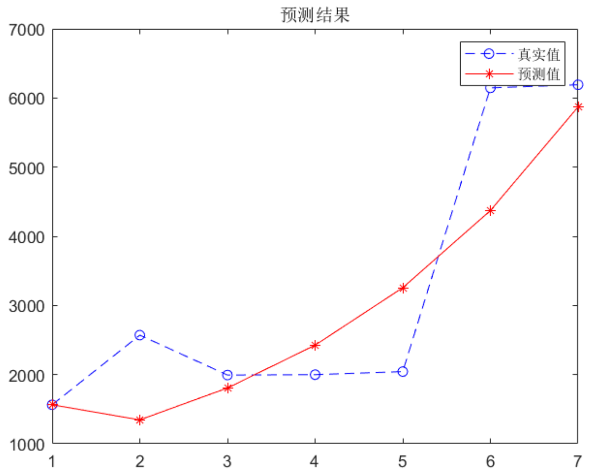
Extreme precipitation is predicted for Chengdu in the 10th month, with the possibility of monthly precipitation exceeding 3000 mm, and the predicted values all fall within the 95% confidence interval, which is a good forecast.



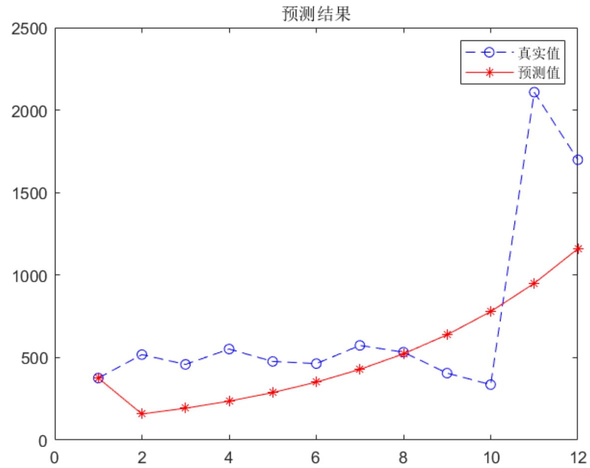
Extreme precipitation is predicted for Shenzhen in the 10th month, with the possibility of monthly precipitation exceeding 2000mm, and the predicted values all fall within the 95% confidence interval, which is a good forecast.



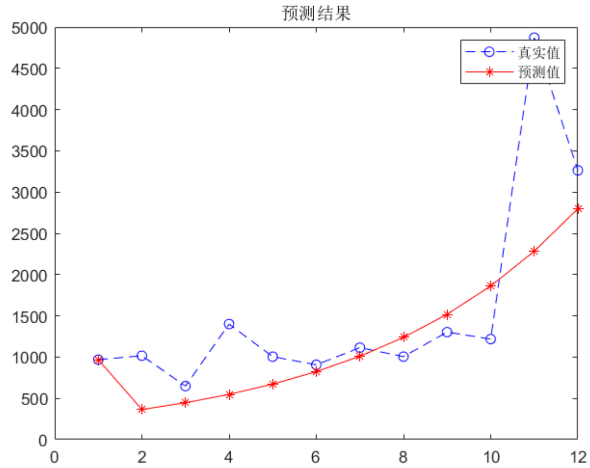
Extreme precipitation is predicted for Taiyuan in the 11th month, with the possibility of monthly precipitation exceeding 1000mm, and the predicted values all fall within the 95% confidence interval, which is good.



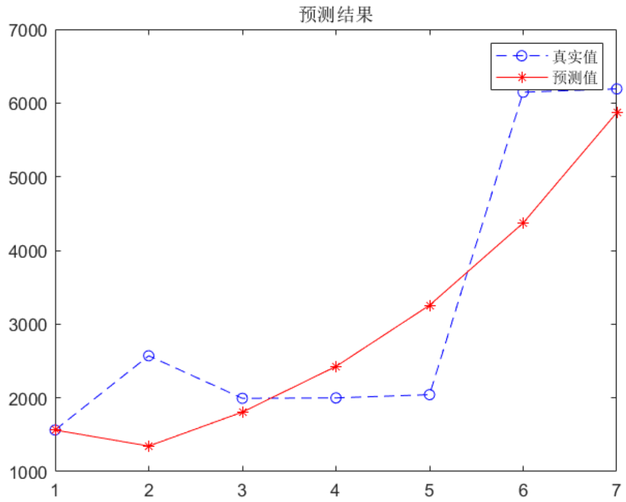
Grey forecast of annual precipitation in Shenzhen



Grey forecast of annual precipitation in Taiyuan



Grey forecast of annual precipitation in Chengdu



Grey forecast of annual precipitation in Beijing

The grey forecast analysis can only be used to describe annual precipitation trends and does not accurately reflect monthly precipitation trends. For this reason, the data derived from the seasonal time series forecasting model was chosen as the final data for our forecast results for the above cities.

