

Application of Improved Gray Clustering Method in Ming-guan Campus

Fuhui Du^a, Yuan Wang^b

College of Water Conservancy and Hydropower Engineering, Hebei University of Engineering,
Handan 056021, China

^adufuhui@hebeu.cn, ^b15670832032@163.com

Keywords: Grey clustering; Weight; Fuzzy evaluation; Water assessment

Abstract. In order to verify the feasibility and innovation of grey clustering method for the introduction of weight theory in water quality evaluation. Ming-guan campus of Hebei University of Engineering set up 8 monitoring points to continue water quality monitoring in Handan for 7 days as a cycle of October, November and December in 2016. According to the local water pollution, selected 7 indicators as evaluation factors, such as DO, BOD₅, COD_{Cr}, COD_{MN}, TP, TN and NH₃-N. The results show that: the standard of grade 2 account for 22% of the overall assessment, the standard of grade 3 account for 67% and the standard of grade 4 account for 11%. The water quality in October is significantly better than that in November and December. Compared with the traditional fuzzy water quality evaluation results, result shows that the grey clustering method based on weight is more convenient and reasonable than the traditional fuzzy evaluation.

Introduction

Water quality assessment is a non-linear and complex integrated system controlled by various factors. In order to analyze water quality comprehensively, a variety of factors affecting water quality should be considered. Therefore, the water quality evaluation of the water system is comprehensive [1,2]. In recent years, domestic and foreign scholars have put forward a variety of comprehensive evaluation of water quality methods combined with the comprehensive and complex of the water quality evaluation, such as: single-factor evaluation method, fuzzy evaluation method, BP neural network method, gray clustering analysis method and so on [3,4,5]. The advantages and disadvantages of the method rely on the processing of the data, the complexity of the algorithm itself and its applicability [6,7]. This paper established a water quality evaluation method based on gray system weight clustering analysis, and taking the comprehensive evaluation of the water quality of the experimental field of ming-guan campus of hebei engineering university as an example to evaluate the water quality to verify the feasibility of the method.

Evaluation Method

Gray quality analysis was used to evaluate the water quality of the test site, weight of the gray clustering method specific steps are as follows:

Composition of Cluster Samples. There are k monitoring points, each of it has i pollutant indexes and each pollutant index has j gray, k sample points and i indexes can form matrix C_{ki} , which represents the concentration value of the i th contaminant index at the k th sample monitoring point, $k \in [1, 2, \dots, m]$, $i \in [1, 2, \dots, n]$, $j \in [1, 2, \dots, s]$.

Establishment of Whitening Function of Gray. According to the water quality classification standard of our country, the water quality is divided into s class, so there has s grey class. According to the different clustering indicators, the different gray class can determine the different whitening function [8]. The whitening function formula of gray class 1, gray class j and gray class s of i th indexes are:

- (1) 1class water ($j=1$), the membership function is:
- (2) j class water ($j=2, 3, \dots, s-1$), the membership function is:
- (3) s class water ($j=s$), the membership function is:

$$f_{i1}(x) = \begin{cases} 1 & x \leq x_1 \\ \frac{x_2 - x}{x_2 - x_1} & x_1 < x < x_2 \\ 0 & x \geq x_2 \end{cases} \quad f_{ij}(x) = \begin{cases} (x - x_{j-1}) / (x_j - x_{j-1}) & x_{j-1} < x \leq x_j \\ (x_{j+1} - x) / (x_{j+1} - x_j) & x_j < x < x_{j+1} \\ 0 & x \leq x_{j-1} \text{ or } x \geq x_{j+1} \end{cases} \quad f_{is}(x) = \begin{cases} 1 & x \geq x_s \\ \frac{x - x_{s-1}}{x_s - x_{s-1}} & x_{s-1} < x < x_s \\ 0 & x \leq x_{s-1} \end{cases} \quad (1)$$

x is the sample value, $x_1, x_2, x_{j-1}, x_j, x_{j+1}, x_{s-1}, x_s$ are represent 1, 2, $j-1, j, j+1, s-1, s$ class water.

