Modeling and Prediction of Rainfall-runoff Relationship in the Yuanjiang, Red River Basin

Xiuli Sang, Song Gao, Jianxin Xu
Quality Development Institute, Kunming University of Science and Technology, Kunming, P. R. China
Engineering Research Center of Metallurgical Energy Conservation and Emission Reduction Ministry of Education, Kunming University of Science and Technology, Kunming, P. R. China
Email: kmustsunny2008@126.com, gaosongyouxiang@126.com, xujianxina@163.com

Hua Wang
Engineering Research Center of Metallurgical Energy Conservation and Emission Reduction Ministry of Education, Kunming University of Science and Technology, Kunming, P.R. China
Email: kmustgao@163.com

Kun Zhang
Hydrology & Water Resources Bureau of Yunnan, Kunming, P.R. China
Email: zhangkun323@163.com

***Abstract*—This paper found rainfall and runoff was in a state of chaos in Poincare section and developed a method for constructing a relationship model of rainfall-runoff based on time series analysis technology. It aimed at estimating and discussing the relationship of rainfall-runoff time series by Euclidean distance (n>2). A linear model was well built to illustrate time-correlation characteristics with the correlation coefficient 0.994, which was able to predict each other by obtaining one rainfall or runoff value. Analytical results verified the accuracy of this model. It is complementary research to our previous work which discussed the special situation (n=2), and the previous work can't be applied to predicting rainfall or runoff. So it was an in-depth study and got improved.**

***Index Terms*—Distance, Relationship, Rainfall-runoff, Similarity, Time series**

I. Introduction

In *Yunnan* province, water shortage and drought have prevailed in four consecutive years. Therefore, it is vital to investigate the time-correlation characteristic between rainfall and runoff to have an effective management and utilization of scarce water resources.

Yuanjiang, Red River Basin is of great importance in Yunnan's water system, whose water pollution, changes of watercourse and hydrologic regime have attracted international attention [1]. Researches of *Yuanjiang* have been a hot topic in recent years. As we all know, the main source of water resources is Precipitation. Rainfall and runoff is correlated and consistent. The study on correlation is helpful and of great significance for Yunnan's water resource.

Previous studies [2-6] mainly concentrated on constructing the relational model which involved simulation of runoff in a given rainfall sequence. These studies illustrated that they shared the similar variability. In [7, 8], plots of standardized runoff and rainfall per basin revealed similar variability. In [9, 10], although rainfall and runoff trends were similar on the whole, the correlation coefficient of rainfall and runoff showed some fluctuations [11].

Our previous study was about the relationship model based on Euclidean distance (n=2) [12], it demonstrated time-correlated characteristics of rainfall and runoff in *Yuanjiang, Red River Basin*, and presented a detailed evaluation of the time correlation of rainfall-runoff similarity. Rainfall-runoff similarity analysis was used to determine the optimal similarity. The results showed that a time-correlated model was found to be capable of well predicting the rainfall-runoff similarity in the *Yuanjiang, Red River Basin*. Based on preliminary research had limitations, a method with Euclidean distance (n>2) for analyzing time-correlated characteristics was proposed. It was complementary research to our previous work. A linear model was well built to illustrate time-correlation characteristics with the correlation coefficient 0.994, which was able to predict each other by obtaining one rainfall or runoff value. Thus, it is comprehensive and meaningful to construct relationship model of rainfall and runoff.

Our main contributions are as follows:

1. We found the rainfall and runoff was all in a state of chaos in Poincare section.
2. We constructed a relationship model based on Euclidean distance (n>2) between rainfall and runoff. The goodness of fit of new model was higher than that of previous one.

Manuscript received September 5, 2013; revised September 21, 2013.
Corresponding author: Jianxin Xu.

doi:10.4304/jsw.9.2.437-442

3. This method could predict the other value by obtaining one rainfall or runoff value.

Your goal is to simulate the usual appearance of papers in a Journal of the Academy Publisher. We are requesting that you follow these guidelines as closely as possible.

II. Htudy Area

Yuanjiang, Red River Basin lies in mainly central and southeastern *Yunnan* as well as the southwest of *Guangxi Zhuang Autonomous Region*. Rich in water resources, Yuanjiang can provide water for field irrigation, industry and urban residents. It is making a significant contribution to local economic development. Meanwhile, it's an international river via *Vietnam* into the *South China Sea*. The amount of water flowing out reaches at 47.6×109 m^3/yr. So it's an important river [15].