4. 3. 2 Solving the model

After reducing the actual problem to a certain mathematical model, it is time to use the mathematical model to solve the proposed practical problem. Generally, you need to use computer software to solve the problem, such as Matlab, Spss, Lingo, Excel, Stata, Python, etc. After the solution is completed, the results should be standardized, accurate and eye-catching, and if they are too long, they should be compiled in an appendix. (Note: If you are using an intelligent optimization algorithm or a numerical computation method, you need to briefly explain the computational steps of the algorithm.)

4. 4 Problem 4 modeling and solving

4. 4. 1 Modeling

Mathematical models for factor analysis

Let there be m original indicators to be factor analyzed, denoted as $x_1, x_2, \ldots, x_m$. There are n samples with corresponding observations of x_{ik}, i=1 ,2,... ,n, k=1,2,...,m. ,m. After making the normalized transformation, transform xk to x_k^* i.e.

$$x_k^* = \frac{x_k - \bar{x}_k}{s_k}, \text{ the}$$

where $\bar{x}_k$ is the mean of x_k and s_k is the $_k$standard deviation of x.

The common factors in factor analysis are non-directly observable but objectively common influences, and each variable can be expressed as a linear function of the common factor and the sum of the special factors, i.e.

$$X_i = \mu_i + a_{i1}F_1 + a_{i2}F_2 + \ldots + a_{im}F_m + \varepsilon_i, (i = 1,2,\ldots, \quad p)$$

where is $F_1, F_2, \ldots, F_m$ called the common factor and is ε_i called the special factor of Xi. The model can be represented by the matrix as.

$$X - \mu = AF + \varepsilon$$

The following equation.

$$X = \begin{bmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_3^* \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{p1} & a_{p2} & \cdots & a_{pm} \end{bmatrix}, \quad F = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_3 \end{bmatrix}$$

4. 4. 2 **Solving the model**

Eigenvalues of the correlation matrix: Total = 10 Mean = 1

	eigenvalue (math.)	**differential**	**proportion**	**accumulate**
1	5.49826657	3.74731482	0.5498	0.5498
2	1.75095175	0.57488804	0.1751	0.7249
3	1.17606372	0.26829408	0.1176	0.8425
4	0.90776964	0.59904551	0.0908	0.9333
5	0.30872413	0.13570527	0.0309	0.9642
6	0.17301886	0.03622448	0.0173	0.9815
7	0.13679437	0.09011941	0.0137	0.9952
8	0.04667497	0.04522015	0.0047	0.9998
9	0.00145482	0.00117365	0.0001	1.0000
10	0.00028117		0.0000	1.0000

The results show that the cumulative contribution exceeds 80% and three factors should be selected and the contribution of each of the three factors is

0.5498, 0.1751, 0.1176.

Standardized scoring factor

	Factor1	**Factor2**	**Factor3**
x1	0.03647	0.29627	0.37306
x2	0.20997	-0.15092	0.04463
x3	-0.19108	0.04843	0.00538
x4	-0.19848	0.06661	-0.01168
x5	-0.06410	0.57688	-0.22214

	Standardized scoring factor		
	Factor1	**Factor2**	**Factor3**
x6	0.03146	-0.10357	0.66526
x7	0.19771	-0.04340	0.28667
x8	0.17657	-0.11054	-0.09196
x9	-0.08751	0.35940	0.00029
x10	0.11677	0.21348	-0.10827

The results show that the first common factor is closely related to X2, X7, X8, representing Po, Tx, VV, respectively. the second common factor is closely related to X5 and X9, representing Ff, Td, respectively; the third common factor is closely related to X1 and X6, T, Tn, respectively.

	Eigenvalues of the correlation matrix: total = 9 mean = 0.9			
	eigenvalue (math.)	**differential**	**proportion**	**accumulate**
1	4.42278772	2.52120666	0.4914	0.4914
2	1.90158107	0.88171023	0.2113	0.7027
3	1.01987083	0.17599582	0.1133	0.8160
4	0.84387501	0.35885158	0.0938	0.9098
5	0.48502344	0.25410828	0.0539	0.9637
6	0.23091515	0.13769493	0.0257	0.9893
7	0.09322022	0.09050921	0.0104	0.9997
8	0.00271101	0.00269546	0.0003	1.0000
9	0.00001555	0.00001555	0.0000	1.0000
10	0.00000000		0.0000	1.0000

The results show that the cumulative contribution exceeds 80% and three factors should be selected and the contribution of each of the three factors is

0.4914, 0.2113, 0.1133.

