Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

ISSN: 0972-6268	Vol. 14	No. 3	pp. 655-661	2015

Original Research Paper

Correlation Analysis and Forecasting Changes in Yongding River Water Quality Based on Information Entropy and Gray System Theory

Men Baohui†, Long Rishang, Zhao Yawei, Wang Anze, Hu Sha and Wu Shuaijin
Renewable Energy Institute, North China Electric Power University, Beijing, China
†Corresponding author: Men Baohui

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com
Received: 12-8-2014
Accepted: 16-10-2014

Key Words:
Yongding river
Water environment
Grey system GM(1,1) model
Information entropy

ABSTRACT

The Yongding River is the mother river of Beijing. However, due to the environmental pollution caused by the economic development, the water and coastal environment of this river has suffered from great destruction. The ecological restoration of the Yongding River is imperative. In this paper, we analysed seven basic water quality indicators in Yongding River based on Information Entropy and found that the main factors for affecting water quality were ammonia and CODMn. Then the basic water quality indicators were predicted, based on Grey System GM(1,1) model and we concluded that turbidity and conductivity would grow fastest in the next 20 years. Finally, we made some reasonable ideas and methods in Yongding River ecological restoration.

INTRODUCTION

Yongding River, the largest river in Haihe River area, has a total basin area of 47,000 km^2 (Long 2013). Its drainage area is approximately 3200km^2 in Beijing. During the Beijing section, it flows through five areas including Mentougou, Shijingshan, Fengtai, Daxing and Fangshan. Yongding River is the mother of Beijing since ancient times, giving birth to the Beijing rich cultural heritage and unique cultural resources (Miao 2012).

In recent decades, with many years of drying up, riverbed desertification, overgrown with weeds and environmental degradation, Yongding river has become dry landfills and the largest "sandstorm source" in Jingxi. Beijing has repeatedly wanted governance, but it failed to achieve as result of lack of water. In 2010, Beijing launched the Yongding River green corridor plan and brought hope for the Yongding river. And a new opportunity was provided due to Beijing Garden Expo. Water environment effect depends on the changes in ecological restoration of water quality situation, therefore, the water quality analysis is one of the basic methods of evaluating changes in water quality situation.

Water quality evaluation is an important aspect of environmental quality assessment and there is a large number of domestic and foreign researches. The common methods for water quality assessment include: fuzzy comprehensive evaluation method, the composite index method, gray clustering method, the maximum entropy method, etc. Currently, the information entropy effect on water quality evaluation model is better. Zhao et al. (2012) believed that the comprehensive evaluation model of water quality fuzzy maximum entropy principle could overcome the shortcomings of the mean, while the calculation was simple, highly reliable, which was an ideal model for water quality assessment (He & Chen 2001). Studies showed that the calculations resolution, sensitivity and reliability, and evaluation of information in entropy model had greatly improved. Han et al. (2011) were using the maximum entropy principle to study the risk of water shortage rate risk, vulnerability, recoverability and risk of the return period.

Water quality prediction is another important aspect of water quality evaluation. The main current methods are neural networks, time series, regression analysis and the gray model method. In recent years, gray prediction method is widely used in the field of water quality prediction. Wang et al. (2002) pointed out that the applicability of the gray model to predict the water quality was good and the predictions were consistent with environmental conditions in the study of DAWU resources. And in the study of agricultural irrigation water forecast analysis, Shen (2007) also pointed out that the gray model required less data, moderate calculation and prediction results agreed well with the local conditions.

In order to evaluate and predict the Yongding River water quality, our group collected water samples from four sections: from the beginning of the Marco Polo Bridge (the interval of sections is 100m) and then we conducted experi-

Fig. 1: Water sampling locations.

ments, measured and obtained data about Yongding river's seven basic indicators of water quality. The water quality indicators included pH, turbidity, hardness, CODMn, CODCr, conductivity and ammonia. On the basis of entropy law on information, entropy theory and gray GM (1,1) model, we did a correlation analysis and forecasting changes in water quality, and thus provide support data on the Yongding governance.

RESEARCH METHODS

Entropy law: Water environment research is an important job in environmental assessment. Currently, most of the water environment evaluation is comprehensive water quality evaluation, which is a fundamental and important work to carry out the investigation and management of water environment. However, the water level primarily reflects a comprehensive water quality of the water environment of each index, which can not be carried out at the water pollution control for one or more of the excessive factors, bringing pollution blindness, uncertainty, thereby increasing the cost of pollution. This paper finds a way to find one or several factors about more serious effects of water pollution in water environmental assessment process so that it can be targeted in the future governance and reduce pollution control costs.

