Water Quality Evaluation of the Yellow River Basin Based on Gray Clustering Method

X Q Fu, Z H Zou*
School of Economics and Management, Beihang University, Beijing 100191, China
623688881@qq.com
*Corresponding Author

Abstract. Evaluating the water quality of 12 monitoring sections in the Yellow River Basin comprehensively by grey clustering method based on the water quality monitoring data from the Ministry of environmental protection of China in May 2016 and the environmental quality standard of surface water. The results can reflect the water quality of the Yellow River Basin objectively. Furthermore, the evaluation results are basically the same when compared with the fuzzy comprehensive evaluation method. The results also show that the overall water quality of the Yellow River Basin is good and coincident with the actual situation of the Yellow River basin. Overall, gray clustering method for water quality evaluation is reasonable and feasible and it is also convenient to calculate.

1. Introduction
With the development of social economy, water quality has become the focus of people's concern, people are increasingly demanding water resources, scientific and reasonable assessment of water quality is particularly important. At present, there are many methods to evaluate water environment quality, such as matter element analysis, analytic hierarchy process, nonlinear optimization and set pair analysis, grey relation analysis, artificial neural network, fuzzy comprehensive evaluation, etc[1-6]. Since the quality of water environment data obtained are limited in time and space obtained monitoring information is incomplete or inaccurate[7-9], that is some information in the system is known and some information is unknown, so the principle of gray system should be applied to evaluate the water quality[10-11]. Taking the Yellow River Basin as the research object, the application of gray clustering method in water environment quality evaluation is described.

2. Gray clustering method
Gray clustering method is based on the generation of definite weighted functions of gray numbers, and aggregates some of the observed indicators or objects into several definable categories.

2.1. Determination of the cluster whitening number matrix
There are k samples (monitoring section), and each have i indicators, each indicator has j gray class, by the k number of indicators of the whitening number of matrix (dki)m×n.Where dki is the initial value of the i-th clustering index for k clustering samples, among them, k=1,2,···;n,i=1,2,···;m,j=Ⅰ,Ⅱ,···,c.
2.2. The Determination of gray definite weighted functions $f_i(x)$

The $n$ objects on the indicator $i$ value of the corresponding surface water quality as $j$ gray class, pollution standards at all levels are gray. Determining definite weighted functions $f_i(x)$ is index $j$ subclass. The threshold $\lambda_{ij}$ of the definite weighted functions is taken from the water quality grading standard with a peak of 1.

2.3. Determination of clustering weight $Z_{ij}$

According to the purpose of the use of ground water and protection objectives to determine the reference standard for the waters. The number of gray categories for non-dimensional processing.

$$Z_{ij} = \frac{V_{ij}}{\sum_{i=1}^{n} V_{ij}}$$  \hspace{1cm} (1)

$$V_{ij} = \frac{S_{ij}}{S_{0i}}$$  \hspace{1cm} (2)

In the formula, $V_{ij}$ is the dimensionless number of the $j$-th gray category of the $i$-th pollution index, $S_{ij}$ is the gray number (standard value) of the $j$-th gray category of the $i$-th pollution index, and $S_{0i}$ is the reference criterion of the $i$-th pollution index.

2.4. The determination of the clustering coefficient $W_{ij}$

The clustering whitening number $d_{ki}$ is substituted into the respective definite weighted functions $f_i(x)$ to obtain $f_i(d_{ki})$, and the clustering coefficient $W_{ij}$ is obtained.

$$W_{ij} = \sum_{k=1}^{n} f_i(d_{ki}) \times Z_{ij}$$  \hspace{1cm} (3)

2.5. Clustering

Construct the clustering vector first: $(W_{i1},W_{i2}, \ldots, W_{ic})$,The maximum clustering coefficient $W$ is found in the clustering row vectors $(W_{i1},W_{i2}, \ldots, W_{ic})$, and the gray class $j$ corresponding to the largest clustering coefficient is the gray class of the clustering object $k$.

3. Evaluation of water quality in the Yellow River Basin

3.1. Data Sources

Select the 12 monitoring sections of the Yellow River as the research object, such as Lanzhou New Town Bridge, Zhongwei Xindun, Shizuishan Ma Huanggou, Wuhai Haibo Bay, Baotou Painter Yingzi, Xinzhou Wanjiazhai Reservoir, Jiyuan Xiaolangdi, Jinan Luokou, Haidong Minhe Bridge, Yuncheng Hejin Bridge, Tianshui Niubei, Weinan Tongguan suspension bridge, and they are numbered 1-12.