	Standardized scoring factor		
	Factor1	**Factor2**	**Factor3**
x1	0.35409	0.31317	-0.16821
x2	-0.26972	-0.05246	0.13666
x3	0.03321	-0.32840	0.09688
x4	0.03888	-0.32297	0.09116

	Standardized scoring factor		
	Factor1	**Factor2**	**Factor3**
x5	0.17706	-0.02735	0.32556
x6	0.24141	0.58457	-0.04705
x7	-0.08723	0.11322	0.34719
x8	-0.25793	-0.08212	0.00503
x9	0.00000	0.00000	0.00000
x10	-0.05548	-0.10681	0.62053

The results show that the first common factor X1 and X2 are closely related to T and Po, respectively; the second common factor is closely related to X6, X1, X4, and X3 with Tn, T, U, and P, respectively; and the third common factor is closely related to X10 with DD, respectively.

Comparing the above indicators related to each factor, it can be seen that the precipitation in Zhengzhou in July is the same as that in Taiyuan in October in that the precipitation is related to the atmospheric pressure Po at the level of weather stations.

Heavy rains in Shanxi show the following characteristics.

First, the accumulated rainfall of the storm was high. Fifty-one counties (cities and districts) in Shanxi Province received between 100 and 200 mm of precipitation, with a maximum cumulative precipitation of 285.2 mm.

Second, the rainstorm lasted for a long time. Starting at 23:00 on October 2, the precipitation continued for four days and ended at 23:00 on the 6th, with the heaviest precipitation periods occurring from the 4th to the 5th, with regional heavy rainfall occurring for two consecutive days in central Shanxi.

Third, precipitation extremes were prominent. A total of 59 national meteorological observation stations in the province had daily precipitation exceeding historical extremes for the same period since the establishment of the stations, and 63 national meteorological observation stations had cumulative precipitation exceeding historical extremes for the same period.

Finally, the storm was accompanied by lightning, strong convection and cooler temperatures. This

The heavy rainfall caused a total of 15 deaths in Shanxi due to the disaster, 3 people missing, crop damage of 3,576,900 mu, 19,500 collapsed houses, 18,200 severely damaged, direct economic losses of 5.029 billion yuan.

Heavy rainfall in Henan shows the following characteristics.

1., long duration: four consecutive days of widespread heavy precipitation, and heavy precipitation will continue.

2. High cumulative rainfall: the province's average rainfall was 113.5 mm.

3. The range of heavy precipitation is wide: there are 4,098 rainfall stations in the province with more than 50 mm of precipitation, 1,923 with more than 100 mm and 606 with more than 250 mm.

4, heavy precipitation period concentration: heavy precipitation period mainly from the night of the 18th, the northern, central and western areas of the general heavy rainfall, heavy rainfall, local very heavy rainfall.

5. Extreme in nature: precipitation amounts reached historical extremes.

According to official media reports, the heavy rains in Zhengzhou have caused the subway to stop running, several highways to be closed, trains and planes to stop running, and other such traffic is almost paralyzed! This rainstorm has brought huge economic losses to Zhengzhou. According to statistics, the direct economic losses in Zhengzhou, Henan Province, from the rare heavy rainfall exceeded 540 million yuan.

In addition, since July 16, this round of heavy rainfall caused 1240737 people affected in 560 townships in 89 counties (cities and districts) in Henan province, with 25 people killed and 7 people lost due to extreme rainstorms. Crop damage area of 75 thousand hectares, disaster area of 25.2 thousand hectares, 4.7 thousand hectares of crop failure.

4. 5 **Problem 5 modeling and solving**

4. 5. 1 **Solving the model**

The heavy rainfall in Zhengzhou has posed a very serious problem to urban construction, how can cities cope with extreme rainfall events in the future? First of all, we take Zhengzhou, the city with the most severe disaster, as a typical city for analysis and make relevant recommendations.

Zhengzhou is a national central city, located in the transition zone between the pre-mountainous hills in the southwest and the alluvial plain of the Yellow River in the east, with complex geographical and climatic conditions and frequent flooding, making it one of the 31 key cities in China for flood control.

First, the bottom line of national security must be safeguarded. At present, including Zhengzhou, a large number of power distribution facilities in the city, including strong and weak electricity, are arranged in the underground space, and in the event of an over-standard flood and water in the underground space, the city's major infrastructure, such as electricity and communications, will be interrupted, even leading to the paralysis of the entire city.

Second, corresponding standards should be established for graded classification. Graded and classified flood safety design standards should be established for various types of buildings, such as urban subways, residential communities, large office buildings, subways and underground garages, in order to ensure the safety of people's lives. At the same time, the planning and construction of urban subways, residential communities, large office buildings, subways, underground garages, etc. should be carried out by the relevant management departments for flood risk safety evaluation.

Third, a distinction should be made between the concept of "sponge city" construction and urban flood control. A "sponge city" is a city that can act like a sponge, absorbing, storing, seeping and purifying water when it rains, and releasing and using the stored water when needed, and is a low-impact development of urban rainwater systems. Therefore, we should attach great importance to the planning and design of urban flood control, and to the flood control of a national central city like Zhengzhou, we should attach importance to not only "drainage" but also to the water system connection for water consumption and balance, and more importance to "storage", which should be combined with The construction of "sponge cities" should be combined with the construction of wetland parks and urban flood control in low-lying areas on the outskirts of the city to create flood diversion and storage projects.

