ISSN 2616-5775 Vol. 6, Issue 1: 32-41, DOI: 10.25236/AJCIS.2023.060106

Study on Rainfall Distribution in Yunnan Province Based on ARIMA Model by Big Data Computation

Kaimin Li*, Guang Li, Yuanyuan Zhang

Baoshan University, Baoshan, Yunnan, 678000, China
**Corresponding author*

Abstract: *Taking the monthly rainfall data of 16 prefectures and cities in Yunnan Province from 2012 to 2020 as the research object, an ARIMA time series model is established based on big data calculation. First, we use the cluster analysis method to divide all regions into three categories according to the distribution characteristics of rainfall and select a representative city from each category to study the rainfall distribution. Then, according to the results of the ADF test, the method using phase and seasonal differences is used to eliminate the non-stationary and seasonal trends of the series. Finally, the ARIMA prediction model of rainfall distribution in Kunming, Dali, and Pu'er is obtained by combining the sequence autocorrelation and partial autocorrelation analysis diagram to determine the values of various parameters in the ARIMA model. The prediction accuracy of the model is high, and the residual sequence is a white noise sequence, which has a good fitting effect. It effectively predicts the fluctuation law of rainfall and provides an early warning mechanism for drought, flood, debris flow, and other disasters.*

Keywords: *rainfall, cluster analysis, ADF test, ARIMA model by big data computation*

1. Introduction

With the increasing trend of global warming, extreme climates such as El Nino and La Nina occur alternately, and the resulting natural disasters also occur frequently. The frequency, area, and intensity of regional natural disasters in Yunnan Province show an increasing trend. Natural disasters lead to a significant increase in the rate of direct and indirect economic losses. Yunnan has a large monsoon climate and a three-dimensional temperature difference between day and night. Due to the large monsoon climate and the small temperature difference between day and night in South Asia, Yunnan has a clear three-dimensional distribution in the dry season and the rainy season. The rainy season is from May to October every year, and the rainfall accounts for about 85% of the whole year. The dry season is from November to April, and the rainfall accounts for about 85% of the whole year [1]. The rainfall shows obvious seasonal fluctuation differences and uneven spatial distribution. Therefore, the study on the distribution of rainfall in Yunnan Province plays a positive role in promoting the early warning of regional drought, flood, debris flow, and other disasters. ARIMA model is a common method to study time series in statistics, as it predicts the future change trend of the system [2]. Many scholars have applied the ARIMA model to the research of Meteorology and achieved fruitful results. Ma [3] used long-term meteorological data to establish the prediction model of time series. The simulation results of precipitation and wind speed of meteorological data show that the performance of the ARIMA seasonal model is better than the ARMA model. Xi [4] established a time series prediction model based on the precipitation data of Wuwei County from 1957 to 2016. Wang [5] used the wavelet analysis method to establish the ARMA-GARCH model (W-A-G) based on wavelet analysis for precipitation in the Tongyu area from 1955 to 1999. The prediction results show that W-A-G model is a feasible method. Therefore, using the monthly rainfall data of 16 prefectures and cities in Yunnan Province from January 2012 to December 2020 as the research object (data source: Yunnan Meteorological Bureau), we establish ARIMA model to analyze the distribution of rainfall in Yunnan Province.

ISSN 2616-5775 Vol. 6, Issue 1: 32-41, DOI: 10.25236/AJCIS.2023.060106

2. Cluster Analysis of Rainfall Distribution in Yunnan Province

2.1. Cluster analysis

First, the number of clusters is determined. Using the method of systematic clustering, each observation first forms a-class. The average distance between each observation in the sub-class is defined as the between-groups linkage distance, and the individuals with the smallest between-groups linkage distance are clustered into a class. Next, the distance between the remaining individuals is measured and the subclass again to condense the currently closest individuals or subclasses into one class. For n individuals, the condensation process is repeated n-1 times until all individuals gather into a large class.