The time series of daily rainfall and runoff were obtained from the hydrology stations of *Da-dong-yong* which was located in *Najian County of Yunnan Dali Prefecture*, at 100° 34 ' east longitude, 25° 04 ' north latitude. It was a typical provincial important hydrometric station in *Yuanjiang, Red River Basin*. The station was mainly responsible for collecting basic hydrological information, providing services for the protection and exploitation of water resources.

III. Modeling

A. Euclidean Distance

Euclidean distance is a common model to measure absolute distance between the respective points in the multidimensional space. The smaller the distance is, the higher the degree of similarity is. It represents the real distance between two points in the *n*-dimensional space. While *n*=2, the Euclidean distance is the straight-line distance between two points [12, 16]. Where d is the value of Euclidean distance, and x_1, y_1, x_2, y_2 represent horizontal and vertical coordinate values of two different points.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (1)$$

Euclidean distance formula (*n*>2):

Each point X can be expressed as $(X_1, X_2, \cdots\cdots, X_n)$, among which $X_i (i = 1, 2, 3, \cdots\cdots, n)$ is a real number and it refers to the *n*-th coordinate of X. The distance between $A = (x_1, x_2, \cdots, x_n)$ and $B = (y_1, y_2, \cdots, y_n)$ is recorded as $d(x, y)$, and is defined as the following equation [17].

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_i - y_i)^2} \quad (i = 1, 2, \cdots, n) \quad (2)$$

B. Data Processing

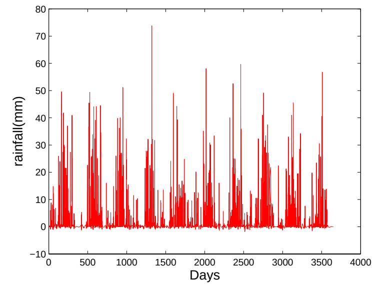
The rainfall and runoff in *Yuanjiang, Red River Basin* from 2001 to 2010 were obtained. In order to reflect real trend of hydrological data, the noise was eliminated by wavelet analysis [9, 10, 12-14]. We defined Signal-to-noise ratio (often abbreviated *SNR*) as

$$SNR = \frac{|X|}{|X - XD|} \quad (3)$$

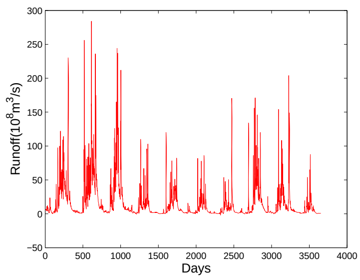
X represented the vector signal, and the XD indicated the de-noised version of input signal.

By selecting different parameters in wavelet function, a higher *SNR* was determined. The *SNR* of rainfall time series was 18.31, and the other was 64.25.

Obviously, rainfall and runoff presented periodic change [11, 18]. But runoff values showed oscillation and instability, which indicated that runoff was influenced by some factors such as vegetation coverage and evaporation capacity [18].



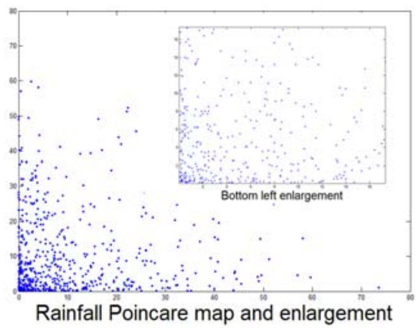
(*SNR*=18.31)



(*SNR*=64.25)

Figure 1. Time series plots of the de-noised rainfall (up) and runoff

Then this torrent of data was observed in Poincare surface of section, the trajectory of all points was shown in Figure 2.



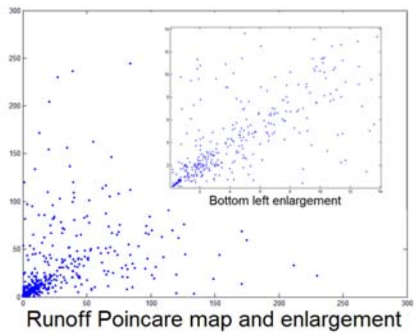


Figure 2. Poincare surface of section about rainfall (up) and runoff (down)

The movement characteristics of these points were obtained from Fig.2, points of density were patchy in Poincare section and they had obvious hierarchical structure. Therefore, rainfall and runoff was in a state of chaos.

Because it was hard to calculate with these data, and the total annual data as a unit was more convenient.. The real total annual data was shown in Table I.

TABLE I.