Calculate Cluster Right. In the process of water quality evaluation, due to the different clustering units, we cannot directly calculate the indicators and should deal with the gray class, the introduction of clustering right to measure the weight of same gray matter is extremely important. The formula is:

$$z_{ij} = \frac{v_{ij}}{\sum_{i=1}^n v_{ij}} \quad \text{Among them} \quad v_{ij} = \frac{S_{ij}}{\sum_{i=1}^n S_i} \quad (2)$$

z_{ij} Represents the quantization value of the j th gray class of i th pollution index;

S_{ij} Represents the standard value of the j th class representing i th pollution index;

S_i Represents a reference standard for i pollution target; $i \in [1, 2, \dots, m], j \in [1, 2, \dots, s]$.

Weight Method Correction Factor. The weight can measure the validity of the provided data information and the proportion of the effectiveness. As follows: the smaller the value of the sample, the smaller the effective amount of information provided by the indicator; if not, the smaller the value of the sample. Specific steps are as follows:

(1) Normalize Original Data Matrix

For the larger value is the excellent indicators (DO) and the smaller value is the excellent indicators (COD):

$$v_{ij} = \frac{x_{ij} - \min_j \{x_{ij}\}}{\max_j \{x_{ij}\} - \min_j \{x_{ij}\}} \quad v_{ij} = \frac{\max_j \{x_{ij}\} - x_{ij}}{\max_j \{x_{ij}\} - \min_j \{x_{ij}\}} \quad (3)$$

n_{ij} Represents the standardized value of i th pollution index in j th grey class;

x_{ij} Represents the original data of i th pollution index in j th grey class;

$\max_j \{x_{ij}\}$ Represents the Max value of i th pollution index in j th grey class;

$\min_j \{x_{ij}\}$ Represents the Min value of i th pollution index in j th grey class.

(2) Determine Weight

The formula of the weight of the i th index is:

$$g_i = -k \sum_{j=1}^n f_{ij} \ln f_{ij} \quad \text{Among them} \quad f_{ij} = n_{ij} / \sum_{j=1}^n n_{ij}, k = 1 / \ln n, i = 1, 2, \dots, m \quad (4)$$

(3) Calculate Weight

The formula of the right of the i th index is:

$$l_i = \frac{1 - g_i}{m - \sum_{i=1}^m g_i} \quad \text{Among them} \quad 0 \leq l_i \leq 1, \sum_{i=1}^m l_i = 1 \quad (5)$$

(4) Using the weight correction index to calculate the weight coefficient matrix is:

$$z'_{ij} = l_i z_{ij} / \sum_{i=1}^m l_i z_{ij} \quad (6)$$

Calculate Cluster Coefficient. The clustering coefficients are brought into the respective whitening functions, and it can reflect the degree of clustering samples. The formula is:

$$x_{ij} = \sum_{i=1}^n f_{ij} (C_{ki}) z_{ij} \quad (7)$$

Constructe Cluster Vector. According to the principle of maximum classification, determine the classification of objects belonging to which categories. The vector is: $X_{ij} = \{x_{i1}, x_{i2}, \dots, x_{is}\}$.

Calculate the Weighted Average of the Comprehensive Evaluation. Because the gray clustering coefficient has the disadvantage of losing a lot of information data, the idea of introducing

the weighted average can avoid the influence of this error on the result. The formula is:

$$H_k = \sum_{j=1}^s X_{ij} / \sum_{j=1}^s (h_i \cdot X_{ij}) \quad (8)$$

H_k represents the rank of the k th sample monitoring point; h_i represents the weight of the evaluation result of ξ_{ij} , $h_i=(1, 2, \dots, s)$.

Application Example

Sample Settings and Sample Collection.

Sample Setting. The pilot area set 8 monitoring points in the Ming-guan campus of HeBei Engineering University. The wide, long and deep of the pool are 8m, 4m and 0.5m. The sampling points are located in the variety of plant space.

Sample Collection. Samples were monitored at 8 monitoring sites every 7 days from October to December and monitored for water quality. The collection, transportation and preservation of the samples are carried out in accordance with the regulations [9].

Monitoring Methods and Monitoring Indicators.

Monitoring Method. The methods and instruments for monitoring water samples are monitored in the fourth edition of "water and wastewater monitoring and analysis methods" published by China Environmental Science Press, prepared by the State Environmental Protection Administration. Tables (refer with: Table 1) is the specific water quality monitoring methods.