After determining the number of clusters, the K-Means cluster analysis method is used to determine the specific regional members in each class. K-Means clustering analysis method is to divide the observations into K groups into the most vicious classes according to the given rules. The rainfall data is regarded as a point in p-dimensional space, and the square Euclidean distance is used to measure the degree of affinity and alienation between all individuals. The steps are as follows.

The first step is to specify the number of clusters K and select k initial center points.

The second step is to calculate the distance between each data point and K centers and assign it to the center point closest to it.

The third step is to recalculate the average value of the distance from the point in each class to the center point of the class and take the average point as the center of K classes to complete an iteration.

The fourth step is to judge whether the conditions for terminating cluster analysis are met.

Steps 2 and 4 are repeated until all observations have been allocated or the given maximum number of iterations has been reached.

2.2. Cluster analysis results

The monthly rainfall data of 16 prefectures and cities in Yunnan Province from January 2012 to December 2020 is selected to calculate the Pearson correlation coefficient of these regions. The correlation coefficient is between 0.484−0.873, and several data show a high correlation. Next, the cluster analysis method is used to classify the regions with similar rainfall distribution characteristics. Finally, according to the condensed state table in the results of systematic cluster analysis, the gravel map is drawn.

With the continuous aggregation of each sub-category, the distance between different categories is gradually increasing, and the number of clusters is gradually decreasing (Fig. 1). Before clustering into 3 classes, the distance between different classes increases slightly, and the distribution of condensation points in the gravel map is relatively steep. However, after clustering into 3 classes, the distance between classes increases rapidly, and the distribution of condensation points in the gravel map has been relatively flat. According to the principle that the distance between regions with large similarity is small, and the distance between regions with small similarity is large. Taking the "inflection point" in the gravel map as the basis of classification number, 16 regions are considered to be clustered into three categories.

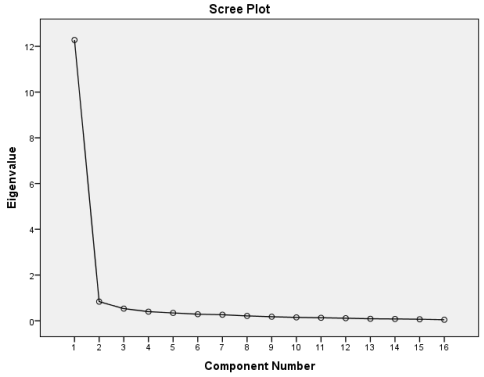

Figure 1: Cluster analysis of 16 districts in Yunnan Province.

ISSN 2616-5775 Vol. 6, Issue 1: 32-41, DOI: 10.25236/AJCIS.2023.060106

Using the effective algorithm proposed by Hartigan and Wong [6], the following equation is obtained.

$$ss(k) = \sum_{i=1}^{n} \sum_{j=1}^{p} (x_{ij} - \overline{x}_{kj})^2 \quad (1)$$

where x_{ij} represents the value of the j-th variable in the i-th observation value, $\overline{x}_{ij}$ represents the mean value of the j-th variable in the k-th class, p indicates the number of variables, and $ss(k)$ represents the distance between each observation and the center point in steps 2 and 4.

Table 1: Results of K-means cluster analysis

	Districts
Class I	Pu'er Dehong
Class II	Wenshan Nujiang Dali Xishuangbanna Lincang Baoshan
Class III	Kunming Zhaotong Qujing Honghe Yuxi Lijiang Shangri La Chuxiong

According to Table 1 the results of the K-means cluster analysis, category I is the sunny hot valley area represented by Pu'er, which belongs to the type of subtropical monsoon climate, with an annual average rainfall of about 1400−1700 mm. The II type is the low latitude plateau monsoon climate area represented by Dali, with an average annual precipitation of about 1000 mm, belonging to the region with medium rainfall in the province. Category III is the north subtropical low latitude plateau mountainous monsoon climate area represented by Kunming, with an average annual precipitation of less than 1000 mm. Therefore, we select Kunming, Dali, and Pu'er as representatives to further study the rainfall distribution characteristics of these three regions.

