Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

ISSN: 0972-6268	Vol. 14	No. 3	pp. 699-702	2015

Original Research Paper

Application of Regression Analytical Method in Dynamic Prediction of River Water Quality

Jianbo Pan, Gao Zhou and Dedong Liu*
North China University of Water Resources and Electric Power, Zhengzhou 450045, China
*Shandong Survey and Design Institute of Water Conservancy, Shandong 250013, China

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com
Received: 25-3-2014
Accepted: 22-5-2014

Key Words:
Regression Model
River water quality
Prediction

ABSTRACT

It is very important to accurately predict the river water quality. Prediction of river water quality has been closely watched in water resources evaluation, and is the primary work of scientific planning and management of water resources and exploitation. The accuracy of prediction will directly influence whether we can work out a reasonable plan and management measures. According to the relationship of river water quality with the influencing factors, regression method is used to predict the tendency of river water quality. The influencing factors include rainfall, sediment and runoff. This study could provide reference and guidance for further exploitation of river water.

INTRODUCTION

Quality prediction is to build the corresponding relationship between water quality indicators and land pollution in water pollution control unit. There are many methods to predict water quality at present, which include certain mathematical models and stochastic statistical methods, such as limited unit method, finite difference method, regression analysis, wave analysis, time series analysis, probability method of average and so on (Dao 2004, He 2009). Every method has its own characters, because there are many factors affecting the water quality in river, and the factors have nonlinear relations with water quality. Water environment system has the characteristics of openness, complexity and uncertainty, and deterministic mechanism model has certain limitations in applying to water quality prediction. River water quality prediction is the basis of environmental planning, assessment and management (Fu 1985, Gao 1999). It has important practical significance to promote the sustainable use of groundwater resources efficiently. Regression analysis method is a kind of mathematical statistics method to deal with relationship among variables. It cannot only provide mathematical expressions among variables, but also can use the probability statistical knowledge for analysing to judge its effectiveness (Wang 1997, Chen 2002, Yang 2007). Using the relationship, one or more values can be used to forecast and control the value of the dependent variable, which can further know that at what extent the forecast and control can be made. Regression analysis method is based on statistical regression concept, using a variety of regression methods to establish forecasting equation including linear, multiple linear and nonlinear, etc. The authors used multiple linear regression method to analyse the river water quality with a high prediction accuracy.

REGRESSION ANALYSIS MODEL

Linear regression is to explain the dependent variable changes using a major influencing factors as independent variables. In practice, the change of variables is often affected by a number of important factors, then you need to interpret the changes of independent variables according to the variable factors with two or more, which is also known as multiple regression return. When it is a linear relationship between multiple independent variables and the dependent variable, the regression analysis is performed diversity regression.

If y is the dependent variable, $x_1, x_2, \cdots\cdots, x_k$ are independent variables. And it is the linear relationship between the independent variables and the dependent variable, the multiple linear regression model can be expressed as follows:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k + e$$

Where, b_0 is constant, $b_1, b_2, \cdots b_k$ are regression coefficients, b_1 is $x_2, x_3, \cdots x_k$, the effect of x_1 with one unit increase on y, which is also called partial regression coefficient of x_1 to y; the effect of x_2 with one unit increase on y, which is also called partial regression coefficient of x_2 to y. If the relationship of two independent variables x_1, x_2 with

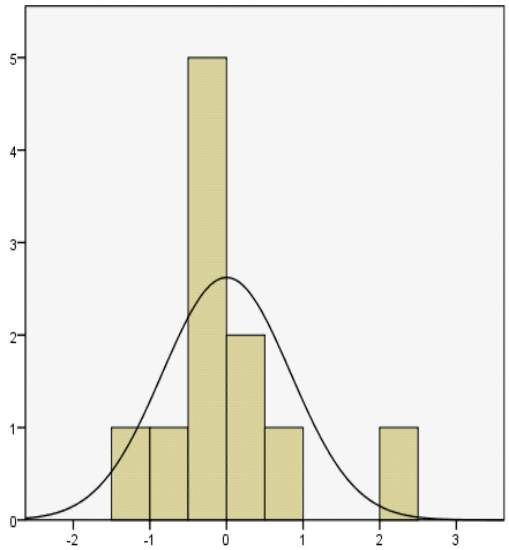

Fig. 1: Standarized regression residuals histogram.