To determine the factors for the plurality of water pollution "contribution", it is necessary to find a quantization factor for all the standard treatments. Visual comparisons can be performed to obtain the relationship between each other and the big right is the major impact for water pollution. Currently, the ways to handle this problem are mainly AHP and information entropy method. AHP involves mutual impact between a certain level of factors, and it is conducted by human evaluation factors, with a subjective evaluation method which is not suitable for quantitative analysis. This paper empowerment is based on improved information entropy method through max-min normalization process.

Originally entropy, a concept in physics, is used to study the relationship between the thermodynamic conversion of energy. In 1948, American mathematician Claude Elwood Shannon introduced the concept of entropy to information theory, a measure used to solve the problem and quantify information. Entropy is a method to determine the information amount of each member in the system, which is defined as the formula (Huang 2012).

$$H(x) = -C\sum_{i}^{n} p(x_i)\, lgp(x_i) \quad ...(1)$$

Wherein, x_i represents the adjustment coefficient of the system, which means a system of independent random events, $p(x_i)$ denoting the probability of this event.

The size of the entropy of each factor reflects the degree of order of information. When entropy is large, it means that the degree of order of the factor is small but uncertainty is large. Conversely, when the entropy is small, it means that the degree of order of a factor is great but uncertainty is small. Depending on the size of the uncertainty factor, it can reflect the difference to the overall impact and determine the size of the relative weights.

From the above principle, this method can be applied to the evaluation of water environment. However, the differences between factors are not only on the magnitude but also on the dimension, which makes the evaluation inconvenient. Therefore, the data should be normalized. There are many methods for data normalization, such as "max-min normalize", "Z-score normalize", etc. The Yongding River water quality is influenced by season and human factors. At a specific time, there may be a specific location error in original data. In the Z-score normalize, the large data will affect the results, while the max-min normalize is the easiest way to eliminate dimensionless. The principle of the max-min normalize is that the attributes values are processed differently in the different type of property. Information entropy consists of mainly four attributes, namely, efficiency, cost-based, fixed, interval. In this paper, pH is the type of fixed attribute, hardness (mg/L), turbidity (NTU), conductivity (μS/cm), CODCr (mg/L), CODMn (mg/L) and ammonia (mg/L) are the type of cost-based attributes.

For the fixed type (pH), using formula (2) to standardize, and it is not necessary to standardize for the cost-based type.

$$b`_{ij} = |b_{ij} - 7| \quad ...(2)$$

$$x_{ij} = \frac{b`_{ij} - min_i}{max_i - min_i} \quad ...(3)$$

(i = 1,2,3……m; j = 1,2,3……n)

After treatment, according to the standardized indicators of water quality, we get standardized matrix:

$$X = (x_{ij})_{m\times n} \quad ...(4)$$

Where, x_{ij} is dimensionless, and $x_{ij} \in [0,1]_0$ is calculated by the normalized matrix (4):

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}} \quad ...(5)$$

By the formula (5), to obtain the probability distribution of the matrix:

$$P = (p_{ij})_{m\times n} \quad ...(6)$$

(i = 1,2,3……m; j = 1,2,3……n)

According to the probability distribution matrix (6), using the formula (7) to calculate the information entropy of each water quality indicators:

$$H_j = -C\sum_{i=1}^{m} p(x_i)\,lgp(x_i) \quad ...(7)$$

Where, the adjustment coefficient C = 1/ln m; then the formula (7) is converted to:

$$H_j = -\frac{1}{\ln m}\sum_{i=1}^{n} p(x_i)\,lgp(x_i) \quad ...(8)$$

Then, calculate the coefficient of variation of the evaluation indicators:

$$F_j = 1 - H_j \quad ...(9)$$

Where, $H_j \in [0,1]$, i = 1, 2, 3......m.

The larger the difference in the coefficient, the more important indicators will be and the greater impact on water quality. Express formula (9) in weight:

$$w = \frac{F_j}{\sum_{j=1}^{n} F_j} \quad ...(10)$$

Formula (10) represents that the larger weight factor is, the larger the index on the whole. Such entropy calculation does not involve any subjective factors; the result respects the objective laws, and has good manoeuvrability.