3. Time series analysis of rainfall distribution in Yunnan Province

Due to the randomness and uncertainty of rainfall data in time, a family of random variables depends on time t. ARIMA model founded by box and Jenkins [7] can effectively predict the short-term fluctuation of random time series.

Figure 2 shows that from January 2012 to December 2020, the rainfall in Kunming, Dali, and Pu'er showed a cyclical and seasonal change trend. Therefore, the ARIMA model was selected to study the structure and characteristics of rainfall time series.

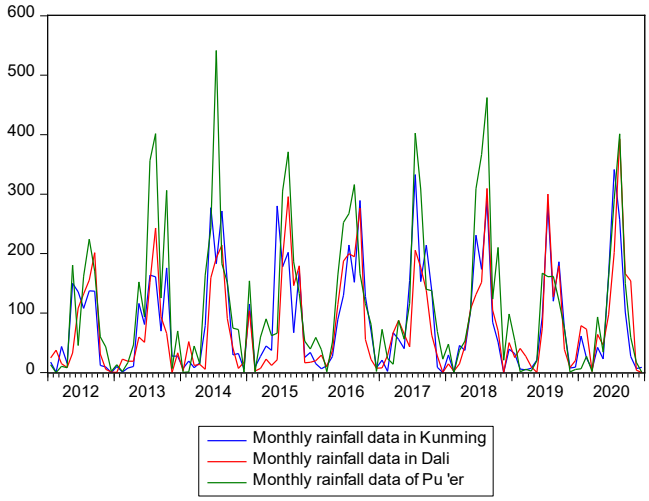

Figure 2: Sequence diagram of rainfall in Kunming, Dali, and Pu'er

3.1. Basic assumptions of ARIMA model

If a time series y_t is expressed by the linear combination of its previous value and random term,

ISSN 2616-5775 Vol. 6, Issue 1: 32-41, DOI: 10.25236/AJCIS.2023.060106

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t \tag{2}$$

The time series y_t is a p-order autoregressive model [8], which is called $AR(p)$, and the parameter $\phi_1, \phi_2, \cdots, \phi_p$ is called the autoregressive coefficient of $AR(p)$ model.

When B^k is a k-step lag operator, even if $B^k y_t = y_{t-k}$,

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \tag{3}$$

Then, Eq. (2) can be abbreviated as $\phi(B) y_t = u_t$.

If a time series y_t is represented by a linear combination of its current and previous random error terms,

$$y_t = u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \cdots - \theta_q u_{t-q} \tag{4}$$

The time series y_t is a q-order moving average model [9], which is called $MA(q)$, and the parameter $\theta_1, \theta_2, \cdots, \theta_p$ is called the moving average coefficient of $MA(q)$ model.

With a lag operator, $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ (5)

Then, Eq. (4) can be abbreviated as $y_t = \theta(B) u_t$.

If a time series y_t is expressed as a linear combination of its previous value and the random error term of the current period and the previous period. That is, it is a mixture of $AR(p)$ model.

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \cdots - \theta_q u_{t-q} \tag{6}$$

The time series y_t is a (p,q)-order autoregressive moving average model, which is called $ARMA(p,q)$.

The precondition of establishing $ARMA(p,q)$ model is that the time series is stable [10]. If a time series y_t is non-stationary and stationary after d-order phase by phase difference, the $ARMA(p,q)$ model established by the stationary series $z_t = \nabla^d y_t, t > d$ can be expressed as $ARIMA(p,d,q)$ model.

$$\phi_p(B)(1-B)^d y_t = \theta_q(B) u_t \tag{7}$$

$$\text{or } \phi_p(B) \nabla^d y_t = \theta_q(B) u_t \tag{8}$$

Because the time series data have both tendency and seasonality and have the correlation with the integer multiple of the season as the length, it is necessary to convert the d-order phase by phase difference and the D-order seasonal difference with the cycle length of s into a stable time series. For such time series, $ARIMA(p,d,q)(P,D,Q)^s$ model can be established as follows.