Table1 Indicators and methods of water quality monitoring

Index	Method	Location
DO	Electrochemical Probe Method	Site
BOD ₅	Dilution and Inoculation Method	Laboratory
COD _{Cr}	Potassium Dichromate Method	Laboratory
COD _{MN}	Acid Method	Laboratory
TP	Ammonium Molybdate Spectrophotometric Method	Laboratory
TN	Alkaline Potassium Persulfate Digestion-UV Spectrophotometric Method	Laboratory
NH ₃ -N	Nessler's Reagent Spectrophotometric	Laboratory

Monitoring Indicator. According to the monitoring results of the test site, the following seven evaluation factors were determined: DO, BOD₅, COD_{Cr}, COD_{MN}, TP, TN and NH₃-N. The evaluation criteria are strictly in accordance with the "Surface Water Environmental Quality Standard". Therefore, the monitoring concentration of each monitoring point is the whitening number; the cluster sample number; the class indicators and the number of gray categories respectively are $m=8$ ($k=1,2,\dots,8$), $n=7$ ($i=1,2,\dots,7$) and $s=5$ ($j=1,2,\dots,5$).

Evaluation Results and Analysis. Due to too much data, only 10.04 whitening numbers are listed, Tables (refer with: Table 2) is the data.

Table2 10.04 Water quality monitoring value of each monitoring point

Date	Point	DO	BOD ₅	COD _{Cr}	COD _{MN}	TP	TN	NH ₃ -N
10.04	1#	7.52	7.2	32	6.52	0.228	1.54	2.86
	2#	7.44	6.7	24	5.31	0.202	1.21	2.69
	3#	5.97	8.2	20	5.96	0.195	1.35	2.08
	4#	6.81	7.9	28	6.31	0.184	1.45	2.56
	5#	7.36	6.1	20	6.67	0.089	1.42	2.00
	6#	7.91	5.5	16	6.14	0.102	1.51	2.59
	7#	7.59	4.9	12	5.81	0.087	1.40	1.62
	8#	7.33	3.7	8	5.31	0.091	1.31	1.78

Clustered Whitening Number.

Establishment of Whitening Function and Derivation of Value. According to the whitening function formula (1), we can get t 7 kinds of evaluation factors. such as COD_{MN}:

$$f_1(x) = \begin{cases} 1 & x \leq 2 \\ \frac{4-x}{2} & 2 < x < 4 \\ 0 & x \geq 4 \end{cases} \quad f_2(x) = \begin{cases} \frac{x-2}{2} & 2 < x \leq 4 \\ \frac{6-x}{2} & 4 < x < 6 \\ 0 & x \leq 2 \text{ or } x \geq 6 \end{cases} \quad f_3(x) = \begin{cases} \frac{x-4}{2} & 4 < x \leq 6 \\ \frac{10-x}{4} & 6 < x < 10 \\ 0 & x \leq 4 \text{ or } x \geq 10 \end{cases} \quad f_4(x) = \begin{cases} \frac{x-6}{4} & 6 < x \leq 10 \\ \frac{15-x}{5} & 10 < x < 15 \\ 0 & x \leq 6 \text{ or } x \geq 15 \end{cases} \quad f_5(x) = \begin{cases} 1 & x \geq 15 \\ \frac{x-10}{5} & 10 < x < 15 \\ 0 & x \leq 10 \end{cases}$$

As the other indicators are smaller values that the water quality better, but DO is contrast, the whitening function should be adjusted accordingly:

By substituting the data in table 1 into the above formula, we can obtain all kinds of standard whitening function values, such as the whitening function value of each monitoring cross point of class IV standard ($f_{4j}(c_{kj})$).

Calculate the Cluster Right. The dimensionless treatment is carried out with the class I water as the parameter. According to the formula (2) to calculate the cluster right, results is:

