THE COMPARISON OF TWO PRECIPITATION PREDICTION METHODS

Xuqi Mo^{1, 2}, Guiwen Lan^{1, 2*}, Yongliang Du^{1, 2}, Zixin Chen^{1, 2}

¹ Guangxi Key Laboratory of Spatial Information and Geomatics, Guilin University of Technology, Guilin, China – (760634990,

23955461, 291296038, zxchen055@163)@qq.com

² College of Geomatics and Geoinformation, Guilin University of Technology, Guilin, China

KEY WORDS: Precipitation predict; Regression analysis model; Time series model; Comparative analysis

ABSTRACT:

Precipitation forecasts play the role in flood control and drought relief. At present, the time series analysis and the linear regression analysis are two of most commonly used methods. The time series analysis is relatively simple as it only requires historical precipitation data. The model of the linear regression analysis can ensure high accuracy for causality analysis and short, medium and long-term prediction. Guilin is the region of the heavy rain center in Guangxi, which frequently suffers serious losses from rainstorms. Selecting a better model to predict precipitation has the important reference significance for improving the accuracy of precipitation weather forecast. In this research, the two methods are used to predict precipitation in Guilin. According to data of the monthly maximum precipitation, monthly average daily precipitation and monthly total precipitation from 2014 to 2016, this paper establishes the time series model and linear regression analysis model to predict precipitation in 2017 and compare the forecast results. The results show that the monthly average daily precipitation model is best with the accuracy of the time series model, and the residual error of predicted precipitation is 3.08mm, but the change trend of predicted precipitation is not accord with the actual situation. The residual error is only 0.45 mm through using inter-annual linear regression equation established by different seasons is used to predict the precipitation with residual error of 3.25 mm, and it is coincident for the predicted precipitation trend with the actual situation. Furthermore, the predictions fitting errors of spring, summer, autumn and winter are all less than 20%, which are within the scope of the specification prediction error.

1. INTRODUCTION

Heavy precipitation is a disastrous weather that occurs frequently in summer in China. This is catastrophic and easy to cause urban water logging, landslides, farmland water logging, house collapse and other disasters. The northern Guangxi region centered on Guilin city is one of the rainstorm centers in Guangxi. Therefore, heavy rain and rainstorm are one of the major disastrous weather in Guilin. Heavy rain and rainstorm not only cause losses to social economy, agricultural production and tourism, but also poses a serious threat to people's life and property safety (Tao, 1980, Write, 2012). Hence, the study of precipitation prediction play the great significance role in forecast and early warning of heavy precipitation weather in Guilin. In a large number of climate prediction problems, future weather changes are not only affected by many factors of the weather process, but also necessarily related to the periodic changes of the forecast itself (Jin et al., 2003). Many scholars have done relevant research for the analysis of precipitation prediction, Liu et al. (2017) classified the research objects and established a prediction model for the disaster year by establishing a fuzzy evaluation index for flood disasters. However, the prediction effect of this model is good only at two points behind the origin, and reduced reliability of forecasts after 2015, which is only suitable for short-term prediction. Ou et al. (2009) used rainfall data to analyze the characteristics, trends and cause of rainstorm in Guangxi. At the same time, they studied the relationship between Guangxi and South China Sea summer monsoon. In order to explore the trend of precipitation in Guangxi, Yao et al. (2013) adopted the standard index, reversed the orthogonal function of the torsion-history and Mann-Kendall method to explore the characteristics of drought and flood in Guangxi. In addition, Liu et al. (2014) established a time series ARIMA model to

analyze the time series of precipitation in Beijing from 2004 to 2015, and forecast the precipitation in Beijing from 2015 to 2020. The study shows that the model study alone does not guarantee the accuracy of the results. However, most of these scholars only use the modeling methods to research precipitation tends to forecast precipitation, while comparative analysis of time series precipitation prediction models and linear regression precipitation prediction have less literature on predicting precipitation, for precipitation changes an precipitation in Guilin. There are fewer explorations for prediction. In this paper, a time series model is established according to change period of precipitation itself. Combined with the monthly precipitation and temperature prediction factors of Guilin, a linear regression method is adopted to study the precipitation prediction, and the effects of the two prediction models are compared and analyzed. In particular, the characteristics of the torrential rain and heavy precipitation change, and understand the changing laws of precipitation, which will help improve the prediction ability of disaster prevention and mitigation.