$$\phi_p(B) \Phi_P(B^s)(1-B)^d (1-B^s)^D y_t = \theta_q(B) \Theta_Q(B^s) u_t \tag{9}$$

$$\text{or } \phi_p(B) \Phi_P(B^s) \nabla^d \nabla_s^D y_t = \theta_q(B) \Theta_Q(B^s) u_t \tag{10}$$

where P is the autoregressive order of seasonal factors, Q is the moving average order of seasonal factors, and D is the order of seasonal difference.

ISSN 2616-5775 Vol. 6, Issue 1: 32-41, DOI: 10.25236/AJCIS.2023.060106

3.2. Stationary test of sequence

As shown in Fig. 2, the time series of rainfall in Kunming, Dali, and Pu'er show obvious periodic and seasonal fluctuations, and the autocorrelation function shows an increasing trend with the increase of lag value, which are non-stationary time series. In this study, unit root (ADF) is used to test the stationarity of time series.

Table 2: ADF test results of each sequence

variable	ADF statistics	Closeout probability	1% threshold	5% threshold	10% threshold	conclusion
km	-1.871651	0.3443	-3.500669	-2.892200	-2.583192	nonstationary
dl	-1.197141	0.6732	-3.500669	-2.892200	-2.583192	nonstationary
pe	-6.714928	0.0000	-3.494378	-2.889474	-2.581741	steady

In Table 2, the value of t statistic in the ADF test result of km sequence is -1.871651, and the value of t-statistic in the ADF test result of dl sequence is -1.197141. Given the significant level of 1%, both are greater than the critical value of t-statistics, and the original assumption that the series have unit roots is not rejected. That is, the series km and dl are non-stationary time series. In the ADF test results of pe series, the value of t- statistic is -6.714928, which is less than the critical value of the t-statistic with a given significance level of 10, 5, and 1%. Therefore, pe series has no unit root and is a stationary series.

3.3. Adjustment of nonstationary time series

Because the series of Kunming and Dali have the characteristics of trend and seasonal fluctuation at the same time, the moving average difference method is used to adjust the precipitation data of the three regions stably and seasonally while the seasonal trend of the precipitation time series of Pu'er is obvious. The difference operator $d(y,n,s)=(1-B)^n(1-B^s)y$ represents the first-order phase-by-phase difference and the first-order seasonal difference with step s for the sequence y n times.

Table 3: ADF test results of each sequence after difference

variable	difference operator	ADF statistics	Closeout probability	1% threshold	5% threshold	10% threshold	conclusion
dkm	d(km,1,12)	-3.514914	0.0099	-3.510259	-2.896346	-2.585396	steady
ddl1	d(dl,1,12)	-3.333019	0.0165	-3.511262	-2.896779	-2.585626	nonstationary
ddl2	d(dl,2,12)	-11.71334	0.0001	-3.510259	-2.896346	-2.585396	steady
dpe	d(pe,0,12)	-8.680126	0.0000	-3.500669	-2.892200	-2.583192	steady

Table 3 shows that the sequence of km sequence after one-time first-order phase by phase difference and one-time seasonal difference with step size of 12 is recorded as dkm. According to the ADF test results, dkm is a stable time series. However, the time series $ddl1$ after the first-order phase by phase difference and the first-order seasonal difference with a step size of 12 for dl series are still non-stationary time series, so it is necessary to difference the data again. The dl sequence $ddl2$ is a stationary time series after a 2-order phase by phase difference and a 12 step seasonal difference. Because the pe sequence is a stable time series, there is no need to eliminate the stationarity of the sequence, only the seasonal difference with a step size of 12 is needed, and the obtained sequence dpe is a stable time series.