$$f_{4j}(c_{kj}) = \begin{bmatrix} 0.1429 & 0.0512 & 0.0238 & 0.0095 & 0.0047 \\ 0.1429 & 0.0640 & 0.0476 & 0.0477 & 0.0584 \\ 0.1429 & 0.0640 & 0.0476 & 0.0477 & 0.0467 \\ 0.1429 & 0.1279 & 0.1071 & 0.1193 & 0.1313 \\ 0.1429 & 0.3198 & 0.3571 & 0.3580 & 0.3503 \\ 0.1429 & 0.1599 & 0.1786 & 0.1790 & 0.1751 \\ 0.1429 & 0.2132 & 0.2381 & 0.2387 & 0.2335 \end{bmatrix} = \begin{bmatrix} 0 & 0.7 & 0.8 & 0.13 & 0.28 & 0.92 & 0 \\ 0 & 0.83 & 0.4 & 0 & 0.02 & 0.42 & 0 \\ 0 & 0.45 & 0 & 0 & 0 & 0.7 & 0 \\ 0 & 0.53 & 0.8 & 0.08 & 0 & 0.9 & 0 \\ 0 & 0.98 & 0 & 0.17 & 0 & 0.84 & 0 \\ 0 & 0.75 & 0 & 0.03 & 0 & 0.98 & 0 \\ 0 & 0.45 & 0 & 0 & 0 & 0.8 & 0.76 \\ 0 & 0 & 0 & 0 & 0 & 0.62 & 0.44 \end{bmatrix}$$

Calculate Weight1 and Weight2 of Indicator. According to formula (3), the original data matrix was standardized. The weight1 is calculated according to the formula (4), and the weight2 is calculated according to the formula (5). Tables (refer with: Table 3) is that.

Table3 Entropy value and entropy weight result of evaluation index

Date	Index	DO	BOD ₅	COD _{Cr}	COD _{MN}	TP	TN	NH ₃ -N
10.04	Weight1	0.9243	0.8405	0.7557	0.6747	0.6214	0.5963	0.5527
	Weight2	0.0372	0.0784	0.1201	0.1599	0.1861	0.1984	0.2199

Calculate Weight Coefficient. According to the formula (6) to calculate the new clustering weight, the results are as follows:

Calculate Cluster Coefficient. According to the formula (7) to calculate the clustering coefficient, we can get the monitoring of the water quality of the clusters of coefficients formed by the matrix:

0.0372	0.011	0.0049	0.0019	0.001	0.0372	0	0.3453	0.3325	0.31
0.0784	0.0290	0.0205	0.0204	0.0251	0.0357	0.0406	0.5526	0.1179	0.2854
0.1201	0.0444	0.0315	0.0312	0.0307	0	0.0302	0.5299	0.1447	0.2949
0.1599	0.1181	0.0942	0.1039	0.115	0.0201	0.06	0.4197	0.2183	0.2929
0.1861	0.3437	0.3657	0.363	0.3569	0.0599	0.2965	0.1409	0.2002	0.2816
0.1984	0.1832	0.195	0.1935	0.1903	0.0372	0.3723	0.1102	0.2081	0.2849
0.2199	0.2707	0.2881	0.2860	0.2811	0.1871	0.2993	0.1361	0.3813	0.0675
					0.1737	0.3559	0.1507	0.2458	0.1574

Comprehensive Evaluation. According to the formula (8) to calculate the evaluation results and the results and analyzed it. Tables (refer with: Table 4) is the result.

Comparison with Fuzzy Evaluation Method. In order to verify the feasibility of the gray comprehensive evaluation method, the fuzzy evaluation method is selected as the control[10]. Fuzzy evaluation method is a variety of factors that affect things to do a comprehensive evaluation of the method. Tables (refer with: Table 5) is the result.

Table4 Water quality grade results

Date	1#	2#	3#	4#	5#	6#	7#	8#
10.04	4	4	4	4	3	3	3	3
10.11	4	3	3	3	3	3	2	2
10.18	3	3	3	3	3	3	2	2
10.25	3	3	3	3	3	3	3	2
11.02	4	4	4	3	3	3	3	2
11.09	4	3	3	3	3	3	3	3
11.16	4	3	3	3	3	3	3	2
11.23	4	4	4	3	3	3	3	2
12.01	4	4	4	3	3	3	3	3

Table5 Traditional fuzzy comprehensive evaluation results

Date	1#	2#	3#	4#	5#	6#	7#	8#
10.04	4	4	4	4	4	4	4	4
10.11	3	3	3	3	2	2	2	2
10.18	3	3	3	3	3	3	3	3
10.25	3	3	3	3	3	2	2	2
11.02	4	4	4	3	3	3	3	2
11.09	4	3	4	3	3	3	3	3
11.16	4	3	3	3	3	3	3	2
11.23	4	4	4	3	3	3	3	2
12.01	4	4	4	3	2	3	3	3

Results Analysis

Using the gray clustering method and the traditional fuzzy comprehensive evaluation method to monitor the water samples, the results are basically the same, which shows that the water quality of the 8 monitoring points are between II ~ IV water. The result is consistent with the situation at that moment, it shows that the gray clustering method with weight can be applied in practice.