2. RESEARCH AREA AND DATA

2.1 Research Area

Guilin is located in the northeastern part of Guangxi (25°15′23.3″N;110°19′25.5″E) and belongs to the subtropical monsoon climate zone, the climate is hot and rainy in summer, which is affected by the warm and humid air stream in southwest, tropical cyclone or denature-changing cold air mass in north China. The annual average temperature is 19.2°C, the highest of temperature appeared in June to September, and the lowest temperature occurs in January to February. The annual average relative humidity is 76%-82%, and the distribution of annual precipitation is uneven, mainly concentrated in summer half year (April-September), the annual precipitation is about 1500mm, accounting for 80% of the annual precipitation,

^{*} Corresponding author e-mail: 23955461@qq.com

rainstorm disaster are common. Autumn and winter rain less, more dry.

2.2 Data Sources

The study uses weather data provided by the Weather Company An IBM Business (https://www.wunderground.com/history/): monthly maximum precipitation data from 2014-2017 in Guilin, and monthly average precipitation and total monthly rainfall data, average daily temperature data from 2014 to 2017.The weather website observations are updated every hour, and the weather is updated more frequently in bad weather, and with high data quality.

The SPSS statistical software originated from the United States is used to predict and analyze the precipitation in Guilin, which is one of the best statistical analysis software in the world and has a wide range of applications (Cao et al., 2012). In this paper, the average daily precipitation of Guilin from 2014 to 2017was the minimum 0.2mm (in October 2014) and the maximum precipitation was 456.94mm (in March 2015).

3. ESTABLISHMENT OF TWO PRECIPITATION PREDICTIONMODELS

3.1 Establishment of the Time Series Precipitation Prediction Model

3.1.1 Time Series Precipitation Prediction Model

The expert modeler modeling in SPSSS statistical software was

selected when establishing the time series prediction model. The expert modeler automatically identifies and estimates the best fit ARIMA or exponential smoothing model for one or more dependent variable sequences without having to repeatedly verify to identify the most appropriate model. It also automatically finds the best fit model for each dependent sequence to select a statistically significant relationship model with the dependent sequence. The historical data of precipitation was stored in the Excel table, define the data of guilin in 2017.

The precipitation prediction model for the time series of maximum monthly precipitation, average daily precipitation and total precipitation of Guilin in 2014-2016 is shown in the following chart. From Table 1 shows that the smooth R-square of the monthly average daily precipitation and monthly total precipitation models in Guilin are greater than 0.5, and the effect is significant, while the monthly maximum precipitation model has a stable R-square of 0.21 with less correlation. In addition, the average daily absolute error (MAE) of the monthly average precipitation model is only 1.91, which better reflects the smaller prediction error and higher model accuracy.

Model	Number of prediction variables	Stable R square	MAE	Statistics	DF	Significance
Monthly daily maximum precipitation model	1	0.21	20.98	22.79	18	0.19
Average daily precipitation per month model	1	0.56	1.91	24.63	18	0.13
Total precipitation per daily per month model	1	0.57	57.81	25.28	18	0.11

Table 1. Statistical of Guilin precipitation model from 2014 to 2017

3.1.2 Analysis of Precipitation Change

Though the visual analysis of the single factor of monthly maximum precipitation, monthly average precipitation and monthly total precipitation in Guilin, it can be seen from Figure 1 that precipitation does not change periodically with time series from 2014 to 2015, there was no obvious peak change in precipitation, and the trend of precipitation was inconsistent with the trend of the fitted curve. It is predicted that the precipitation in 2017 will only change from the middle of March to July in the confidence interval. To predict rainfall only about the middle of March to July precipitation change in a confidence interval. Found the maximum precipitation modeling reliability is not high, and precipitation changes are not affected by a single factor. From the visual analysis of Figure 2 and Figure 3, it can be concluded that the predicted precipitation is the same as the actual precipitation, and the increase and decrease are consistent. The peak of real precipitation is about the same time. The maximum precipitation peak in three years occurs in July 2015. The peak of precipitation in 2014-2015 was in May, and heavy rains occurred in May of these two years, causing floods. The precipitation trend in 2017 is predicted to change periodically, with obvious peak and valley values of precipitation. However, the two figures cannot distinguish the accuracy effect of the precipitation prediction model, it is necessary to select the most suitable data to establish a time series model to predict precipitation.