Fourth, urban upstream reservoirs should also be thoroughly strengthened and rehabilitated. Small, medium and large reservoirs where the risk of dam failure affects urban safety should be completely rebuilt to improve engineering resilience and mitigate natural disaster risks and losses.

Fifth, public awareness of risk avoidance and non-engineering early warning should be strengthened. In-depth information on the risk of flooding should be provided to the public through various means, and the set of flood monitoring, prediction and early warning, dispatching and decision-making, and disaster publicity should be strengthened.

Fourth, the city's upstream reservoirs should also be thoroughly strengthened and renovated. Small, medium and large reservoirs whose risk of dam failure affects the safety of the city should be thoroughly renovated to improve engineering resilience and mitigate natural disaster risks and losses.

Fifth, public awareness of risk avoidance and non-engineering early warning construction should be strengthened. In-depth information on the risk of flooding should be provided to the public through various channels to strengthen the set of flood monitoring, prediction and warning, scheduling In-depth information on the risk of flooding should be provided to the public through various channels to strengthen the set of flood monitoring, prediction and warning, scheduling, and disaster publicity. The Government should also ensure that the public is aware of the risk of flooding and that the public is informed of the risks of flooding.

五、 Analysis, testing and improvement of models

In problem one, the correlation coefficients between each meteorological indicator are calculated and the order of magnitude of the correlations is then obtained by grey correlation analysis.

In question two, find relevant information and analyze rainfall trends.

In problem three, the seasonal time series prediction model and the neural network prediction model were used to predict the monthly rainfall for possible heavy rainfall, and the annual rainfall analysis was predicted by grey prediction, the seasonal time series prediction model passed the t-test and F-test, the residuals passed the test, verifying that the model was reasonable and the predicted values were all within the 95% confidence interval, which was good; the neural network model was a "black box" through machine learning. The training characteristics of the sample to predict; gray prediction of annual rainfall ko shows a growing trend, can not be a good prediction of the possible extreme will be in the case. For this reason it is more reasonable to choose a seasonal time series forecasting model.

In question four, the same characteristics were obtained by factor analysis, the coefficient scores indicate the magnitude of the effect, the models both passed the t-test and the F-test, and the models were reasonable.

In question five, urban warning mechanisms and arrangements were made for the deployment of possible heavy rainfall by reviewing a large amount of information and analysis.

六、 bibliography

[1] Zhou S. F., Liu T. X., Duan L. M., Zhang W. R., Ji R., Sun L.. Analysis of simulated precipitation prediction based on time series in the Hailiu Rabbit River basin[J]. Soil and water conservation research,2021,28(05):88-94.

[2]Wu Wanqin,Qian Hong. Research on the prediction analysis of monthly precipitation in Kunming city--based on SARIMA model and Holt-Winters summation model[J]. Journal of Yunnan University for Nationalities (Natural Science Edition),2021,30(04):365-370.

[3]Zhou Xiaoxu. Application of hierarchical clustering-based LSTM neural network model for precipitation prediction in Jiangsu Province[D]. Shandong University,2020.

[4]Zhang Caihong. Analysis and prediction of precipitation trends in Weinan City based on ARIMA model[J]. Value Engineering,2019,38(34):197-199.

[5]Li Wenhui,Chen Yan,Liang Jiawen,Chen Guozhen. Application of time series homogeneous function model in precipitation prediction[J]. Guangdong Meteorology,2019,41(04):11-14.

[6]Chen Husheng,Zhou YL,Zhou P,Jin JL. Analysis and prediction of annual precipitation in Huangshan City based on wavelet and ARIMA[J]. South-North Water Diversion and Water Conservancy Science and Technology,2019,17(05):50-55.

[7]Yi LW,Wu S.A.,Zhang YJ. Research application of climate field principal component regression prediction model in flood precipitation prediction on Hainan Island[J]. Journal of Hainan University (Natural Science Edition),2019,37(02):165-171.

[8]Zhou S. F.,Liu T. X.,Duan L. M.,Zhang W. R.,Ji R.,Sun L.. Analysis of simulated precipitation prediction based on time series in the Hailiu Rabbit River basin[J]. Soil and water conservation research,2021,28(05):88-94.

[9]Wu Wanqin,Qian Hong. Research on the prediction analysis of monthly precipitation in Kunming city--based on SARIMA model and Holt-Winters summation model[J]. Journal of Yunnan University for Nationalities (Natural Science Edition),2021,30(04):365-370.