It can be seen from the table 8 that there are differences in time and space between monitoring sections, which are more obvious in space. From the spatial distribution, 1 # monitoring point to 8 # monitoring point of water quality from the class IV water to type II water to change, there is a clear

trend of change. In terms of time distribution, the water quality is better in October, the water quality was poor in November and December [11]. It is possible that there is more rain in October, and the amount of water can be diluted to make water self-cleaning. During this period, aquatic plants grow well, which can play a strong purification effect on water quality, and can effectively reduce the degree of eutrophication in water, inhibit the growth of phytoplankton in water, maintain water ecological balance, so the water quality improved. In the month of 11 and 12, the rainfall was relatively small and the water temperature was relatively low, which was harmful to the diffusion and volatilization of microorganisms. The aquatic plants wilt and die, and the aquatic plants not be fished out in time. The rotting plant roots release a large number of microorganisms and increase the eutrophication degree of the water body, and then increase the concentration index of the pollutants.

Conclusion

It can be seen from the application examples of this paper that the gray clustering method with weight has a strict theoretical basis, calculation and evaluation method. The algorithm considers the grayness of the system and is an important characteristic of comprehensive evaluation of water quality. Through this algorithm, the influence of various factors can be comprehensively evaluated, and the water quality can be accurately reflected. The grey clustering method can determine the weights of different grades according to different water quality categories, which can avoid the traditional fuzzy algorithm only the introduction of a weight of the unreasonable, making the evaluation results more reliable. The weight coefficient of the index is modified by introducing weights, which can effectively utilize the information left by the sample. The principle of weighted comprehensive evaluation is used to replace the principle of maximum clustering in traditional algorithms, which is more favorable to the superiority of clustering results. The weighted grey clustering method does not increase the calculation of the algorithm, but the application is more convenient and fast, and it can meet the requirements of water quality comprehensive evaluation. It is proved that weight-based gray clustering method is a method for scientific water quality evaluation.

Acknowledgements

This work was financially supported by the research center of water ecological civilization and social governance in Hebei province. (16273604D)

References

- [1] Q.J. Hao. in: Journal of Southwest China University, Vol. 36, No. 4, p. 115-122(2011). (In Chinese)
- [2] X. Z, C.S. Jiang, Q.J. Hao. in: Chinese Agricultural Science Bulletin, Vol. 26, No. 2, p. 241-245(2010). (In Chinese)
- [3] Z.L. Li, Q.M. Ma, S.J. Xu, et al. in: Transactions of Oceanology and Limnology. No. 3, p. 139-144 (2011). (In Chinese)
- [4] Bin Qian, Qiyang Feng, Ting Li, et al. in: Water Saving Irrigation, No. 2, p. 50-53(2014). (In Chinese)
- [5] T. Sun, M.X. Zhang, M.M. Li, et al. in: Environmental Science Technology, Vol. 37, No. 4, p. 185-190(2014). (In Chinese)
- [6] Y. Yun, Z.H. Zou. In: Mathematics in Practice and Theory, Vol. 37, No. 12, p. 65-72 (2007). (In Chinese)
- [7] B. P, Y.X. Li, L. Tong, et al. in: Environmental Science Technology, Vol. 34, No. 11, p. 185-188(2011). (In Chinese)

- [8] Y. Zhao, T. Wang, X.i Xu. In: Yellow River, Vol. 34, No. 7, p. 56-58(2012). (In Chinese)
- [9] F.Y. Yu, M.Z. Zhang, Z.Y. Pang. In:Analysis on Improvement of Water Quality Improvement in Demonstration Area of Baiyun Lake Aquatic Ecosystem[C]. The Sixth National Conference on River Regulation and Ecological Restoration in, 014:217-222(2014). (In Chinese)
- [10] P.Z. Wang. In:Application of fuzzy mathematics[M]. BeiJing: Beijing University of Economics Press,1989. (In Chinese)
- [11] Y. Zhang. In: Journal of Green Science and Technology, No. 3, p. 59-61(2014). (In Chinese)