Gray model law: Gray prediction use the GM model to estimate and predict developed behavioural characteristics of the system, the model can also estimate the time, an exception occurs, as well as make the case studies in the distribution of events in the next period of time to and so on. It essentially regard the "random process" as a "gray process", "random variable" as "gray variables." In this paper, we use gray system GM (1,1) model to make forecast for each basic indicators of water quality trends.

First performed level ratio test. Based on water quality indicators time sequence:

$$x^0 = (x^0(1), x^0(2), x^0(3), \ldots, x^0(24)) \quad ...(11)$$

Calculate the level ratio $\lambda(k)$

$$\lambda(k) = \frac{x^0(k-1)}{x^0(k)} \quad ...(12)$$

If all of the $\lambda(k) \in [0.9,1.1]$, $k = 2,3,\ldots,n$, we can use x^o to be the proper GM (1,1) model. For formula (11) a cumulative time (AGO) of the sequence

$$x^1 = (x^1(1), x^1(2), x^1(3), \ldots, x^1(n))$$
$$= (x^1(1), x^1(1) + x^0(2), \ldots, x^1(n-1) + x^0(n)) \quad ...(13)$$

Make the data matrix B and data vector Y

$$B = \begin{bmatrix} -1/2\,(x^1(1) + x^1(2)) & 1 \\ -1/2\,(x^1(2) + x^1(3)) & 1 \\ \vdots & \vdots \\ -1/2\,(x^1(n-1) + x^1(n)) & 1 \end{bmatrix} \quad ...(14)$$

$$Y = \begin{bmatrix} x^0(2) \\ x^0(3) \\ \vdots \\ x^0(n) \end{bmatrix}$$

Calculate

$$u' = (a, b)^T = (B^TB)^{-1}B^TY \quad ...(15)$$

then get *a*, *b*.

Modelling

Table 1: The time and exact location of water sampling.

Sampling time	Location
April 20	Upstream of the reservoir
May 3	Reservoir downstream
May 11	Marco Polo Bridge west Jiecun
May 25	Green riparian zone 1
June 8	Green riparian zone w
June 22	Green Bank Park

Table 2: The basic indicators of Yongding water quality test results summary.

No.	1	2	3	4	5	6	7	8
Date	4.20	4.27	5.4	5.11	5.18	5.25	6.1	6.8
Section	I	II	III	IV	I	II	III	IV
pH	8.2	8.25	8.23	8.19	8.11	8.21	8.16	8.17
Hardness (mg/L)	201.18	204.5	195.7	207.9	234	219.96	226.02	207.85
Turbidity (NTU)	18.7	21.2	26.6	15.5	19.2	20.4	25.6	20.5
Conductivity (μs/cm)	812	815	817	807	831	823	811	809
CODCr (mg/L)	22.4	40.32	34.72	41.44	17.92	36.96	28	15.68
CODMn (mg/L)	5.7	5.54	5.6	5.44	5.16	5.22	5.36	5.3
Ammonia (mg/L)	0.547	0.66	0.633	0.67	0.493	0.568	0.522	0.56
No.	9	10	11	12	13	14	15	16
Date	6.15	6.22	6.29	7.6	7.13	7.20	8.3	8.10
Section	I	II	III	IV	I	II	III	IV
pH	8.62	8.55	8.56	8.5	8.57	8.44	8.45	8.52
Hardness (mg/L)	231.96	304.91	203.72	207.85	201.7	219.85	177.5	227.9
Turbidity (NTU)	20.7	25.9	21.3	20.1	23.7	23.3	21	25.8
Conductivity (μs/cm)	832	815	830	829	836	820	831	846
CODCr (mg/L)	42.56	50.4	39.2	56	31.36	24.64	35.84	36.96
CODMn (mg/L)	6.94	6.72	6.98	7.04	5.62	5.72	5.64	5.66
Ammonia (mg/L)	0.777	0.769	0.799	0.788	0.513	0.478	0.545	0.532
No.	17	18	19	20	21	22	23	24
Date	8.17	8.24	8.31	9.7	9.14	9.21	9.28	10.5
Section	I	II	III	IV	I	II	III	IV
pH	8.42	8.45	8.45	8.45	8.42	8.41	8.41	8.42
Hardness (mg/L)	203.72	252.13	238.01	201.7	217.84	221.87	226.02	205.73
Turbidity (NTU)	19.3	20.4	27.8	20.3	25.1	22.5	24.4	23.4
Conductivity (μs/cm)	823	835	830	828	819	823	821	819
CODCr (mg/L)	33.6	34.72	41.44	31.36	36.96	29.2	29.12	25.76
CODMn (mg/L)	5.8	5.84	5.75	5.74	6.02	6.13	6.1	6.17
Ammonia (mg/L)	0.432	0.443	0.448	0.437	0.445	0.437	0.46	