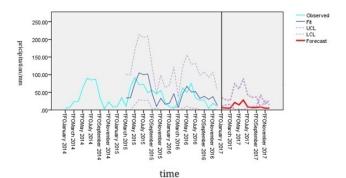


Figure 1. Figure forecast model of monthly maximum precipitation in Guilin from 2014 to 2017

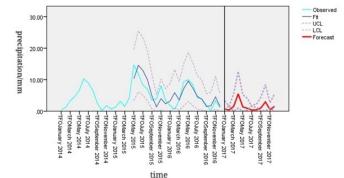


Figure 2. Figure forecast model of average daily precipitation in Guilin from 2014 to 2017

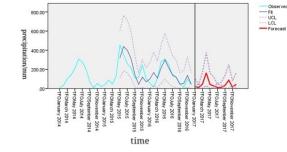


Figure 3 Figure forecast model of total precipitation per day per monthly in Guilin from 2014 to 2017

3.1.3 Precision Test

In order to reduce the prediction error, accuracy verification is required. The accuracy is tested by calculating the residual error between predicted precipitation and actual precipitation in 2017, to select a model with small residuals and within the accuracy range to predict precipitation. As shown in the chart below (unit: mm). From the above Table 2, Table 3 and Table 4 can be derived: the maximum monthly actual precipitation and predicted precipitation residual of Guilin in 2017 is 25.32mm, which does not meet the accuracy requirements; the residual error of average monthly precipitation in Guilin is 3.08mm within the accuracy range (\leq 3.5mm).However, the residual error of monthly total precipitation is 82.38mm, which is much larger than the residual accuracy range.

Month	Actual precipitation	Predict rainfall	BIAS	Monthly average BIAS
1	7.1	6.54	0.56	
2	10.9	5.47	5.43	
3	21.1	5.2	15.9	
4	55.9	21.22	34.68	
5	46	14.52	31.48	
6	146.1	28.15	117.95	25.32
7	65	8.62	56.38	25.52
8	53.1	6.79	46.31	
9	5.1	7.09	-1.99	
10	3	8.34	-5.34	
11	7.9	5.08	2.82	
12	5.1	5.41	-0.31	7

Table 2. Monthly maximum precipitation and forecast precipitation in 2017

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-3/W10, 2020 International Conference on Geomatics in the Big Data Era (ICGBD), 15–17 November 2019, Guilin, Guangxi, China

Month	Actual precipitation	Predict rainfall	BIAS	Monthly average BIAS
1	0.4	0.67	-0.27	
2	1.2	0.38	1.32	
3	4.5	1.52	2.98	
4	4	5.5	-1.5	
5	6	1.38	4.62	
6	14.1	0.92	13.28	3.08
7	8.2	0.38	7.82	5.08
8	7.4	0.41	7.09	
9	1.5	1.05	0.45	
10	1.2	2.98	-1.78	
11	3.9	0.49	3.41	
12	0.6	1.51	-0.91	

Table 3. Monthly daily average precipitation and forecast precipitation in 2017

Month	Actual precipitation	Predict rainfall	BIAS	Monthly average BIAS
1	18.53	20.29	-1.76	
2	48.01	11.28	36.73	
3	140.97	46.48	94.49	
4	118.87	163.4	-44.53	
5	185.67	43.76	141.91	
6	439.45	26.73	412.72	82.38
7	252.73	11.32	241.41	82.38
8	232.92	12.15	220.77	
9	15.99	32.81	-16.82	
10	5.09	87.74	-82.65	
11	27.42	14.91	12.51	
12	18.91	45.14	-26.23	

Table 4. Monthly daily total precipitation and forecast precipitation in 2017

3.2 Linear Regression Analysis of Precipitation Model

3.2.1 Establishment of the Linear Regression Model

Precipitation and temperature are two major elements of climate. The temperature and precipitation vary with the region. Different regions have different temperature and precipitation conditions, and the climatic characteristics are different. Therefore, this paper introduces the factor temperature that affects precipitation, and studies the change of precipitation and temperature to predict precipitation. Wang Jiawen et al. (2017) analyzed the correlation between extreme precipitation and temperature in mainland China from the perspective of time and space through statistical methods. It is found that the maximum extreme precipitation increased with temperature and was positively correlated with temperature. In this paper, the regression analysis method is used to establish inter-annual and seasonal linear regression models with precipitation as the dependent variable and temperature as the independent variables, and analyze the relationship between precipitation and temperature, to predict precipitation. According to the meteorological season division method, winter is from December to February, spring is from March to May, summer is from June to August, and autumn is from September to November.