ANNUAL VALUES OF RAINFALL AND RUNOFF FROM 2001TO 2010

Year	rainfall/mm	runoff/（$m^3 \cdot s^{-1}$）
2001	936.4	8535.08
2002	1087	12368.97
2003	932.4	10728.29
2004	658.1	3358.76
2005	709.4	3960.07
2006	734.6	2803.88
2007	694.8	2468.60
2008	958.7	6479.11
2009	863	5646.47
2010	519	1987.35

Then these data were shown in Fig.3 in order to be more intuitive.

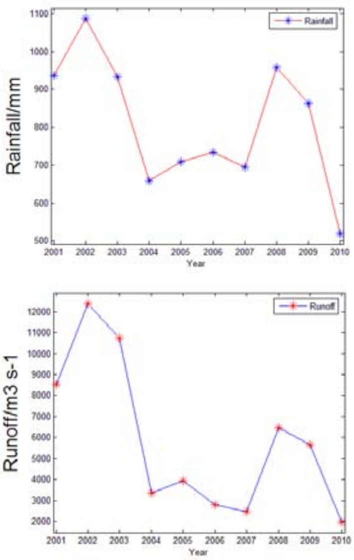


Figure 3. The annual rainfall (up) and annual runoff (down)

An interesting phenomenon was found: the annual rainfall and annual runoff demonstrated similar change trend. Besides, the correlation coefficient between rainfall and runoff was 0.906. For this reason, the time-correlated characteristics were worth studying.

C. Model Construction based on Euclidean Distance (n>2)

Although rainfall-runoff modeling involves many methods [18-22], a model based on the above methods is to be constructed in this paper. Detailed steps as follows:

Step one: a normalized data from 10 years of annual rainfall values was obtained by a linear transformation:

$$Y=20X\text{-}100 \qquad (4)$$

Step two: data calculation of the ten-year rainfall and runoff, as shown in table II.

TABLE II.

DATA PROCESSING OF RAINFALL-RUNOFF

Year	Annual rainfall x_i	Annual runoff y_i	x_i-y_i	$(x_i-y_i)^2$
2001	18628	8535.08	10092.92	101867034
2002	21640	12368.97	9271.03	85951997
2003	18548	10728.29	7819.71	61147864
2004	13062	3358.76	9703.24	94152866
2005	14088	3960.07	10127.93	102574966
2006	14592	2803.88	11788.12	138959773
2007	13796	2468.6	11327.4	128309945
2008	19074	6479.11	12594.89	158631279
2009	17160	5646.47	11513.54	132561488
2010	10280	1987.35	8292.65	68768110

Step three: calculation of the consecutive years' Euclidean distance by adopting Euclidean distance (n>2) formula. The specific process is shown as follows:

$$d_{2001}(x,y)=\sqrt{(x_1-y_1)^2}$$

$$d_{2002}(x,y)=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2}$$

…………………………

$$d_{2009}(x,y)=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\cdots+(x_9-y_9)^2}$$

Step four: the above data were superimposed and results were shown in table III.

TABLE III.

SUPPERPOSITION OF DATA OF RAINFALL-RUNOFF

The cutoff year	Euclidean distance accumulated value
2001	10092.92
2002	13704.71
2003	15778.68
2004	18523.49
2005	21111.48
2006	24179.63
2007	26701.39
2008	29522.8
2009	31688.44
2010	32755.54

The above data were illustrated in Fig.4, and the goodness of fit 0.994 was obtained by linear fit with least square method.

Four steps were calculated with MATLAB 7.0, detailed code was as follows.

```
>> A=[936.4 1087 932.4 658.1 709.4 734.6 694.8 958.7 863 519];
>> a=20*A-100;
>> b=[8535.08 12368.97 10728.29  3358.76 3960.07 2803.88 2468.6 6479.11 5646.47 1987.35];
>> C=a-b;
>> c=power(C,2);
>> d(1)=sqrt(c(1));
>> d(2)=sqrt(c(1)+c(2));
>> d(3)=sqrt(c(1)+c(2)+c(3));
>> d(4)=sqrt(c(1)+c(2)+c(3)+c(4));
>> d(5)=sqrt(c(1)+c(2)+c(3)+c(4)+c(5));
>> d(6)=sqrt(c(1)+c(2)+c(3)+c(4)+c(5)+c(6));
>>d(7)=sqrt(c(1)+c(2)+c(3)+c(4)+c(5)+c(6)+ c(7));
>>d(8)=sqrt(c(1)+c(2)+c(3)+c(4)+c(5)+c(6)+c(7)+c(8));
>>d(9)=sqrt(c(1)+c(2)+c(3)+c(4)+c(5)+c(6)+c(7)+c(8)+c(9));
>>d(10)=sqrt(c(1)+c(2)+c(3)+c(4)+c(5)+c(6)+c(7)+c(8)+c(9)+c(10));
>> x=[2001 2002 2003 2004 2005 2006 2007 2008 2009 2010];
>> y=d;
>> p=polyfit(x,y,1);
>> xx=2001:2010;
>> yy=polyval(p,xx);
>> plot(x,y,'o',xx,yy);
>> plot(x,y,'*',xx,yy);
>> legend('accumulated value','Linear fitting' );
>> xlabel('Cutoff Year');
>> ylabel('Euclidean distance accumulated value');
>> sprintf('Linear fitting coefficient: \n(%g,%g,%g)+ (%g,%g,%g)t\n',p)
```