$$\frac{dx^1}{dt} + ax^1 = b \qquad \text{...(16)}$$

Solving (16):

$$x^1(k+1) = \left(x^0(1) - \frac{b}{a}\right)e^{-ak} + \frac{b}{a} \qquad \text{...(17)}$$

Calculate the generating column values $x^1(k+1)$ and model reduction value $x^0(k+1)$.

So $k = 1,2,3,\ldots,n$. From the above time response function it can be considered $\hat{x}^1$, make $\hat{x}^1(1) = x^0(1)$. From $\hat{x}^0(k) = \hat{x}^1(k) - \hat{x}^1(k-1)$, so $k = 2,3,4,\ldots,n$, we can get $\hat{x}^0 = [\hat{x}^0(1), \hat{x}^0(2), \ldots, \hat{x}^0(24)]$.

Then after testing the accuracy of the model, it can be used to forecasts and forecast in level of I.

CASE STUDY

Data: From April 20 to June 22, 2013, about two months time, the research team went to different locations of the Yongding River to collect water samples and water quality monitoring. Specific water sampling locations are shown in Fig. 1, the sampling time is detailed in Table 1. Based on the collected water samples, the water quality indicators were measured (Table 2).

The results: According to entropy model, using equation (4), calculated normalized matrix is given in Table 3.

The water quality index probability distribution matrix is given in Table 4.

Then we get the entropy of each index, coefficient of variation and the impact of the size of weight , as depicted in Table 5. According to the Gray model, for the pH, for example, the results obtained by the calculation are:

$$u' = (a,b)^T = (B^TB)^{-1}B^TY = \begin{pmatrix} -0.0014 \\ 8.2462 \end{pmatrix}$$

a= -0.0014, b=8.2462

$$x^1(k+1) = \left(x^0(1) - \frac{b}{a}\right)e^{-ak} + \frac{b}{a} = 5997.72e^{0.00137676\,k} - 5989.52$$

Model checking: Check the pH data in aspect of residuals, relative error and level deviation (exponential rate difference value). The test results are given in Table 6. Gray model checking is classified in Table 7 (Li 2006).

Usually taken relative error indicators prevail. Upon examination, the accuracy of the model I, can be used to predict and forecast.

Repeat the above steps for the other six basic indicators of gray prediction and detection accuracy (All are I) to obtain

Table 3: Normalized matrix.

pH	Hardness	Turbidity	Conductivity	CODCr	CODMn	Ammonia
0.1392	0.0000	0.0000	0.0000	0.4532	0.1867	0.5462
0.0000	0.5645	0.2761	0.2805	0.0000	0.0000	0.2790
1.0000	1.0000	0.4478	0.6707	1.0000	1.0000	1.0000
0.8418	0.1270	0.8806	1.0000	0.3399	0.2410	0.2243
0.7089	0.6200	0.4328	0.7927	0.4784	0.3148	0.0000
0.6392	0.4468	1.0000	0.3780	0.2527	0.5090	0.0335
3.3291	2.7583	3.0373	3.1220	2.5243	2.2515	2.0830

Table 4: Probability distribution matrix.

pH	Hardness	Turbidity	Conductivity	CODCr	CODMn	Ammonia
0.05	0	0	0	0.1796	0.0829	0.2622
0	0.2046	0.0909	0.0898	0	0	0.1339
0.3004	0.3625	0.1474	0.2148	0.3962	0.4441	0.4801
0.2529	0.06	0.2899	0.3203	0.1347	0.107	0.1077
0.2129	0.2248	0.1425	0.26	0.1895	0.1398	0
0.192	0.162	0.3292	0.1211	0.1001	0.2261	0.0161

Table 5: Entropy value of each index, coefficient of variation and the impact of weight.

Item	pH	Hardness	Turbidity	Conductivity	CODCr	CODMn	Ammonia
Entropy/Hj	0.3648	0.3616	0.3642	0.3678	0.3613	0.3435	0.31
The coefficient of variation/Fj	0.6352	0.6384	0.6358	0.6322	0.6387	0.6565	0.69
Weight/wj	0.1403	0.141	0.1405	0.1397	0.1411	0.145	0.1524

Table 6: pH gray model test results summary.