The Pearson correlation coefficient is used to judge whether precipitation and temperature are related. The Pearson correlation coefficient is also referred the correlation coefficient. It is a measure to measure the linear correlation between two continuous variables. If the correlation coefficient is r, then:

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{((n\sum x^2) - (\sum x)^2)}\sqrt{((n\sum y^2) - (\sum y)^2)}}$$
(1)

Where r= correlation coefficient n=the number of sample x=the annual sum of temperatures y=total annual precipitation

Among them, the correlation coefficient value is between -1 and 1, the greater the absolute value of r, the closer the linear relationship. It is generally believed that the absolute value of r is greater than or equal to 0.5, and the correlation can be determined by one-dimensional linear regression analysis. In the formula (1): x means temperature as an independent variable, and y means precipitation as a dependent variable. This paper selects the monthly average daily precipitation of Guilin from 2014 to 2016 and corresponding average daily temperature of 36 samples per month. Determine the correlation coefficient between precipitation and temperature for these 36 samples and judge their relationship, where n is the number of sample, as shown in Table 5 (unit: mm)

Year	Number of samples	у	Х	ху	x^2	y^2
2014	12	47.1	237	1132.9	5295	316.55
2015	12	70.9	237	1587	5189	609.67
2016	12	48.5	240	1092.5	5418	324.13
Sum	36	166.5	714	3812.4	15902	1250.35

Table 5. Actual precipitation patterns from 2014 to 2016

It can be seen from the calculation of (1) formula that r=0.588, the correlation coefficient is within the range, and the precipitation is correlated and strongly correlated with its corresponding temperature (0.588>0.5). Therefore, to analyze the trend between precipitation and temperature through SPSS established a linear regression model. Let the linear regression model be:

$$y_i = a + bx \tag{2}$$

Where

a= constant b= a one-time coefficient.

y= predicted value

Using the monthly average daily precipitation of Guilin from 2014 to 2016 and its corresponding temperature data, an inter-annual linear regression equation was established. Explore the causes of precipitation and temperature predicting in Guilin in 2017.

From Table 6, it can be obtained that R=0.558 is equal to the Pearson correlation coefficient calculated in (1) and greater than 0.5, the precipitation is positively correlated with temperature, and the R square is equal to 0.311, indicating that

only 31.1% of the factors affecting precipitation originate from temperature, and there are other influence factors (such as wind speed and humidity, etc.).

From Table 7, it can be concluded that the output non-standardized coefficient is 0.293, the standard error is 0.075, the constant is -1.187, and the standardized error is 1.571.The comparison of standardized regression coefficients will change with geographic location and will only apply to specific geographic locations. Non-standardized regression coefficients are used for natural quantification, and actual prediction should be selected for non-standard coefficients. This paper predicts precipitation by studying the relationship between precipitation and temperature, and the geographical location will change, which is not suitable for choosing the standardized coefficient. Therefore, the non-standardized coefficient is selected. The linear equation can be obtained as follows:

$$y_i = -1.187 + 0.293x \tag{3}$$

R	R Square	Adjusted R square	Errors in standard estimates			
0.558	0.311	0.291	3.119			
Table 6. Model statistic						

	Nonstandardzed coefficient		Standard coefficient	t	Significance
	В	Standard error	Beta		
Temperature	0.293	0.075	0.558	3.920	0.000
Quantity	-1.187	1.571	-	755	0.455

From Table 7.Precipitation regression coefficient

As can be seen from Figure 4 that precipitation changes around linear or linear, the temperature is between 8°C-30°C. The precipitation increases with the increase of its corresponding temperature, and the peak value of precipitation is about 25°C. When the temperature is 8°C, 15°C, 20°C and 27°C, the precipitation is strongly correlated with temperature, and the correlation is good.

In order to reveal the seasonal variation of precipitation and temperature, a linear equation was established for the precipitation data and temperature data of Guilin from 2014 to 2016 in spring, summer, autumn and winter, and the precipitation in 2017 was predicted in different seasons. In hydrological prediction, the fitting error reaches the prediction range, and the absolute value is less than 20% of the error of the "Standard for Hydrological Information And Hydrological forecasting" (Shuilibu, 2009),within the specification license, the smaller the error, the higher the accuracy, and the prediction requirements. The linear relationship between precipitation and temperature is as follows:

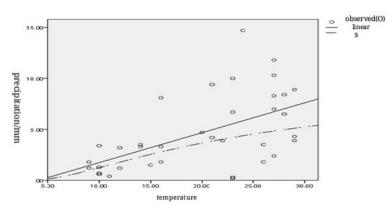


Figure 4. Figure monthly average precipitation and temperature relationship from 2014-2016 in Guilin

It can be seen from Table 8 that the correlation between spring and summer precipitation and temperature is strongly correlated, and the correlation in autumn is good, but there is no obvious correlation in winter. Guilin has low temperature and little rain in winter and plum rain in spring. It has abundant rainfall from April to September every year, and a long rainy season. However, heavy rainfall and rainstorm usually occur in the period April-September every year, which is related to Guilin's subtropical monsoon climate zone (unit: mm).