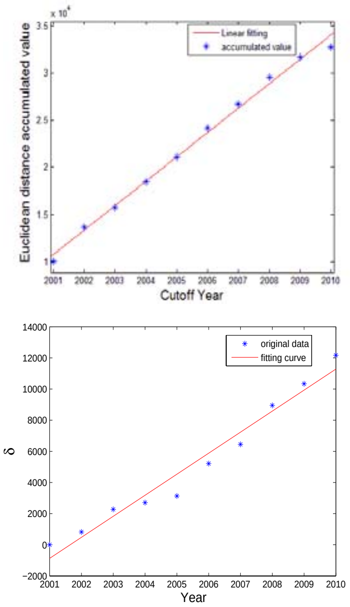


Figure 4. The new model (up) and previous model(down) with the correlation coefficient 0.994 and 0.982.

A model of rainfall-runoff similarity in *Yuanjiang, Red River Basin* was determined. The formula was as follows.

$$y = a \times x - b \tag{5}$$

Then parameters' sensitivities on the results were taken into consideration. In case of data (9 years), where a=2689, b=5370100; While in data of 8 years, where a=2724, b=5440000. So, the parameters of the model showed some fluctuations clearly.

D. Discussion about Euclidean Distance ($n \geq 2$)

It's believed that the similarity of rainfall and runoff still exist in other regions [2-10]. Although this paper focused on *Yuanjiang, Red River Basin*, this method could be applied in different fields. The parameters would have a difference accordingly.

This accuracy of model could be verified by using the data of 2010. The Euclidean distance value of y with above formula was 34701, while the correct Euclidean distance value of 2010 was 32755.54. However, an accumulated error from 2001 to 2010 was obtained, and the annual error was 194.55. The relative error rate was acquired by the following formula. Where ω is the relative error rate, λ predicted value and θ actual value [15].

$$\omega = \frac{\lambda - \theta}{\theta} \tag{6}$$

After calculation, the error rate was 0.59%.

This method could predict rainfall and runoff as well.

$$d_n(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}$$
$$= \sqrt{d_{n-1}^2(x, y) + (x_n - y_n)^2}$$
$$|x_n - y_n| = \sqrt{d_n^2(x, y) - d_{n-1}^2(x, y)} \tag{7}$$

If only one rainfall or runoff value was obtained, other data could be derived with the help of formula (7). $|x_{2010} - y_{2010}|$=13393.7. The value of y_{2010} was 1987.35, and we can get $x_{2010} = 15381.05$. The value of rainfall in 2010 was 774.05, while the true value was 519. Because of an accumulated error from 12 months, the monthly error rate was 4.1% after calculation through (6).

In our previous work, the relationship model of rainfall-runoff was based on Euclidean distance (n=2), and the goodness of fit was 0.982. In addition, an interesting and stable trend was found in table IV. It is seen that corresponding constraint condition was a range or interval in even-numbered years, and for the odd-numbered years, the corresponding constraint condition was just the boundary value of range. This regularity was discovered. But it was still difficult to predict. Therefore, the range of constraint condition in 2010 wasn't obtained accurately [12]. On the premise of the 2010 rainfall value, the runoff value was to be calculated with the model. Suppose runoff value of 2010 was $b+s$. $Min\Phi_{2010}$= $|Q|$+ $|Q+c_1|$+……+$|Q+c_9|$+$|Q+c_{10}|$, $c_{10} = -(8348 + s)$ (s>-83535.08) (c_i the constant, Q the difference of a and b, Φ objective function). Although $Min\Phi_{2010}$ was 11307 by the model, Q and s were variable. So unique runoff value wasn't acquired, and the result differed in different conditions. Based on previous work, this latest research made the goodness of fit improved by Euclidean distance (n>2). It can better measure relationship between rainfall and runoff. In addition, it can predict rainfall or runoff value very accurate, which is a breakthrough.