[10]Kumar V.,Vannan Mani. it takes two to tango: Statistical modeling and machine learning[J]. Journal of Global Scholars of Marketing Science,2021,31(3):

APPENDIX

```
data ex;
input x1-x10 x11$@@;
cards;
2.1  221.3  5944.3  6019.9  517  14  202.4  211.5  8  161.4  0
95   207.6  5948.6  6024.8  605  15  196    216.1  8  167.7  0
84   223.9  5940.8  6016.7  564  15  185.6  221.5  8  171.5  0
0    239.4  5948.4  6023.6  493  21  197.6  272    8  170    0
0    229.1  5955.4  6030.9  506  19  194.7  270.1  8  166.3  0
0    227.4  5950.2  6025.9  591  16  194.6  260.4  8  183.2  0
0    241.3  5954.6  6029.8  551  14  203.6  267.5  8  186.7  0
0    240.7  5946.1  6021.3  488  17  211.3  287.6  8  169.1  0
0.4  253.7  5934.4  6009.4  448  15  186.5  305.2  8  168.3  0
0    252.1  5937.6  6012.2  628  19  214.7  299.3  8  216.2  0
15.6      230.5  5925.8  6000.9  743  21  219.7  164.5  8  219.9  0
4    256.4  5905.3  5979.8  555  19  214.4  275.5  8  197.9  0
0    263.4  5902.4  5976.2  516  22  223.7  305.8  8  197.9  0
36   250.4  5928.8  6003.3  639  21  222.3  303.7  8  213.4  0
88   237    5958.6  6034.1  606  15  198.6  259.4  8  195    0
0.4  235.5  5966.6  6042.3  669  16  214.5  267.3  8  209.6  0
0    222.7  5956.4  6032.3  680  21  208.3  250.7  8  199.2  0
41   207.8  5957.1  6033.3  793  16  196.1  208.6  8  206.6  0
215       200.9  5965.1  6041.7  800  13  195.9  136.3  8  200.9  0
438       189.7  5967.6  6044.6  800  21  187.1  205.9  8  189.7  0
378       197.6  5965    6042    800  19  182.2  159.9  8  197.6  0
114  206.8  5967.3  6043.9  772  16  189.6  196.1  8  201.6  0
0.4  201.6  5972.2  6049    792  13  190.5  168.8  8  200.2  0
0    222.5  5964.8  6041    662  9   191    244.4  8  193.1  0
0    228.2  5948    6023.7  626  10  183.3  266.5  8  189.4  0
0    235.5  5926.3  6001.5  601  8   187.2  274.6  8  190.4  0
0.5  234.4  5925    6000.1  649  16  196    275.6  8  202.2  0
2.4  222.1  5922.8  5997.8  650  24  205.7  230.4  8  190.7  0
0    218.1  5914.6  5989.8  538  20  197.3  250.3  8  163.5  0
0    238.2  5907.2  5982.4  581  12  183.8  267.1  8  190.5  0
0    256.4  5902.9  5977.1  595  23  202.9  295.1  8  209.4  0


proc factor data=ex method=principal rotate=varimax percent=0.8 score outstat=ex1;
var x1-x10 x11$;run;
proc score data=ex score=ex1 out=ex2;
var x1-x10 x11$; run;
proc print data=ex1; proc print data=ex2;
```

```
run;

data ex;
input x1-x10 x11$@@;
cards;

3	111.4	2791.1	3060.8	649	14	9.5	20	8	84.7	0
0	113.2	2791.9	3061.6	415	20	10	17.8	8	28.7	0
0	106.2	2798.9	3070	355	11	6.8	18.4	8	6.7	0
0	99.6	2797.5	3069.8	476	9	4.3	20.2	8	29.7	0
0	83.1	2787.9	3058.2	482	9	4.8	21.5	7	38.3	1
0	114.2	2782.2	3051.5	549	9	6.1	23.2	8	59.5	0
0	125.2	2780.9	3048.5	499	9	9.1	24.7	8	58.5	0
0	129.9	2778.1	3044.7	538	7	9.6	25.1	8	73.3	0
0	123.3	2769.1	3034.8	587	9	9.9	24.9	8	79.4	0
5	114.1	2769.8	3036	597	11	11	23.2	8	77.8	0
0.1	103.3	2789.3	3059.9	383	13	7.7	17.8	8	14.8	0
0	97.7	2796.1	3067.4	459	14	8.8	18.2	8	30.7	0
0	79.8	2790.4	3062.7	508	9	5.9	15.6	7	42.3	1
0	98.1	2793.4	3064.9	311	11	6.9	17.7	7	5.8	1
0	85	2792.4	3065	419	11	1.9	19.3	8	0	0
