

# Identify of Runoff Influence Factors in Coal Mining Area Based on Stepwise Regression Analysis

Wu Xi-jun<sup>a</sup>, Dong Ying<sup>b</sup>

Department of Architecture Engineering, Yulin University, Yulin, China

<sup>a</sup>Wxj0826@163.com, <sup>b</sup>Dying1010@163.com

**Keywords:** Runoff; Stepwise Regression Analysis; influence factors; coal mining.

**Abstract.** The Northern Shaanxi province has severe water shortages, especially in coal mining areas, and it is very important to identify the influence factors of runoff for the similar area environmental protection. This paper take the Kuye River as an example, firstly introduces the principle and calculating process of Stepwise Regression Analysis; then analyze the runoff influence factors for 1961-2010 by this method. Results show that, the main factors affect the Kuye River runoff is disparate in different periods, in 1961-2010 the main factors are rainfall, temperature, soil and water conservation measures, coal mining. This research results can serve as an important reference for the ecological restoration of coal mining area.

## Introduction

In practical problems, people always hope to select some variables that affect the dependent variable as independent variables, and apply the multiple regression method to establish the "optimal" regression equation to predict or control the dependent variable. The so-called "optimal" regression equation, mainly refers to the regression equation to include all independent variables that have significant effect on the dependent variables [1-2]. Stepwise regression analysis is a regression analysis based on this principle. Multiple linear stepwise regression analysis has been widely used in many environmental fields such as environmental prediction, environmental monitoring and environmental assessment [3].

Kuye River basin is the main component of Northern Shaanxi Energy and Chemical Base, water resources are extremely scarce [4]. Some domestic scholars have studied the response relationship of the Kuye River runoff, and considered it has been affected by the coal mining in recent years [5-6]. But to what extent the influence of coal mining on river runoff, and how it relates to other influencing factors, this study also is less. In this paper, stepwise regression analysis will be used to analyze and calculate the influence degree and mutual relations of each factor on the Kuye River runoff.

## Study Area and Research Methods

### Study Area.

Kuye River is a tributary of Yellow River in the middle reaches, it originated in Ordos city of Inner Mongolia province, into the Yellow River in Shenmu County of Shaanxi province. Kuye River total length of 242 km, the basin area of 8706 km<sup>2</sup>, 4499 km<sup>2</sup> belongs to Shenmu County, the average runoff into Yellow River is 5.11×10<sup>8</sup> m<sup>3</sup>. The basin near Muus Desert, and above Shenmu is mainly sandy area, downstream is loess hilly region, it is one of the major tributaries of the Yellow River sediment transport [7]. Kuye River basin is rich in high quality coal resources, Shenfu-Dongsheng coalfield is located in the basin, and the amount of coal mined each year is more than 10000×10<sup>4</sup> t after 2010.

### Research Methods and Data Sources.

After preliminary analysis [8], there is a strong correlation between the rainfall, air temperature, water for production and living, water and soil conservation area, coal mining and runoff in the Kuye River Basin. Although the correlation between the factors and river runoff is different, it must be linear if

expressed one relationship. Therefore, the stepwise regression analysis can be used to solve the runoff equations. This study used SPSS software to analysis and calculation.

the multiple linear regression model:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_i X_i \quad i=1, 2, 3, \dots, k; \quad (1)$$

$\alpha_i$  for partial regression coefficient of each independent variable;

$X_i$  for different independent variables;

The basic idea of stepwise regression is that variable can enter and go out. The specific method is that all the independent variables  $X_i$  are enter into the regression equation one by one according to their influence on the dependent variable  $Y$ , and when each independent variable is entered, make the statistical tests for selected variables one by one, that is T test, only retain the significant independent variables [9]. Enter a variable or removing a variable from the regression equation is a step of stepwise regression, also do the significance test at each step, which is the F-test. This process is repeated until no significant independent variables are entered into the regression equation, and no insignificant independent variable removed from the regression equation [10].

F test is to test the whole fitting equation generated by each step, T test is to test the partial correlation coefficient of each variable, so it can be said that stepwise regression analysis has a double test process. In order to retain more variables, the significance level  $P$  of the F-test and T-test are both 0.10, if it less than 0.10 is significant, the equation satisfies the significance requirement and the independent variables can stay in the equation. Step by step, F-test is performed first, if it is passed, the T-test results of each variable in the equation are checked until the P-value is less than 0.10, finally the optimal regression equation is obtained [11].