Season	Spring	Summer	Autumn	Winter
Correlation coefficient	0.786	0.654	0.514	0.165

Table 8. Relation between precipitation and temperature from 2014 to 2016

As can be seen from Table 9 that the prediction errors of precipitation prediction for spring, summer, autumn and winter in Guilin from 2014 to 2016 are respectively -8.7%, 19.6%, -15.5% and -16.3%, and the fitting errors are all less than

20%.It can be seen from Table 10 that the residual of the predicted precipitation and actual precipitation is 3.25mm, and the prediction fitting error is 8%. Within the accuracy range, all meet the prediction accuracy requirements.

3.370 and -1	.570 and -10.570, and the fitting errors are an less than meet the prediction accuracy requirements.						
Season	Regression function	Actual precipitation	Predict rainfall	Prediction fitting error%			
Spring	$Y_i = 9.45 + 0.817x$	14.5	15.77	-8.7			
Summer	$Y_i = 58.67 - 1.83x$	30.4	24.45	19.6			
Autumn	$Y_i = 7.64 - 0.223x$	7.1	8.2	-15.5			
Winter	$Y_i = -1.2 + 0.16x$	2.2	1.84	-16.3			

Table 9. Regression formula table of seasonal precipitation and temperature

Year	Actual precipitation	Predict rainfall	Residual	Prediction fitting error
2017	53.1	49.85	3.25	8%

Table 10. Statistical table of forecast errors of actual precipitation and forecast precipitation in 2017

Where

3.2.2 F-test Of Precipitation and Temperature

Due to observation and measurement errors, insufficient data, single factors such as research factors, the predicted value and the actual value will produce certain errors. Therefore, the feasibility of the research method is verified by F-test.

$$SS_{T} = SS_{R} + SS_{E}$$

$$F = \frac{MS_{R}}{MS_{E}}$$
(4)
(5)

$SS_{T=}$ sum of squares
$SS_{R=}$ the sum for regression sum of squares
$SS_{E=}$ sum of squared errors
MS_{R} =return to the mean square
MS _{E=} residual mean square

From (4), (5) and Table 11, F=15.364.Check the F value table, $F_{0.05}(1,34)$ =4.13,F> $F_{0.05}(1,34)$, indicating that the difference is statistically significant at the 0.05 level, so the linear relationship of the equation is significant, indicating that the influence of temperature and precipitation is significant, and the model is established.

	Sum of squares	DOF	Mean square	F	F _{0.05} (1,34)
Regression (R)	149.485	1	149.485	15.364	4. 13
Residual	330.803	34	9.729	-	-
Sum	480.288	35	-	-	-

Table 11. Variance analysis table of precipitation from 2014 to 2017

4. COMPARATIVE ANALYSIS OF TWO MODELS OF PRECIPITATION PREDICTION RESULTS

It can be seen from Figure 5 that the actual precipitation and the predicted precipitation increase in the first three monthly months in 2017, but in the rainy season from April to June, the predicted precipitation shows a decreasing trend with no seasonal change. Heavy rain weather in June 2017, the actual

precipitation increased sharply. However, the predicted precipitation did not increase in June, but decreased. This situation is due to the establishment of a time-predicted precipitation model to predict precipitation. Only based on historical precipitation data to predict precipitation, without considering the factors affecting precipitation, such as temperature, humidity and wind speed, the trend of predicting precipitation is not in conformity with the actual situation, it can be seen that this model is used to predict rainfall not rigorous. When selecting the research model, the precision should not only meet the requirement, but also consider the actual situation. As can be seen from Figure 6 that the linear equation is predicted regardless of the season to predict the precipitation peak in August, which delayed by two months than the actual precipitation peak, and the period change is not obvious. It is dry in autumn, but its predicted precipitation is more than spring, while have a significant impact on temperature and precipitation of the changes. It can be seen from Figure 7 that the predicted precipitation in 2017 is consistent with the actual precipitation increase and decrease. When the temperature increases, the precipitation also increases, and the precipitation also decreases when the temperature decreases. Since the rainstorm occurred in Guangxi from June to early July in 2017, the precipitation peak occurred in June. Precipitation is mainly concentrated in the season of high temperature (April-September), and the precipitation increases with temperature. After November, it begins to enter the winter. As the temperature decreased, the predicted precipitation and the actual precipitation showed a decreasing trend. Affected by subtropical monsoon climate, it is hot and rainy in summer and low in winter.