0	96.1	2785.5	3057.3	484	10	3.2	21.1	8	30	0
0	85	2797.5	3069.7	571	9	5.8	21.5	8	43.4	0
0	81.1	2797.5	3071.2	620	12	6.7	16.9	8	50.5	0
0	93.6	2799.3	3073.5	503	13	4.8	14.7	8	29	0
0	81.2	2800.4	3073.7	576	10	4.8	19.9	8	41.9	0
0	94.4	2790.6	3062.4	600	7	8.6	15.4	8	58.9	0
0	99.4	2784.2	3055.8	562	8	5.9	15.6	8	52.7	0
5.4	96.7	2785.3	3055.3	672	12	10.7	19.5	8	75.5	0
7.7	72.4	2792.6	3065.5	637	19	8.5	14	8	45.2	0
0	47.5	2819.8	3100.5	306	20	2.1	12.2	8	-58.8	0
0	36.6	2829.7	3111.7	409	13	3	10.2	8	-40	0
0	43.1	2817.1	3098	539	14	3.1	6.7	8	-5.1	0
0	43.5	2817.4	3098.6	518	11	-0.7	9.9	8	-12.8	0
0	45.1	2817.9	3099.6	517	11	-1.4	12.8	8	-12.1	0
0	49	2814.7	3094.8	517	10	-0.8	13	8	-8.3	0
0	66.1	2806.9	3084.8	457	10	1.8	13.8	8	-1.4	0

proc factor data=ex method=principal rotate=varimax percent=0.8 score outstat=ex1;
var x1-x10 x11$;run;
proc score data=ex score=ex1 out=ex2;
var x1-x10 x11$; run;
proc print data=ex1; proc print data=ex2;
run;
```

```
clear
syms a u;
c=[a,u]';%compose the matrix
A=[539.2
720.7
729.3
608.6
485.6
479.1
702.2
675.4
569.7
436.7
2015.5
2457.4
]';%Enter data that can be modified
Ago=cumsum(A);% The original data are accumulated once to get the 1-AGO sequence xi(1).
n=length(A);%Number of original data
for k=1:(n-1)
    Z(k)=(Ago(k)+Ago(k+1))/2; %Z(i) is the sequence of immediately adjacent mean generators of xi(1)
end
Yn =A;%Yn is a vector of constant terms
Yn(1)=[]; % Starting from the second number, i.e. x(2),x(3)...
Yn=Yn';
E=[-Z;ones(1,n-1)]';% accumulate the generated data to make the mean
c=(E'*E)\(E'*Yn);% Use the formula to find a, u
c= c';
a=c(1);% get the value of a
u=c(2);% get the value of u
F=[];
F(1)=A(1);
for k=2:(n)
    F(k)=(A(1)-u/a)/exp(a*(k-1))+u/a;% find the GM(1,1) model equation
end
G=[];
G(1)=A(1);
for k=2:(n)
    G(k)=F(k)-F(k-1);% both do difference reduction of the original series to get the predicted data
end
t1=1:n;
t2=1:n;
```

```
plot(t1,A,'bo--');
Hold on;
plot(t2,G,'r*-');
title('predicted results');
legend('true value','predicted value');
% post-test
e=A-G;
q=e/A;% relative error
s1=var(A);
s2=var(e);
c=s2/s1;% variance ratio
len=length(e);
p=0; % small error probability
for i=1:len
    if(abs(e(i))<0.6745*s1)
        p=p+1;
    end
end
p=p/len;
clear
syms a u;
c=[a,u]';%compose the matrix
A=[969
1016.5
648.1
1400.5
1005.1
906.6
1114.6
1005.5
1302.6
1218.5
4871.5
3263.6


]';%Enter data that can be modified
Ago=cumsum(A);% The original data are accumulated once to get the 1-AGO sequence xi(1).
n=length(A);%Number of original data
for k=1:(n-1)
    Z(k)=(Ago(k)+Ago(k+1))/2; %Z(i) is the sequence of immediately adjacent mean generators of xi(1)
end
```

```
Yn =A;%Yn is a vector of constant terms
Yn(1)=[]; % Starting from the second number, i.e. x(2),x(3)...
Yn=Yn';
E=[-Z;ones(1,n-1)]';% accumulate the generated data to make the mean
c=(E'*E)\(E'*Yn);% Use the formula to find a, u
c= c';
a=c(1);% get the value of a
u=c(2);% get the value of u
F=[];
F(1)=A(1);
for k=2:(n)
    F(k)=(A(1)-u/a)/exp(a*(k-1))+u/a;% find the GM(1,1) model equation
end
G=[];
G(1)=A(1);
for k=2:(n)