No.	Original value	Model value	Residuals	Relative error	Level deviation(exponential rate difference value)
1	8.2	8.2000	0	0	-
2	8.25	8.2631	-0.0131	0.0016	0.0047
3	8.23	8.2745	-0.0445	0.0054	-0.0038
4	8.19	8.2859	-0.0959	0.0117	-0.0063
5	8.11	8.2973	-0.1873	0.0231	-0.0113
6	8.21	8.3088	-0.0988	0.0120	0.0108
7	8.16	8.3202	-0.1602	0.0196	-0.0075
8	8.17	8.3317	-0.1617	0.0198	-0.0002
9	8.62	8.3431	0.2769	0.0321	0.0509
10	8.55	8.3546	0.1954	0.0228	-0.0096
11	8.56	8.3662	0.1938	0.0226	-0.0002
12	8.5	8.3777	0.1223	0.0144	-0.0084
13	8.57	8.3892	0.1808	0.0211	0.0068
14	8.44	8.4008	0.0392	0.0046	-0.0168
15	8.45	8.4124	0.0376	0.0045	-0.0002
16	8.52	8.4239	0.0961	0.0113	0.0068
17	8.42	8.4355	-0.0155	0.0018	-0.0133
18	8.45	8.4472	0.0028	0.0003	0.0022
19	8.45	8.4588	-0.0088	0.0010	-0.0014
20	8.45	8.4705	-0.0205	0.0024	-0.0014
21	8.42	8.4821	-0.0621	0.0074	-0.0049
22	8.41	8.4938	-0.0838	0.0100	-0.0026
23	8.41	8.5055	-0.0955	0.0114	-0.0014
24	8.42	8.5172	-0.0972	0.0115	-0.0002

Table 7: Classified grey model checking.

	Relative error	Accuracy	Posterior ratio	Small error difference	Correlation probability	Level deviation
I	0.01	0.99	0.35	0.95	0.9	1%
II	0.05	0.95	0.50	0.80	0.8	5%
III	0.10	0.90	0.65	0.70	0.7	10%
IV	0.20	0.80	0.80	0.60	0.6	20%

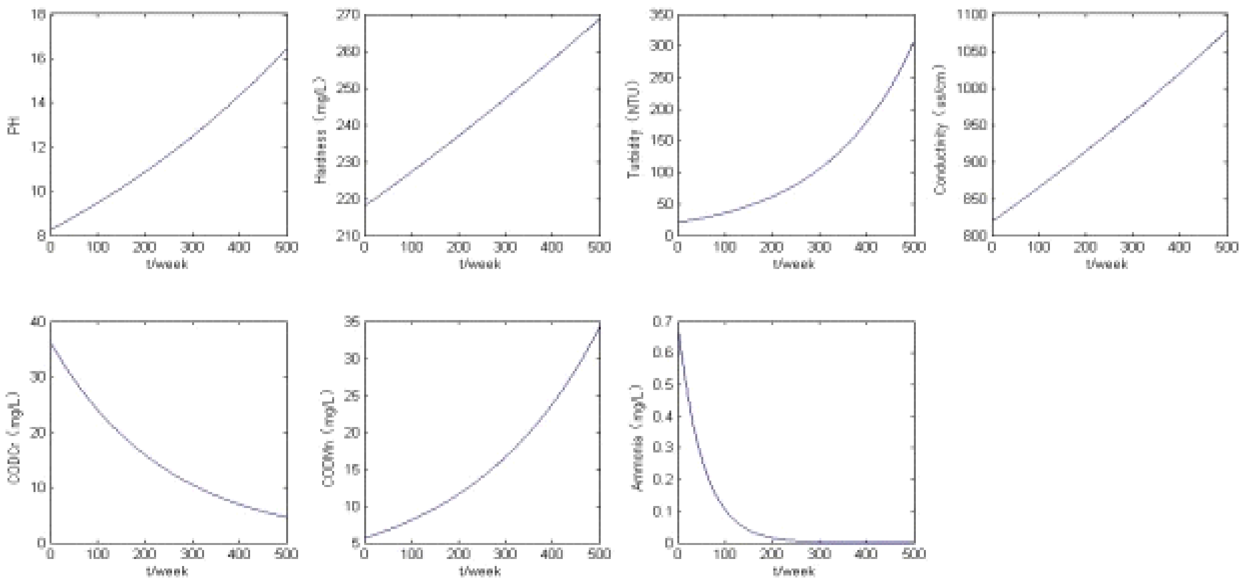

Fig. 2: The next 20 years, the situation changes in water quality indicators.