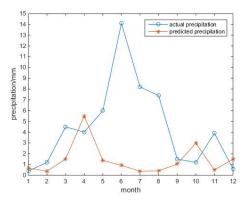


Figure 5. Figure the time series model predicts precipitation changes in 2017

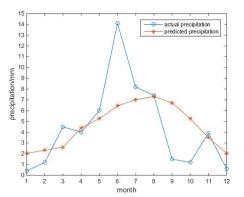


Figure 6. Figure linear regression model inter-annual prediction of precipitation changes in 2017

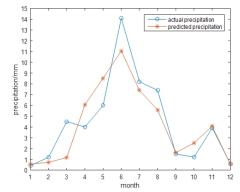


Figure 7. Figure linear regression model seasonal prediction of precipitation changes in 2017

5. CONCLUSION

This paper compares and analyzes the establishment of time series precipitation model and linear regression precipitation model to predict precipitation and draws the following conclusions:

- (1) Obtain a time series model to study the change of precipitation in time series. It is found that using monthly average daily precipitation data from 2014-2016 in Guilin, to establish a model to predict the accuracy of precipitation in 2017 is high (3.08 mm), but due to the model without considering the factors influencing the precipitation (such as temperature, humidity, etc.), predicting the trend of precipitation changes does not conform to the actual situation.
- (2) An inter-annual linear regression model was established to predict precipitation, Pearson correlation coefficient is 0.558.It was found that correlation between precipitation and temperature was significant and positively correlated. The residual error is 0.45mm, and the prediction fitting error is 10.7%, which is within the accuracy range. However, the peak value of precipitation is not obvious in summer, and the peak value is delayed by two months than the actual precipitation, which is not in line with the reality.
- (3) The predicted precipitation residual of the season linear regression prediction precipitation model is 3.25mm, and the prediction fitting error is 8%. The predicted precipitation is consistent with the actual precipitation change trend, the change period is obvious, and the precision is high, which is more practical.
- (4) The data obtained in this paper is not enough, so that the research results may be contingent.

ACKNOWLEDGEMENTS

The research work is supported by National Natural Science Foundation of China under Grant No.41861050.

REFERENCES

Tao, S.Y., 1980. The torrential rain in China. *Science Press*, Beijing.

Editorial Group of Weather Forecast Technology and Method in Guangxi. *China Meteorological Press*, Beijing, 2012, 1-145.

Jin, L., Luo, Y., Wang, Y. Y., 2003. Research on neural network hybrid prediction model for monthly precipitation. *Plateau* Meteorological, 22(6), 618-623.

Liu, H, X., Xu, Q.J., 2007. Fuzzy comprehensive evolution and prediction regional flood disaster risk. *Catastrophology*, 22(4), 38-42.

Ou, Y., 2009: Characteristics of rainstorm in flood season and its relationship with South China Sea summer monsoon in Guangxi. *Tropical Geography*, 29(1), 16-25.

Yao, S., Chen, Z.R., 2013. Analysis of the evolution of drought and collapse characteristics in Guangxi with standard precipitation index. *Journal of Sun Yat-sen University*, 52(2), 42-46.

Liu, X.H., Yang, S., Liu, Z.K., 2014. Analysis of ten years precipitation in Beijing based on ARIMA model. *China University of Mining and Technology, Beijing.*

Cao, M., Shen, L.C., Xiao, Q., 2012. Change of karst groundwater quality and its affect factors based on SPSS--Taking the surface karst groundwater in Jinfoshan, Chongqing as an example. *Journal of Natural Science of Hunan Normal University*, 35(2), 82-87.

Wang, J.W., Liao, W.L., Wang, D.G., 2017. Correlation between extreme precipitation and temperature in mainland China. *Journal of Sun Yat-sen University*, 29(6), 0529-6579.

Ministry of Water Resources, 2009. GB/T 22482-2008. *Standard for Hydrological Information And Hydrological forecasting*, Beijing, 9–10.