## **Stepwise Regression Analysis on Kuye River**

### **Model**

Based on the previous data accumulation and research of the author [12], the main influencing factors for the runoff in the Kuye River are rainfall, temperature, area of water and soil conservation, water consumption and coal mining. The flow data are from the "Yellow River Basin Hydrology Yearbook" published by the Yellow River Committee. The rainfall and temperature data come from China Meteorological Science Data Sharing Service Network. Other data come from Shaanxi Water Resources Statistical Data and Shenmu County Statistical Yearbook. Kuye River runoff equation is set to the following formula (1), to calculate the runoff fitting equation.

$$y = \alpha_0 + \alpha_1 R + \alpha_2 T + \alpha_3 U + \alpha_4 W + \alpha_5 M \quad (1)$$

In the formula,  $\alpha_i$  is the partial regression coefficient of each variable;

R is the rainfall of the Kuye River Basin in each year, mm;

T is the temperature and of the Kuye River Basin in each year, °C;

U is the production and living water consumption of each year in Shenmu County,  $10^4 \text{ m}^3$ ;

W is the water and soil conservation area in each year in Shenmu County,  $\text{km}^2$ ;

M is the amount of coal mined in each year in Shenmu County,  $10^4 \text{ t}$

### **The main influencing factors of Kuye River runoff**

The main influencing factors of Kuye River runoff is different during different periods. In order to make a overall analysis in the past 50 years, we need to conduct a stepwise regression analysis of 1961~2010.

Regression analysis result by SPSS software can be obtained in Table 1, carried out four steps, and entered into four independent variables, respectively is precipitation R, temperature T, the area of water and soil conservation W and coal mining M, the final equation is:

$$y = 80999.99 + 146.21R - 7888.77T - 11.73W - 1.34M, r^2 = 0.81 \quad (2)$$

Table.1 Stepwise regression computation sheet in 1961~2010

Modle		Partial Correlation Coefficient	Standard Coefficient	t	P <sub>1</sub>	P <sub>2</sub>	r <sup>2</sup>
1	α <sub>0</sub>	267174.03		8.46	0.00	0.00	0.50
	T	-25357.48	-0.70	-6.87	0.00		
2	α <sub>0</sub>	188364.87		6.64	0.00	0.00	0.70
	T	-22315.80	-0.62	-7.60	0.00		
	R	127.00	0.46	5.62	0.00		
3	α <sub>0</sub>	97432.85		3.15	0.00	0.00	0.79
	T	-9681.68	-0.27	-2.63	0.01		
	R	148.64	0.54	7.63	0.00		
	W	-14.47	-0.46	-4.59	0.00		
4	α <sub>0</sub>	80999.99		2.61	0.01	0.00	0.81
	T	-7888.77	-0.22	-2.14	0.04		
	R	146.21	0.53	7.73	0.00		
	W	-11.73	-0.38	-3.50	0.00		
	M	-1.34	-0.18	-2.00	0.05		

The T-test was performed on each variable, where P<sub>1</sub> values for variables R, T, W and M were all less than 0.10, and P<sub>1</sub> for U was greater than 0.10, so the water consumption U was discarded. After the test of equation F, the P<sub>2</sub> value is far less than 0.01, so the equation is significant. The weights of variables R, T, W and M respectively are 40%, -17%, -29% and -14%, the natural factors R and T account for more than half, the human factors water and soil conservation W weight is twice the coal mining. In these factors, only the rainfall R is positive for the river runoff, the all rest is negative. There are other factors that also have a impact on river runoff, but these four are the most significant ones.

The measured flow of Kuye River from 1961 to 2010 was compared with the simulation value of the runoff equation formula (2). The result is shown in Fig.1. After calculation, the relative error Re between the simulated value and measured value is 0.08, and the determination coefficient R<sup>2</sup> is 0.84. The overall simulation results of the runoff equation is good, especially between 1960 and 1998, but the 1999 to 2010 is general, this is mainly because the amount of coal mining in the Kuye River Basin has rapidly risen during this period.