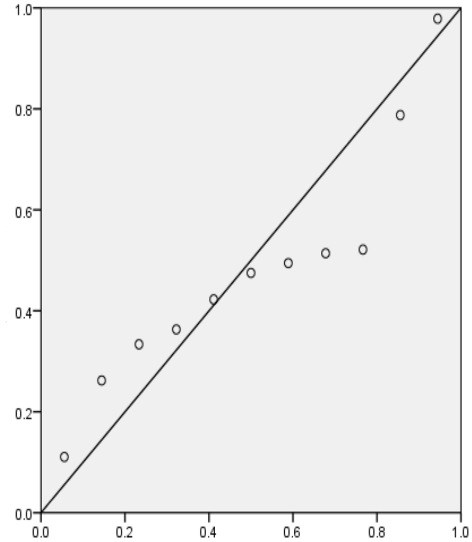

Fig. 2: Normal P-P plot of regression standardized residual.

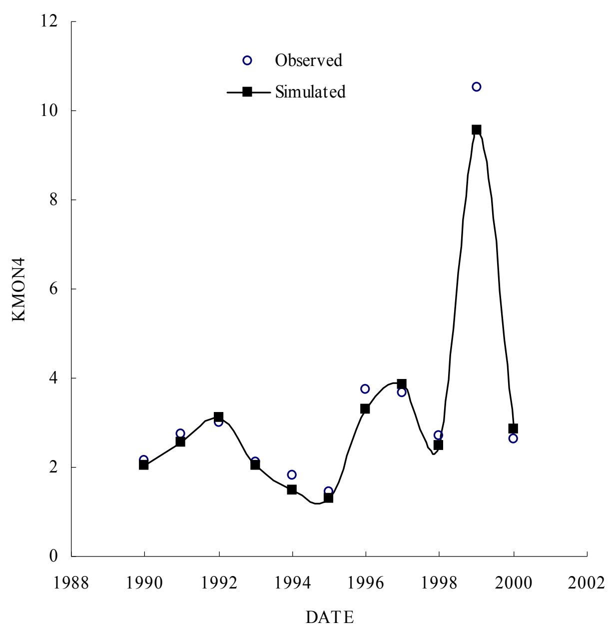

Fig. 3: Comparison of observation with simulation.

dependent variable is linear, the binary linear regression model is described as:

$$y = b_0 + b_1x_1 + b_2x_2 + e$$

When creating diversity regression model, in order to ensure the regression model has an excellent ability to explain and predict, we should first pay attention to the choice of the independent variables, the criteria are:

1. Independent variable must have a significant impact on the dependent variable, and was closely associated with linear relationship.
2. Linear correlation between independent variables and the dependent variable must be true, rather than formal.

Table 1: Descriptive statistics of evaluated water quality with its influencing parameters.

	Mean	Standard deviation	N
$KMnO_4$	3.3157	2.49546	11
Rainfall	576.4727	155.28048	11
Sediment	1.4918	1.69377	11
Runoff	6.2989	4.55913	11

Table 2: Correlation of $KMnO_4$ with its influencing parameters.

		$KMnO_4$	Rainfall	Sediment	Runoff
Pearson Correlation	$KMnO_4$	1.000	-0.397	-0.289	-0.067
	Rainfall	-0.397	1.000	0.711	-0.284
	Sediment	-0.289	0.711	1.000	-0.197
	Runoff	-0.067	-0.284	-0.197	1.000
Significance	$KMnO_4$	.	0.113	0.194	0.423
	Rainfall	0.113	.	0.007	0.199
	Sediment	0.194	0.007	.	0.280
	Runoff	0.423	0.199	0.280	.
N	$KMnO_4$	11	11	11	11
	Rainfall	11	11	11	11
	Sediment	11	11	11	11
	Runoff	11	11	11	11

Table 3: Coefficient[a] of $KMnO_4$ with its influencing parameters.