Data from weekly monitoring in May 2016 were used and averaged for monitoring data in May.DO, COD$_{Mn}$, and NH$_3$-N were used as indicators for water quality evaluation. Table 1 shows all the data from the website of China's Ministry of Environmental Protection Data Center. With "Surface Water Environmental Quality Standard" (GB3838-2002) \cite{14} for the evaluation criteria levels (Table 2).The water quality level is set to 5, so the gray class is divided into 5 categories, the application of gray clustering method to evaluate the water quality.

| Monitoring sections | DO (mg/L) | COD$_{Mn}$ | NH$_3$-N |
|---------------------|-----------|------------|----------|

Table 1. Water quality of the Yellow River Basin monitoring sections in May 2016 (Unit: mg / L).
Table 2. Indicator limits for water quality assessment (Unit: mg / L).

| project                        | Class I | Class II | Class III | Class IV | Class V |
|--------------------------------|---------|----------|-----------|----------|---------|
| DO≥6                           | 7.5     | 6        | 5         | 3        | 2       |
| CODMn≤2                        | 4       | 4        | 6         | 10       | 15      |
| NH₃−N≤0.15                     | 0.5     | 1        | 1.5       | 2        |

3.2. Evaluation analysis

3.2.1. Determination of the cluster whitening number matrix The 12 monitoring sections of the Yellow River Basin were used as cluster samples (k=1,2,…,12), The DO, CODMn and NH₃−N values as clustering indicators (i=1,2,3). The clustering whitening number $d_{ki}$ is the measured value of each monitoring index on each monitoring section.

3.2.2. Gray definite weighted functions According to the definite weighted functions can be obtained the $f_i(x)$ of DO, CODMn and NH₃−N. Such as CODMn gray definite weighted function:

$$f_1(x) = \begin{cases} 
1 & x \in [0, 2) \\
\frac{x-4}{4-2} & x \in [2, 4) \\
0 & x \in [4, \infty)
\end{cases}$$

Because DO is different from other pollutants, the higher the DO value, the better the water quality, and the definite weighted function calculation should be changed accordingly, such as:
3.2.3. Calculate the clustering weight $Z_{ij}$ Class I water quality standards as a reference standard, with formula (2) for non-dimensional treatment, the results shown in Table 3.

Table 3. Clustering indicators dimensionless processing results.

| Clustering index | I   | II  | III | IV  | V   |
|------------------|-----|-----|-----|-----|-----|
| DO               | 1.00| 0.80| 0.67| 0.40| 0.27|
| CODMn            | 1.00| 2.00| 3.00| 5.00| 7.50|
| NH3-N            | 1.00| 3.33| 6.67| 10.00|13.33|
| Total            | 3.00| 6.13|10.33|15.40|21.10|

Use the formula (1) to compute the cluster weight $Z_{ij}$

$$Z_{ij} = \begin{pmatrix} 0.333 & 0.130 & 0.065 & 0.026 & 0.013 \\ 0.333 & 0.326 & 0.290 & 0.325 & 0.355 \\ 0.333 & 0.543 & 0.645 & 0.649 & 0.632 \end{pmatrix}$$

3.2.4. Calculate the clustering coefficient Substituting the $d_{ki}$ in the matrix $d$ into the definite weighted functions $f_i(x)$, the definite weighted function value $f_i(d_{ki})$ of each section is obtained. For example, the definite weighted function value $f_i(d_{ki})$ of each monitoring section relative to class I is:

$$f_{ii}(d_{ki}) = \begin{pmatrix} 1.00 & 0.338 & 0.663 \\ 1.00 & 1.00 & 1.00 \\ 1.00 & 0.613 & 0.734 \\ 1.00 & 0.188 & 1.00 \\ 1.00 & 0.000 & 0.986 \\ 0.373 & 0.863 & 0.691 \\ 0.563 & 0.700 & 1.00 \\ 1.00 & 1.00 & 1.00 \\ 0.637 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 1.00 & 0.988 & 0.820 \\ 0.000 & 0.000 & 0.000 \end{pmatrix}$$

Calculate the clustering coefficient $W_{ij}$ according to the formula (3), and obtain the matrix $D$ of the water quality of the monitoring section.
3.2.5. Comprehensive evaluation  According to the clustering coefficient of the maximum classification principle, get the following results (see Table 4).