    G(k)=F(k)-F(k-1);% both do difference reduction of the original series to get the predicted data
end
t1=1:n;
t2=1:n;
plot(t1,A,'bo--');
Hold on;
plot(t2,G,'r*-');
title('predicted results');
legend('true value','predicted value');
% post-test
e=A-G;
q=e/A;% relative error
s1=var(A);
s2=var(e);
c=s2/s1;% variance ratio
len=length(e);
p=0; % small error probability
for i=1:len
    if(abs(e(i))<0.6745*s1)
        p=p+1;
    end
end
p=p/len;

clear
syms a u;
c=[a,u]';%compose the matrix
```

```
A=[1563.7
2571
1993.3
1999.2
2043.5
6146.7
6191.4
]';%Enter data that can be modified
Ago=cumsum(A);% The original data are accumulated at once to obtain the 1-AGO sequence xi(1).
n=length(A);%Number of original data
for k=1:(n-1)
    Z(k)=(Ago(k)+Ago(k+1))/2; %Z(i) is the sequence of immediately adjacent mean generators of xi(1)
end
Yn =A;%Yn is a vector of constant terms
Yn(1)=[]; % Starting from the second number, i.e. x(2),x(3)...
Yn=Yn';
E=[-Z;ones(1,n-1)]';% accumulate the generated data to make the mean
c=(E'*E)\(E'*Yn);% Use the formula to find a, u
c= c';
a=c(1);% get the value of a
u=c(2);% get the value of u
F=[];
F(1)=A(1);
for k=2:(n)
    F(k)=(A(1)-u/a)/exp(a*(k-1))+u/a;% find the GM(1,1) model equation
end
G=[];
G(1)=A(1);
for k=2:(n)
    G(k)=F(k)-F(k-1);% both do difference reduction of the original series to get the predicted data
end
t1=1:n;
t2=1:n;
plot(t1,A,'bo--');
Hold on;
plot(t2,G,'r*-');
title('predicted results');
legend('true value','predicted value');
% post-test
e=A-G;
q=e/A;% relative error
```

```
s1=var(A);
s2=var(e);
c=s2/s1;% variance ratio
len=length(e);
p=0; % small error probability
for i=1:len
    if(abs(e(i))<0.6745*s1)
        p=p+1;
    end
end
p=p/len;
clear
syms a u;
c=[a,u]';%compose the matrix
A=[375.8
518.1
458.5
551.8
476.5
463.3
573.7
532.6
405.6
336.2
2108.6
1698.7



]';%Enter data that can be modified
Ago=cumsum(A);% The original data are accumulated at once to obtain the 1-AGO sequence xi(1).
n=length(A);%Number of original data
for k=1:(n-1)
    Z(k)=(Ago(k)+Ago(k+1))/2; %Z(i) is the sequence of immediately adjacent mean generators of xi(1)
end
Yn =A;%Yn is a vector of constant terms
Yn(1)=[]; % Starting from the second number, i.e. x(2),x(3)...
Yn=Yn';
E=[-Z;ones(1,n-1)]';% accumulate the generated data to make the mean
c=(E'*E)\(E'*Yn);% Use the formula to find a, u
c= c';
a=c(1);% get the value of a
```

```
u=c(2);% get the value of u
F=[];
F(1)=A(1);
for k=2:(n)
    F(k)=(A(1)-u/a)/exp(a*(k-1))+u/a;% find the GM(1,1) model equation
end
G=[];
G(1)=A(1);
for k=2:(n)
    G(k)=F(k)-F(k-1);% both do difference reduction of the original series to get the predicted data
end
t1=1:n;
t2=1:n;
plot(t1,A,'bo--');
Hold on;
plot(t2,G,'r*-');
title('predicted results');
legend('true value','predicted value');
% post-test
e=A-G;
q=e/A;% relative error
s1=var(A);
s2=var(e);
c=s2/s1;% variance ratio
len=length(e);
p=0; % small error probability
for i=1:len
    if(abs(e(i))<0.6745*s1)
        p=p+1;
    end
end
p=p/len;
goptions vsize=7cm hsize=10cm;/* less data in Guiyang*/
data a;
input x@@;
dif=dif(x);
time=_n_;
cards;
2.9
22.3
15.7
32.9
89.5
```