$x^1_{turbidity}\ (k+1) = 3913.95e^{0.00536475\,k} - 3895.25$

$x^1_{hardness}\ (k+1) = 521069.0e^{0.000418382\,k} - 520868.0$

$x^1_{conductivity}\ (k+1) = 1487666.0e^{0.000550335\,k} - 1486855.0$

$x^1_{CODCr}\ (k+1) = -8800.47e^{-0.00411213\,k} + 8822.87$

$x^1_{CODMn}\ (k+1) = 1567.85e^{0.00360378\,k} - 1562.15$

$x^1_{ammonia}\ (k+1) = -36.6238e^{-0.0188454\,k} + 37.1708$

According to the gray prediction model established above, we can draw water quality indicators to predict future changes in the curve as shown in Fig. 2

RESULTS ANALYSIS

1. Calculate the water quality index weights based on entropy method known (Table 5), ammonia content and potassium permanganate index is the main factor affecting the Yongding River water quality changes. These two indicators are related to the organic content of the water. Accordingly, we can make investigation in respect of the source of nitrogen and other organic pollutants and transmission to determine the appropriate treatment plan.
2. As shown in Fig. 2, under the same conditions in the current environment, in the next 20 years, CODCr and ammonia concentration showed a downward trend, while the pH, hardness, turbidity, conductivity and CODMn showed upward trend in concentration; turbidity and conductivity changes are the fastest.
3. According to the analysis of information entropy model, it shows that Yongding water quality is associated with the highest degree of CODMn, and ammonia is decreased, while several other correlations are not very different. So the next 20 years should strengthen the CODMn, turbidity and conductivity monitoring and take appropriate control measures.
4. In its natural state, the self-purification capacity gradually restored Yongding River, with upstream reservoir

recycling water, by which water quality can be improved in the Yongding River.

CONCLUSION

Through analysis of information entropy correlation, seven basic water quality indicators were calculated and analysed. We come to the conclusion that ammonia and CODMn are the highest degree associated with water quality of Yongding. On the basis of gray system forecasting model, the status and development of the law of the Yongding River water quality was analysed, and found that turbidity and conductivity changed quickest in the next two decades. The results show that the current Yongding water quality is ordinary and has a certain water purification capacity, but authorities need to increase monitoring efforts and take effective control measures to prevent further deterioration of the Yongding River water quality. Through the establishment of the theoretical analysis and mathematical models, we give the actual and reasonable opinion about governance of the Yongding River.

ACKNOWLEDGMENTS

This work was supported by the funds for undergraduate innovative experiment plan of North China Electric Power University (national level), and the Open Research Fund Program of State key Laboratory of Hydroscience and Engineering, No. sklhse-2013-A-03.

REFERENCES

Han, Y.P., Wang, Y.B. and Feng, X.M. 2011. Comprehensive evaluation on regional water shortage risk maximum based on entropy principle. Anhui Agricultural Sciences, 1: 398-399.

He, L. and Chen, X.H. 2001. Groundwater evaluation model based on the principle of maximum entropy. Advances in Water Science, 1: 62-64.

Huang, S.B. 2012. Application of river water quality assessment based on entropy theory. Environmental Science and Management, 8: 179-194.

Li, F.Q. 2006. Stability and accuracy of gray model modeling. Wuhan University of Technology.

Long, R.S., Men, B.H. and Zhao, Y.W. 2013. The investigation about bioremediation to metal pollution of water and soil environment in Yongding river. Nature Environment & Pollution Technology, 4(12): 727-731.

Miao, J. 2012. Yongding river water pollution bioremediation. Beijing Hydrological Science and Technology Seminar Papers, 3: 75.

Shen, F.S. 2007. Irrigation water demand forecast based on GM(1,1) model. Water Resources Protection, 3: 33-34.

Wang, K.Z., Liu, F.S. and Sun, M. 2002. Grey model in water DAWU prediction. Shandong Agricultural University (Natural Science), 1: 67-69.

Zhao, Q.L. 2012. River water environment assessment and application based on the theory entropy. Environmental Science and Management, (8): 179-194.