Model	Non-standardized coefficients		Standard Coefficient	t	Sig.	Confidence interval of B in 95.0%		Correlation			Collinearity statistics	
	B	Standard error	Trial Version			Lower Limit	Upper limit	Zero-order	Partial	Section	Tolerance	VIF
(constant)	8.124	4.440		1.830	.110	-2.375	18.623					
Rainfall	-.007	.008	-.443	-.899	.399	-.026	.012	-.397	-.322	-.305	.474	2.11
1												
Sediment	-.018	.711	-.012	-.026	.980	-1.700	1.663	-.289	-.010	-.009	.495	2.020
Runoff	-.107	.194	-.195	-.551	.599	-.565	.352	-.067	-.204	-.187	.920	1.088

a. The dependent variable: $KMnO_4$

Table 4: Correlation coefficient[a] of $KMnO_4$ with its influencing parameters.

MODEL			Runoff	Sediment	Rainfall
1	Correlation	Runoff	1.000	-.006	.208
		Sediment	-.006	1.000	-.696
		Rainfall	.208	-.696	1.000
	Covariance	Runoff	.038	-.001	.000
		Sediment	-.001	.506	-.004
		Rainfall	.000	-.004	6.287E-005

a. The dependent variable: $KMnO_4$

3. Since independent variables should have some mutually exclusive, i.e., the degree of correlation between the independent variables and dependent variables should not be higher than those of the dependent variable degree of correlation.

4. Independent variable should have complete statistical data, and its predictive value is easy to be determined.

Parameter estimation of regression model is similar to linear regression equation. Least squares method is used to solve parameters under the conditions of minimum squared error $\sum e^2$. The standard regression parameters for solving equations are as following.

$$\sum y = nb_0 + b_1\sum x_1 + b_2\sum x_2$$
$$\sum x_1 y = b_0\sum x_1 + b_1\sum x_1^2 + b_2\sum x_1 x_2$$
$$\sum x_2 y = b_0\sum x_2 + b_1\sum x_1 x_2 + b_2\sum x_2^2$$

Solving this equation, the parameters can be obtained as follows.

$$b == (x'x)^{-1} \cdot (x'y)$$

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 \\ \sum x_2 & \sum x_1 x_2 & x_2^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y \\ \sum x_1 y \\ \sum x_2 y \end{bmatrix}$$

APPLICATION AND CONCLUSIONS

The model was applied in Kuangmenkou station, which is located in the southwest of Hebei province. The monitoring site covered the entire city, which eventually led to multifarious difficulties in water quality evaluation. The descriptive statistics of water quality index can be seen in Table 1. In Table 1, mean, standard deviation and N are given. Table 2 is the correlation of evaluation parameters. Seen from Table 2, Pearson correlations of $KMNO_4$ with rainfall, sediment and runoff are 0.397, 0.289 and 0.067. Table 3 is Coefficient of $KMNO_4$ with its influencing parameters. Table 4 is correlation coefficient of $KMNO_4$ with its influencing parameters. Fig. 1 is regression standardized residual, which is used to judge whether the standardized residuals are subject to normal distribution. Fig. 2 is normal P-P plot of regression standardized residual, which indicated that the standardized residuals are subject to normal distribution. Fig. 3 is the comparison of observation with simulation.

Using regression analysis model to forecast the river water quality has high precision and great practical value. There are many factors influencing river water quality. So the multiple linear regression is the effective method to predict the river water quality.

Seen from the results, it is reliable that the three factors of rainfall, runoff and sediment can be seen as the main affecting factors.

REFERENCES

Chen Dongjing, Ma Anqing, Xu Zhongmin and Cheng Guodong 2002. The application of factor analysis in water quality evaluation. Hydrology, 22(3): 29-31.

Dao Ligang, Tang Ya and Wang An 2004. The application of factor analysis method in water quality assessment of Jiang'an river, Sichuan. Environment, 23(2): 75-77.

Fu Guowei 1985. Water Pollution Control and Planning. Tsinghua University Press, Beijing.

Gao Tingyao 1999. Water Pollution Control Engineering. Higher Education Press, Beijing.

He Chenggang, Feng Yan and Wang Jieping 2009. Application of factor analysis method to the water quality evaluation of water source protection area. Yunan Geographic Environmrnt Research, 21(1): 99-103.

Wang Xudong and Shao Huihe 1997. The algorithm and its application for location of Steiner points of sewerage pipeline network. Information and Control, 26(4): 272-276.

Yang Wei, Lu Wenxi, Li Ping and Yang Zhongping 2007. Application of factor analysis method to the water quality evaluation of Yitong River. Research of Soil and Water Conservation, 14(1): 113-114.