35.5
184
82.3
62.7
0
0
0
12.5
0
16.2
22.2
117.6
265.6
169.8
63.4
28.7
21.5
3.2
0
0
8.2
55.2
31.3
104.7
284.2
58.4
79.3
22.3
78.2
7.5
2.7
4
9.9
5.9
24.5
91.4
232.3
119.9
93.3
24.5
0.2
0
0
4.9

0.9
21.4
34.7
96.6
123
59.9
129.8
14.4
0
0
0.8
11
8
35.2
34.4
48.4
113.4
82.8
92.3
22.1
29
1.7
0.2
8.9
0
5.5
24.7
77.8
350.3
76.7
75.6
73.3
9.2
0
0.2
8
11.4
0
31.2
120.3
114.9
311.9
3.3
74.2

0
0
0
0
3.7
54.6
9.9
34.2
323.9
111.3
26.8
4.4
0.7
0.2
0
2
2.3
40
65.3
10.8
94.4
69.3
91.4
22.7
21.2
17.3
13.9
111.3
34.8
49.1
199.9
128.7
404.8
666.8
285.3
8
112.9
0
3.2
26.3
121.6
6.4
65
150.8

```
991.7
566
391.5
96.7
38.2
.

proc gplot;
plot x*time dif*time;
symbol c=black i=join v=star;
proc arima;
identify var=x(1,24) minic p=(0:5) q=(0:5);
estimate p=1 q=1 noint;
forecast lead=12;
run;
goptions vsize=7cm hsize=10cm;
data a;
input x@@;
dif=dif(x);
time=_n_;
cards;


0.3
3.9
22.5
24.7
74.3
113.9
242.8
298
154.9
18.5
7.2
8
16.6
8.2
38.8
8.4
237.6
76
287.5
114.5
153.8
```

46.3
20.9
7.9
15.6
7.5
26.6
18.3
125.6
82.2
151.6
65.7
82.3
59.2
10.7
2.8
1.2
4.6
8.2
45.1
104.1
187.2
559.1
229.7
201.5
28.6
23.4
7.8
1.1
11.4
13.5
52.6
55.1
139.1
392.5
180.5
53.8
78.7
21.2
5.6
2.7
3.7
10
92
62.2

90.4
97.3
291.5
208.1
23
9.7
16
12.2
27.4
34.5
62.6
89.8
81.3
362.8
273.3
131.8
16.9
20.3
1.7
4.2
20.7
30
71.8
36.1
73.7
386.3
267.8
39.4
61
12.9
1.6
2.5
13.1
38
53.5
129.4
257.2
383.3
180.7
176.4
35.4
19.3
13.8
6.1

```
10.8
37.2
46.6
129.6
44.8
362.4
226.3
257.3
73.5
23.6
0.3
39.6
12.3
85.5
99.8
240.8
416.1
483.5
2972.5
214.4
235.3
15.7
56
10.7
68.9
200
221.1
218.5
186.2
503.2
1221.4
433.4
176.3
23.9

.
proc gplot;
plot x*time dif*time;
symbol c=black i=join v=star;
proc arima;
identify var=x(1,24) minic p=(0:5) q=(0:5);
estimate q=1 noint;
forecast lead=12;
run;
```

```
goptions vsize=7cm hsize=10cm;
data a;
input x@@;
dif=dif(x);
time=_n_;
cards;

54.2
16.6
11.5
63.2
311.1
140.5
281.9
238.1
181
170.4
15.2
80
247.7
40.8
123.8
221.6
386
218.4
220.5
503.4
137.6
353.5
113.1
4.6
5
33.6
54.6
68.8
229.2
448.7
508.2
314.3
207
103.6
20.3
0
86.6
```

4.7
17
83.5
119.9
437
306.6
546.6
306.1
23.7
58.7
8.8
86.6
4.7
17
83.5
119.9
437
306.6
546.6
306.1
23.7
58.7
8.8
98.8
166.3
131.9
362.8
917.6
802.7
293.5
1864.2
1386.7
115.4
6.7
0.1
0
220.9
68
121.6
338.7
1662.5
1444.4
1133.7
445.5

752

```
.
proc gplot;
plot x*time dif*time;
symbol c=black i=join v=star;
proc arima;
identify var=x(1,24) minic p=(0:5) q=(0:5);
estimate q=2 noint;
forecast lead=12;
run;
goptions vsize=7cm hsize=10cm;
data a;
input x@@;
dif=dif(x);
time=_n_;
cards;
```

Monthly precipitation
0
8.5
6.5
27.4
25.9
4.1
51
124.4
106.8
21.2
0
0
1.9
7.2
1
7.8
33.7
44.9
140.1
101.1
100.1
38.1
41.7

0.5
1.1
0.3
8.7
19
11.3
61.5
213.2
29
90.4
12.4
5.8
5.8
2.1
1.5
0
33.7
13.6
94.8
222.5
57.1
106.5
14
6
0
0
11
19.6
23.1
53.4
76.3
104.9
87.5
84.4
9.2
7.1
0
6
19.1
1
40.7
36.8
11.8
49.3

155.6
84.1
29.3
25.8
3.8
2.7
5.6
3.5
53.1
23.5
77.2
219.9
79.4
20.7
63
7.4
17.7
2.2
9.9
5.8
25.3
15.4
47.1
146.6
125.7
1.2
151.9
0
1.5
2.5
0
4.5
23
31.8
31
182.5
63.9
60.5
2.9
2.8
0.2
0
1.1
0

```
52.3
2.5
29.9
45.7
62.3
97
34.1
5.1
6.2
49.2
15.9
31.5
66.8
230
60.8
385.8
983.2
126.9
38.7
114.9
4.9
0
53
125.1
41.6
137.2
194.5
234.5
191.1
249.6
444.2

.
proc gplot;
plot x*time dif*time;
symbol c=black i=join v=star;
proc arima;
identify var=x(1,24) minic p=(0:5) q=(0:5);
estimate p=2 noint;
forecast lead=12;
run;
```