3.4. Time series analysis results of rainfall distribution in Yunnan Province

According to the ADF test results of the difference series of rainfall data in Kunming, the $ARIMA(p,d,q)(P,D,Q)^s$ time series model needs to be established. Firstly, the autocorrelation analysis diagram and partial autocorrelation analysis diagram of dkm sequence are drawn to determine the specific values of p and q in $ARIMA(p,d,q)(P,D,Q)^s$ model.

ISSN 2616-5775 Vol. 6, Issue 1: 32-41, DOI: 10.25236/AJCIS.2023.060106

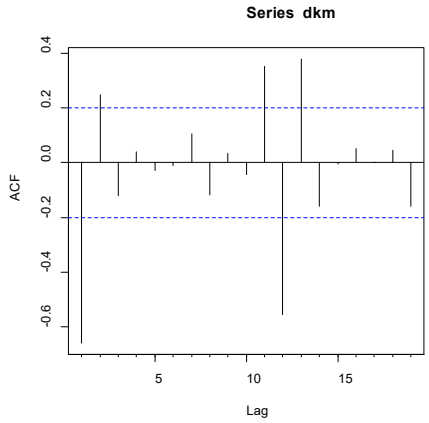

Figure 3: Autocorrelation diagram of Precipitation Series in Kunming

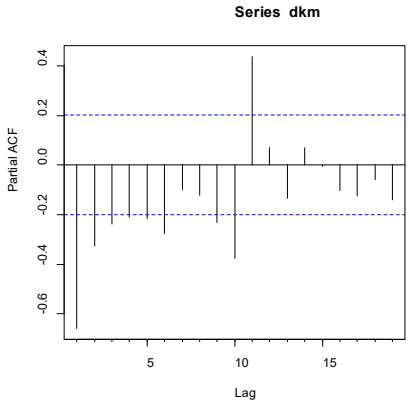

Figure 4: Partial autocorrelation of Precipitation Series in Kunming

According to the autocorrelation diagram and partial autocorrelation diagram of Precipitation Series in Kunming(Fig. 3 Fig. 4), in model (9), $p=2$ or $p=3$, $q=3$ or $p=4$. After the first-order phase-by-phase difference, the trend of the series is eliminated, so $d=1$. When k=12, the autocorrelation function ACF and partial autocorrelation function PACF of the sequence are significantly not equal to zero, so $P=Q=1$, after the first-order seasonal difference, the seasonality of the sequence is eliminated, so $D=1$. Next, time series models with different p,q combinations are established. The test results of the model are as follows.

Table 4: Test results of precipitation series models in Kunming

variable	AdjustedR2	AIC	SC	p-Q	MAPE
$ARIMA(2,1,3)(1,1,1)^{12}$	0.757986	11.11345	11.32851	0.812	3.34
$ARIMA(2,1,4)(1,1,1)^{12}$	0.7755591	11.05327	11.29522	0.911	2.37
$ARIMA(3,1,3)(1,1,1)^{12}$	0.773358	11.06168	11.30363	0.975	1.24
$ARIMA(3,1,4)(1,1,1)^{12}$	0.773596	11.07141	11.34023	0.899	1.66

Table 4 shows that the above four models meet the stationarity and reversibility conditions of $ARIMA$ model. The reciprocal roots of each order of lag polynomial are within the unit element, so the model setting is reasonable. The adjusted decision coefficients R2 are all above 0.75, and the accompanying probability (P-Q) of the white noise test of the residual sequence is greater than 0.8. This indicates that the residual sequence of each model meets the assumption of independence, and the model fitting is good. The MAPE value of the trial prediction is less than 5, and the prediction accuracy of the model is high. By comprehensively comparing the AIC and SC values of each model, $ARIMA(3,1,3)(1,1,1)^{12}$ model is more appropriate.

The fitting results of $ARIMA(3,1,3)(1,1,1)^{12}$ model are as follows.