TABLE IV.

THE CORRESPONDING CONSTRAINTS OF SIMILARITY MODEL IN 10 YEARS.

2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
0	[0,821]	821	[389,821]	389	[0,389]	0	[-35,0]	-35	[-35,0]

IV. CONCLUSION

In this paper, it was found rainfall and runoff was in a state of chaos in Poincare section and a relationship model of rainfall-runoff similarity analysis was presented. Firstly, the normalized rainfall and runoff values were obtained by a simple linear transformation. Then the time-correlated characteristics between rainfall and runoff were explored. A model for illustrating rainfall-runoff relationship on Euclidean distance (n>2) was built. Original time series of *Yuanjiang, Red River* were used in this method. New data verified this new model. Results showed that the relationship between rainfall and runoff was of better time-correlated characteristics with a higher correlation coefficient. Besides, this method can well predict the other value through obtaining one rainfall or runoff value. The previous research which based on Euclidean distance (n=2) could not be applied to predicting rainfall or runoff, and the goodness of fit was 0.982. This latest research was a complementary and in-depth study. In this new model, the goodness of fit reached 0.994, which was better than the previous model. Therefore, this research was breakthrough for the research of time-correlated characteristics on rainfall-runoff.

Studying on time-correlated characteristics between rainfall and runoff in *Yuanjiang, Red River Basin* was of great significance. According to the research results, the relationship of rainfall-runoff always followed a linear distribution, and the next annual relationship of rainfall-runoff could be predicted. In particular, the relationship reflected the impacts of external environmental changes on hydrology and water resources system. In addition, another value can be obtained by one rainfall or runoff. It is conducive to protecting, exploiting and utilizing water resources, and providing reference to controlling floods or droughts.

The rainfall-runoff similarity also existed in other regions. Although this paper investigated *Yuanjiang, Red River Basin*, this method also could be applicable to the calculation of rainfall-runoff time-correlated characteristics of other regions. The model parameters were different accordingly. In a word, this research is of great significance.

ACKNOWLEDGMENT

This research is supported by the National Natural Science Foundation of China (Grant No. 51174105), and the project of Flood Control and Drought Relief Countermeasure in the *Yuanjiang-Red River Basin* under rainfall-runoff collaborative mechanism. We also deeply appreciate the helpful review comments and suggestions by anonymous reviewers.

APPENDIX A LIST OF SYMBOLS

d	Euclidean distance	n	year(2001-2010)
x_i	annual rainfall value	ω	error rate
θ	actual value	λ	predicted value
Y	normalized data	y_i	annual runoff

REFERENCES

[1] D. M. He, S. H. Wu, H. Peng, "A study of ecosystem changes in Longitudinal Range—Gorge Region and transboundary eco-security in southwest China", *Advances in Earth Science*, vol.20, no.3, 2005, pp.338-344.
[2] M. P. Rajurkar, U. C. Kothyari, U. C. Chaube, "Modeling of the daily rainfall-runoff relationship with artificial neural network", *Journal of Hydrology*, vol.285, 2004, pp. 96–113.
[3] F. H. S. Chiew, D. G. C. Kirono, D. M. Kent, et al, "Comparison of runoff modelled using rainfall from different downscaling methods for historical and future climates", *Journal of Hydrology*, vol.387, 2010, pp.10-23.
[4] D. Labat, R. Ababou, A. Mangin, "Rainfall–runoff relations for karstic springs. Part I: convolution andspectral analyses", *Journal of Hydrology*, vol. 238, pp.123-148, 2000.
[5] W. A. Welderufael, P. A. L. Le Roux , M. Hensley, "Quantifying rainfall–runoff relationships on the Dera Calcic Fluvic Regosol ecotope in Ethiopia", *Agricultural water management*, vol.95, no.11, 2008, pp.1223-1232.
[6] O. Kisi, J. Shiri, M. Tombul, "Modeling rainfall-runoff process using soft computing techniques", *Computers & Geosciences*, vol.51, 2013, pp.108-117.
[7] R. Makungo, J. O. Odiyo, J. G. Ndiritu, B. Mwaka, "Rainfall–runoff modelling approach for ungauged catchments: A case study of Nzhelele River sub-quaternary catchment", *Physics and Chemistry of the Earth*, vol.35, no.13, 2010, pp. 596-607.
[8] W. S. Lavado Casimiro, J. Ronchail, D. Labat, J. C. Espinoza, J. L. Guyot, "Basin-scale analysis of rainfall and runoff in Peru (1969–2004): Pacific, Titicaca and Amazonas drainages", *Hydrological Sciences Journal*, vol.57, no.4, 2012, pp.625-642.
[9] J. C Zhang, J. Jiang, D. Liu, D. L. DeAngelis, "Vegetation coverage influence on rainfall-runoff relation based on wavelet analysis", *Journal of American Science*, vol.5, no.2, 2009, pp.97-104.
[10] T. Partal, M. Küçük, "Long-term trend analysis using discrete wavelet components of annual precipitations measurements in Marmara region (Turkey)", *Physics and Chemistry of the Earth, Parts A/B/C*, vol.31, no.18, 2006, pp.1189-1200.
[11] W.C. Boughton, "Effect of data length on rainfallerunoff modeling", *Environmental Modelling & Software*, vol.22, no.3, 2007, pp.406-413.
[12] X. L. Sang, J. X. Xu, K. Zhang, H. Wang, "Analysis and Modeling of Time-Correlated Characteristics of Rainfall-