### Table 4. Comprehensive evaluation results.

| Monitoring sections                  | Level |
|--------------------------------------|-------|
| 1 Lanzhou New Town Bridge            | I     |
| 2 Zhongwei Xindun                    | I     |
| 3 Shizuishan Ma Huanggou             | I     |
| 4 Wuhai Haibo Bay                    | I     |
| 5 Baotou Painter Yingzi              | I     |
| 6 Xinzhou Wanjiazhai Reservoir       | I     |
| 7 Jiyuan Xiaolangdi                  | I     |
| 8 Jinan Luokou                       | I     |
| 9 Haidong Minhe Bridge               | V     |
| 10 Yuncheng Hejin Bridge             | V     |
| 11 Tianshui Niubei                   | I     |
| 12 Weinan Tongguan suspension bridge | III   |

According to the results of the evaluation, in May 2016, there were 9 monitoring sections of grade I water quality in the Yellow River Basin, Weinan Tongguan suspension bridge is III grade water quality monitoring section, Haidong Minhe Bridge and Yuncheng Hejin Bridge are V grade water quality monitoring section. In May 2016, the overall water quality of the Yellow River basin was better.

The fuzzy comprehensive evaluation of the river water quality is carried out by using the method of exceeding the weighting method, and the evaluation result is compared with the gray clustering method, as shown in Table 5.

### Table 5. Comparison of evaluation results.

| Evaluation results |       |
|--------------------|-------|

\[
D = \begin{pmatrix}
0.667 & 0.399 & 0.000 & 0.000 & 0.000 \\
1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.782 & 0.271 & 0.000 & 0.000 & 0.000 \\
0.729 & 0.265 & 0.000 & 0.000 & 0.000 \\
0.662 & 0.285 & 0.044 & 0.000 & 0.000 \\
0.642 & 0.294 & 0.000 & 0.000 & 0.000 \\
0.754 & 0.155 & 0.000 & 0.000 & 0.000 \\
1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.212 & 0.178 & 0.174 & 0.236 & 0.402 \\
0.000 & 0.000 & 0.058 & 0.146 & 0.831 \\
0.936 & 0.102 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.334 & 0.590 & 0.037 & 0.000 \\
\end{pmatrix}
\]
### Monitoring sections

| Monitoring sections | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---------------------|---|---|---|---|---|---|---|---|---|----|----|----|
| Gray clustering method | I | I | I | I | I | I | I | V | V | I | III |
| Fuzzy comprehensive evaluation method | I | I | I | I | I | I | I | V | V | I | III |

The results of the evaluation are basically the same, which indicates that the gray clustering method can be used for the water quality evaluation of the Yellow River Basin.

### 4. Conclusion

The gray clustering method is used to evaluate the water quality, and its theory is easy to understand and simple and convenient. Gray clustering considers an important characteristic of the degree of water pollution, that is, the grayness of the system, which solves the shortcomings of the single factor evaluation for the individual factors, which can reflect the influence of each evaluation factor more comprehensively and thus provide more abundant of the information, so that more accurate judgments for the water environment quality evaluation will provide a reliable scientific basis.

### Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 51478025).

### References

[1] Li F P and Tian Y 2016 *J. Hubei Agricultural Sciences* 2016 55 4154-57.
[2] Li X and Yuan Z T 2014 *J. Yangtze River* 45 25-27.
[3] Yang X H, Zhang X J, Hu X X. 2001 *J. Nonlinear Processes in Geophysics* 18 599-607.
[4] Wang P and Wang Y F 2013 *J. Water Resources Protection* 29 52-54.
[5] Ni S H and Bai Y H 2000 *J. System Engineering Theory and Practice* 8 124-127.
[6] Fu X Q and Zou Z H 2017 *J. Mathematics in Practice and Theory* 47 181-186.
[7] Deng X, Chen Q H and Zhang J S 2010 *J. Environmental Science and Management* 35 187-191.
[8] An Y X 2014 *J. Water Sciences and Engineering Technology* 2 69-72.
[9] Sze S M 1969 *Physics of Semiconductor Devices* (New York: Wiley–Interscience)
[10] Deng J L 1987 *Gray System Basic Method* (Wuhan: Huazhong Institute of Technology Press)
[11] Wang S L and Zou Z H *J. Mathematics in Practice and Theory* 2012 42 84-89.
[12] Zhang X, Jiang C S and Hao Q J 2010 *J. Chinese Agricultural Science Bulletin* 26 241-245.
[13] Wang H M, Lu W X and Xin G 2007 *J. Water Saving Irrigation* 5 20-22.
[14] Yun Y and Zou Z H 2007 *J. Mathematics in Practice and Theory* 37 65-72.
[15] State Environmental Protection Administration, State Administration of Quality Supervision and Inspection and Quarantine 2002 *GB3838-2002 Surface water environmental quality standards* (Beijing: China Environmental Science Press)
[16] Wang P Z 1989 *Application of fuzzy mathematics* (Beijing: Beijing University of Economics Press)
[17] Sun J N, Zou Z H and Ren G P 2005 *J. Techniques and Equipment for Environmental Pollution Control* 6 45-48.