ISSN 2616-5775 Vol. 6, Issue 1: 32-41, DOI: 10.25236/AJCIS.2023.060106

$$(1-0.315B^{12})(1+0.212B-0.652B^2-0.308B^3)(1-B)(1-B^{12})$$
$$=(1-1.312B+1.26B^2-0.83B^3)(1-0.499B^{12})u_t \quad (11)$$

where B^k is the k -step lag operator and u_t is the random error term.

According to the ADF test results of the difference series of precipitation data in Dali, a time series model in the form of $ARIMA(p,d,q)(P,D,Q)^s$ is established. Firstly, the autocorrelation analysis diagram and partial autocorrelation analysis diagram of dkm sequence are drawn to determine the specific values of p and q in $ARIMA(p,d,q)(P,D,Q)^s$ model.

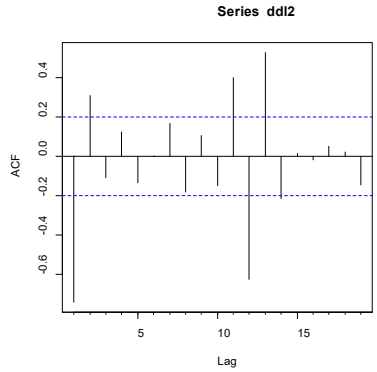

Figure 5: Autocorrelation diagram of Precipitation Series in Dali.

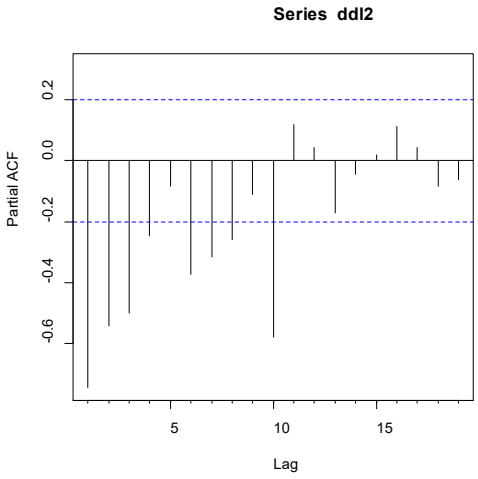

Figure 6: Partial autocorrelation of precipitation series in Dali.

According to the autocorrelation diagram and partial autocorrelation diagram of precipitation series in Dali Prefecture(Fig.5 Fig.6), $p=1$, $q=1$ or $p=2$ in model (9). When k=12, the autocorrelation function ACF and partial autocorrelation function PACF of the sample are significantly non-zero, so $P=Q=1$. Next, time series models with different p,q combinations are established. The test results of the model are as follows.

Table 5: Test results of precipitation series models in Dali

variable	AdjustedR2	AIC	SC	p-Q	MAPE
1,1	0.941571	11.11205	11.24733	0.816	3.07
1,2	0.954038	10.86165	11.02399	0.866	2.85
0,2	0.954465	10.84094	10.97622	0.927	2.41

Table 5 presents that the above three models meet the stationarity and reversibility conditions of $ARIMA$ model, and the model setting is reasonable. The adjusted decision coefficients R2 are all above 0.90, and the accompanying probability (P-Q) of the white noise test of the residual sequence is greater than 0.8. The result meets the assumption of independence, and the model fits well. The MAPE

ISSN 2616-5775 Vol. 6, Issue 1: 32-41, DOI: 10.25236/AJCIS.2023.060106

value of trial prediction is less than 5, indicating that the prediction accuracy of the model is high. The AIC and SC values of $ARIMA(0,2,2)(1,1,1)^{12}$ are smaller than those of the previous two models, so it is appropriate to choose this model.

The fitting results of $ARIMA(0,2,2)(1,1,1)^{12}$ model are as follows.

$$(1-0.406B^{12})(1-B)^2(1-B^{12})dl = (1-1.905B+0.928B^2)(1-0.469B^{12})u_t \quad (12)$$

where B^k is the k -step lag operator and u_t is the random error term.