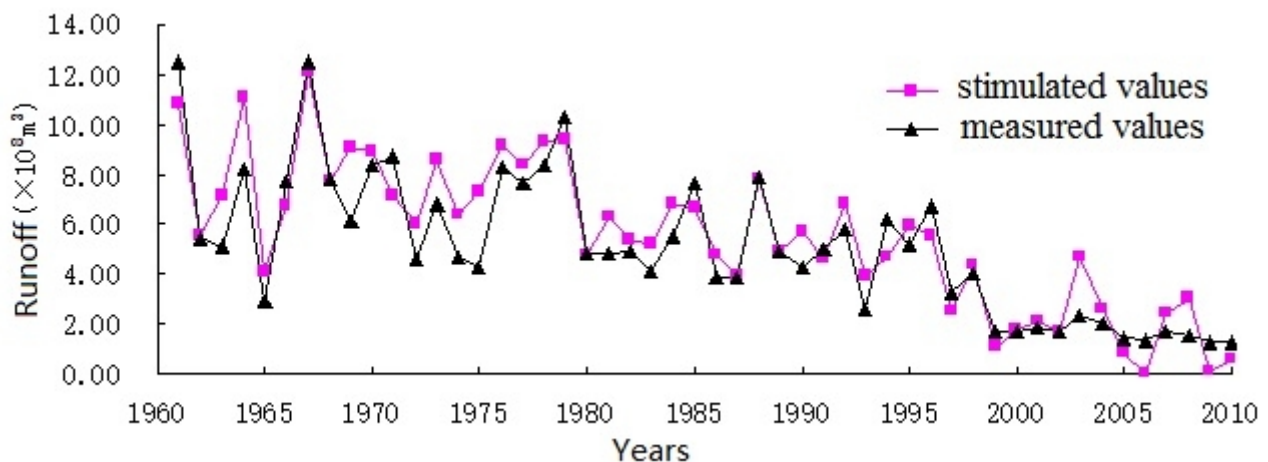


Fig.1 The stimulated values and measured values of Kuye River runoff

**Conclusions**

The main influencing factors during the 50 years from 1961 to 2010 were rainfall, temperature, water and soil conservation measures, coal mining. Rainfall as the main influencing factor has existed for a long time. The large-scale water and soil conservation measures in mid-term and the significant

temperature rise caused the decrease of runoff, The rapid increase of coal exploitation in the later period also had a great impact on the runoff.

### Acknowledgements

This work was financially supported by the Natural Science Basic Research Plan in Shaanxi Province of China (No. 2014JM2-5063) and PhD Scientific Research Funds of Yulin University (16GK07).

### References

- [1] C. Ye, C. H. Li, Q. G. Wang, X. G. Chen. Driving forces analysis for ecosystem health status of littoral zone with dikes: a case study of Lake Taihu. *ActaEcologica Sinica* vol. Vol, 32( 12), (2012), p. 3681-3690, (in Chinese)
- [2] R. L. Wilby, T. M. L. Wigley. Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography*, vol. 21 ( 1997), p. 530-548.
- [3] J. Zhang, M. T. Ma, J. P. Wang, Narration on Regression Analysis Methods Applied in Environmental Fields, *Environmental Science and Technology*, vol. 21( extra 2) , (2008), p. 40-43, (in Chinese)
- [4] Q. Z. Xing, Z. Y. Ren, L. X. Wang, Evaluation on Sustainable Utilization of Water Resources in North Shaanxi Province Based on Ecological Footprint Model. *Arid Zone Reserch*, vol. 26(6), (2009), p. 793-798, (in Chinese)
- [5] X. H. Jiang, X. W. Gu, H. M. He. The Influence of Coal Mining on Water Resource in the Kuye River Basin, *Journal of Natural Resouces*. Vol. 25(2), (2010), p. 300-307, (in Chinese)
- [6] S. F. Zhang, C. Ma, L. Zhang. The regulators of runoff of the Ulan Moron River in Daliuta mine area: The effects of mine coal. *Acta Scientiae Circumstantiae*, vol. 31(4), (2011) p. 889-896, (in Chinese)
- [7] X. K. Zhao, S. J. Wang, X. L. Fan, Analysis on the change trend of runoff and influence factors in Kuye River Basin from 1954 to 1993. *Journal of Water Resources & Water Engineering*, vol. 21(5), (2010), p. 32-36, (in Chinese)
- [8] X. J. Wu, H. E. Li, Y. Dong. the Analysis and identification of Kuye River runoff change characteristics in Northern Shaanxi. *Journal of Ecology and Rural Environment*, vol. 32(4), ( 2016), p. 558-562, (in Chinese)
- [9] H. X. Zhang, J. X. Chang, L. P. Zhang, Future runoff variation of Beiluo River Watershed, *Journal of Northwest A&F University (Nat, Sci, Ed)*, vol. 44(10), (2016), p. 207-213, (in Chinese)
- [10] Q. J. Hu, Y. Zhang, J. W. Li, Influencing factors of runoff in Shule River during the past 50 years, *Journal of Arid Land Resources and Environment*, vol, 25(10), (2011), p.143-148, (in Chinese)
- [11] Sailor D J, Li X, A semiempirical downscaling approach for predicting regional temperature impacts associated with climatic change, *Journal of Climate*, vol. 12, (1999), p. 103-114
- [12] X. J. Wu, H. E. Li, Y. Dong. Quantitative identification of coal mining and other human activities on river runoff in Northern Shaanxi region, *Acta Scientiae Circumstantiae*, vol. 34(3) (2014), p. 772-780, (in Chinese)