Runoff Similarity in the Upstream Red River Basin", *Advances in Meteorology*, vol.2012, 2012, pp.1-7.
[13] H. Zhang, F. Yu, X. Wang, "Condition Monitoring of Rope-less Elevator Braking System Based on Wavelet Denoising", *Journal of Computers*, vol.8, no.3, 2013, pp. 741-748.
[14] X. Chen, Z. Gao, "Data Processing Based on Wavelet Analysis in Structure Health Monitoring System", *Journal of Computers*, vol.6, no.12, 2011, pp.2686-2691.
[15] Y. Li, D. He, C. Ye, "The variation of runoff of Red River basin in Yunnan", *ACTA GEOGRAPHICA SINICA-CHINESE EDITION-*, vol.63, no.1, 2008, pp.41-49.
[16] H. Guo, J. Zhao, "Research on feature extraction for character recognition of NaXi pictograph", *Journal of Computers*, vol.6, no.5, 2011, pp.947-954.
[17] H. Kido, "Classification of isosceles 7-point 3-distance sets in 3-dimensional Euclidean space", *European Journal of Combinatorics*, vol.28, no.3, 2007, pp.685-704.
[18] Lloyd H.C. Chua, Tommy S.W. Wong, "Improving event-based rainfall–runoff modeling using a combined artificial neural network–kinematic wave approach", *Journal of Hydrology*, vol.390, 2010, pp. 92-107.
[19] X. D. Lin, Y. L. Zhang, Z. J. Yao, et al, "The trend on runoff variations in the Lhasa River Basin", *Journal of Geographical Sciences*, vol.18, no.1, 2008, pp.95-106.
[20] J. Z. Li, P. Feng, "Runoff variations in the Luanhe River Basin during 1956-2002", *Journal of Geographical Sciences*, vol.17, no.3, 2007, pp.339-350.
[21] G. F. Lin, C.M. Wang, "A nonlinear rainfall–runoff model embedded with an automated calibration method – Part 1: The model", *Journal of Hydrology*, vol.341, no.3, 2007, pp.186-195.
[22] G. J. Cao, J. Wang, L. J. Wang, Y. Y. Li, "Characteristics and runoff volume of the Yangtze River paleo-valley at Nanjing reach in the Last Glacial Maximum", *Journal of Geographical Sciences*, vol.20, no.3, 2010, pp.431-440.

Xiuli Sang She is a doctor candidate at Quality Development Institute, Kunming University of Science and Technology. She was born in 1980. Her main research interests include data mining, quality engineering and management.

Song Gao He is a master, who now works at Quality Development Institute, Kunming University of Science and Technology. He was born in 1989. His main research interests include data mining, quality engineering and management.

Jianxin Xu He is a doctor at Quality Development Institute, Kunming University of Science and Technology. He was born in 1983. His main research interests include data mining and applied mathematical model.

Hua Wang He is a Ph.D. supervisor at Engineering Research Center of Metallurgical Energy Conservation and Emission Reduction Ministry of Education, Kunming University of Science and Technology. He was born in 1965. His main research interests include energy-saving and emission reduction, applied mathematical model.

Kun Zhang He is a master, who now works at Hydrology & Water Resources Bureau of Yunnan. His main research interests include hydrology and water resource.