As the precipitation time series of Pu'er is stable, it is not necessary to carry out the first-order phase-by-phase difference, and the $ARIMA(p,d,q)(P,D,Q)^s$ model is directly established by using the sequence dpe after seasonal difference.

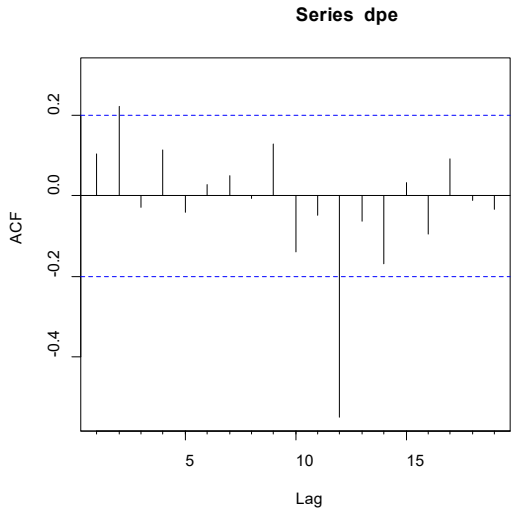

Figure 7: Autocorrelation diagram of Precipitation Series in Pu'er.

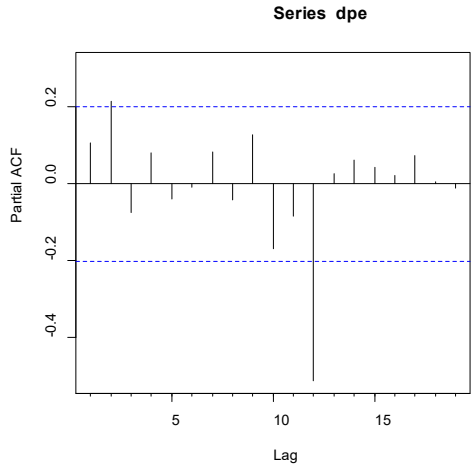

Figure 8: Partial autocorrelation of Precipitation Series in Pu'er.

According to the autocorrelation analysis diagram and partial autocorrelation analysis diagram of dkm series(Fig. 7 Fig. 8), $p=1, q=1, P=Q=1$ in $ARIMA(p,d,q)(P,D,Q)^s$ model.

Table 6: Test results of precipitation series models in Pu'er

variable	AdjustedR2	AIC	SC	p-Q	MAPE
1,1	0.542200	11.48331	11.61687	0.998	1.89

ISSN 2616-5775 Vol. 6, Issue 1: 32-41, DOI: 10.25236/AJCIS.2023.060106

Table 6 reveals that $ARMA(1,1)(1,1,1)^{12}$ model has a good fitting effect. The companion probability value of Q statistic is 0.998. The companion probability of the white noise test of residual sequence meets the assumption of independence. The test passes and the prediction accuracy is high. The fitting results are as follows.

$$(1-0.917B)(1-0.322B^{12})(1-B^{12}) = (1-0.8136B)(1-0.7277B^{12})u_t \quad (13)$$

where B^k is the k -step lag operator, and u_t is the random error term.

4. Discussion

The terrain of Yunnan Province is dominated by mountains and plateaus, which incline from northwest to Southeast, and the altitude difference is large. The characteristics of rainfall in northeast, East, West, and South Yunnan show a seasonal concentration trend with more single-point heavy rainfall, and the spatial and temporal distribution of rainfall is uneven. This leads to the frequent occurrence of regional droughts, floods, mudslides, and other disasters. By using a K-mean clustering method, we divide 16 prefectures and cities into three types of regions from the spatial dimension with rainfall as the variable. Then, the ARIMA prediction model based on big data is established by using the monthly rainfall data of Kunming, Dali, and Pu'er from 2012 to 2020. This method effectively predicts the fluctuation law of rainfall, better grasp the spatial and temporal distribution characteristics of rainfall in Yunnan Province, and provides an early warning mechanism for drought, flood, debris flow, and other disasters.

5. Conclusion

With the monthly precipitation data of Yunnan Province from 2012 to 2020 as the research object, the classification number of regions is determined. Firstly, based on the results of systematic clustering. After applying the K-mean clustering analysis method, 16 prefectures and cities in Yunnan Province are grouped into three categories according to the distribution characteristics of precipitation. A representative city is selected from each category to study the distribution of precipitation. Then, according to the fluctuation of monthly rainfall data in Kunming, Dali, and Pu'er, the ADF test method is used to test the stationarity of the data, and the first-order phase-by-phase difference is performed on the non-stationary data to eliminate the non-stationary. As the rainfall data of the three regions have a seasonal fluctuation trend, the first-order seasonal difference with a step size of 12 is used to eliminate the seasonal fluctuation. Next, the specific form of the ARIMA model is determined according to the data autocorrelation analysis diagram, partial autocorrelation analysis diagram, and the number of differences.

The research shows that the $ARIMA(3,1,3)(1,1,1)^{12}$ model is established for the precipitation data of Kunming City, the $ARIMA(0,2,2)(1,1,1)^{12}$ model is established for the precipitation data of Dali Prefecture, and the $ARMA(1,1)(1,1,1)^{12}$ model is established for the precipitation data of Pu'er. These three models have a good fitting effect, and high prediction accuracy, and the residual sequence of them is a white noise sequence.

Acknowledgment

This study is supported by a Joint special project of local colleges and Universities - Youth Project: Research on Drought Disaster Risk Influencing Factors and Loss Measurement in Yunnan Province (202001BA070001-122)

References

[1] Dong Xuyan, Lu Ying. Study on Spatial-temporal Variation of Rainfall and its Impact on Distribution Pattern of Water Resources in Yunnan Province [J]. China population • resources and environment, 2017,27 (S2): 140-144. (In Chinese)

[2] Dawoud I, Kaciranlar S . An Optimal k of kth MA-ARIMA Models Under a Class of ARIMA

ISSN 2616-5775 Vol. 6, Issue 1: 32-41, DOI: 10.25236/AJCIS.2023.060106

model[J]. Communications in Statistics, 2016, 46(12):5754-5765.
[3] Ma Lihua. Research on Time Series Modeling of Long-term Meteorological Data [D]. Kunming University of technology, 2019. (In Chinese)
[4] Xi Liping, Cai Wenqing, Wu Haiying. Research on Precipitation Prediction Model of Wuwei County Based on Time Series Analysis [J]. Journal of Anhui Vocational and Technical College of water resources and hydropower, 2018,18 (01): 50-53. (In Chinese)
[5] Wang Xihua, Lu Wenxi, Chu Haibo, Chen Sheming. Application of ARMA-GARCH Model Based on Wavelet Analysis in Precipitation Forecast [J]. Water saving irrigation, 2011 (05): 52-56. (In Chinese)
[6] Hartigan, J. A. and M.A.Wong. A K-Means Clustering Algorithm[J]. Applied Statistics, 1979, 28: 100-108.
[7] EA Robinson, Silvia M T. The Box-Jenkins Approach[J]. Nature Reviews Genetics, 1979, 3(11): 883-889.
[8] Yao Dengkui, Duan gonghao. Rainfall trend analysis and prediction of Wuhan mayor sequence based on seasonal SARIMA model [J]. Groundwater, 2022,44 (02): 166-168. (In Chinese)
[9] Chen Jiawei. Prediction and research of landslide displacement based on ARIMA model and PSO-BP neural network algorithm [D]. Three Gorges University, 2020. (In Chinese)
[10] Chen Shan. Research on water quality time series data cleaning and early warning in Qiantang River Basin [D]. Hangzhou University of Electronic Science and technology, 2022. (In Chinese)