

Application of Set Pair Analysis Model Based on Interval Number in Water Quality Evaluation

SI Jundong¹, LU Mian², WU Zhijun³, LI Fanxiu*¹

1.College of Chemistry and Environmental Engineering, Yangtze University, Jingzhou 434023,China;

2.Realhom Appraisal & Consulting Co.ltd, Wuhan 430062,China;

3.Shashi branch of Jingzhou Environmental Protection Bureau, Jingzhou 434000, China

Abstract :

Aiming at the problem of losing information when the monitoring value is averaged during the water quality evaluation process, the method of converting the interval number into the contact number is proposed, and the interval number model of the set pair analysis number is established. The model was applied to the Fen he River for water quality evaluation. The evaluation results were basically consistent with the results of the interval clustering method, and the value of i had no effect on the water quality evaluation results when $[0,1]$. The model calculation process in this paper is simple and fast, and the evaluation results are feasible and effective, which provides a scientific basis for water environment management.

Key words : Water Quality Evaluation ; Set Pair Analysis ; Interval Number ; Connection Number;

Introduction

In recent years, water pollution has become a worldwide pollution problem. The pollution of the water environment is bound to affect the biological integrity of aquatic system, affect the water quality, and directly or indirectly affect human health. Water quality assessment is an important prerequisite for controlling water environment pollution. By establishing a reasonable water quality assessment model, objective assessment on the pollution status of water environment can provide a decision-making basis for water environment management and pollution prevention.

At present, the comprehensive evaluation of water quality mainly includes attribute recognition method, fuzzy comprehensive evaluation method, gray clustering method, matter element analysis method, and artificial neural network method, etc. (Zhang et al.2013,Li 2005,Zhou et al. 2007,Liu et al. 2009,Xie et al.2018 and Hao et al. 2016). Although these methods have achieved good results in the water quality evaluation process, there are also some problems. For example, in application, most existing models can only deal with the data determined by evaluation indexes, but cannot effectively deal with the problem that evaluation indexes are interval numbers. When the water quality is evaluated in a river, lake, and reservoir, etc., the average value of the monitored values is usually calculated in a certain period of time. In the process of averaging the monitoring values, a lot of useful information is lost invisibly, which cannot accurately reflect the situation of water pollution. In order to make full use of more monitoring information, Luo et al.(2010) defined the grade standard value and the actual monitoring value as the interval number,

and used the interval number sorting method to evaluate the water quality of the Dongling Bridge in Hunhe River, and the results were basically consistent with the results of the fuzzy comprehensive evaluation method. Zhou et al. (2013) reduced the data dimension of multiple monitoring values into interval numbers, and used the entropy weight as the weight to evaluate the pollution status of Taihu Lake, which proved the feasibility and effectiveness of the interval water quality evaluation model.

Since the water environment is a comprehensive and complex system, its evaluation factors change dynamically in time and space, and there is a fuzzy, nonlinear and uncertain relationship between each evaluation factor and the evaluation grade, so the water quality evaluation is a complex, changeable and uncertain process. Sampling and testing in the process of water quality monitoring also have some random errors, because these processes often cannot be given in the form of fixed data point values and show the expression form of interval number, which increases the uncertainty of water quality evaluation. So far, there have been few studies on such interval numbers at home and abroad. In this paper, the fluctuation of water quality was represented by the interval number. By means of the advantage of set pair analysis in dealing with uncertainty, the interval number was converted into the number of connection. According to the classification standard of water quality, which grade of a certain measured water environment quality is closest to the evaluation standard. Therefore, a water quality evaluation model based on interval number attribute correlation degree is established, and the feasibility and practicability of the model are verified by case analysis.

1 Set Pair Analysis Model Based on Interval Number

1.1 Set pair analysis principle

Set pair analysis is a new method for the determination of uncertainty theory proposed by Chinese scholar Zhao Keqin in 1989. The core idea is to make mathematical processing for determining uncertainty as a system (determining uncertain systems), and dealing with complex and variable things from the same degree, degree of difference, as well as degree of opposition, and expressing the results of mathematical processing as the same Anti-contact, also known as contact. The connection expression is: $u = a + bi + cj$

Equation above is called identity-discrepancy-contrary connection degree or three-element connection number, where a represents the identical degree of the set pair, b represent discrepancy degree of the set pair, and c is the contrary degree of the set pair. i is the coefficients of the discrepancy degree, and is some uncertain values between -1 and 1 , and is an uncertain value. j is the coefficient of the contrary degree, and is specified as -1 and a, b, c satisfy $a+b+c=1$.

Set the number of connections for set pair $u_1=a_1+b_1i, u_2=a_2+b_2i$, then the number algorithm is as follows (Lu et al. 2009):

$$(1) u_1+u_2=(a_1+b_1)+(a_2+b_2)i$$

$$(2) u_1-u_2=(a_1-b_1)+(a_2+b_2)i$$

$$(3) u_1 \times u_2 = A + Bi, A = a_1 a_2, B = a_1 b_2 + a_2 b_1 + b_1 b_2$$

1.2 Interval number and its operation (Jiang et al. 2015)

Let a and b be two real numbers, and if $a < b$, then $[a, b]$ is the interval number, a is the lower bound of the interval number, b is the upper bound of the interval number, and the normal interval number is denoted as $[a^-, a^+]$. The interval number calculation rule is as follows:

$$(1) a+b = [a^-+b^-, a^++b^+]$$

$$(2) a-b = [a^-, b^-, a^+, b^+]$$

$$(3) \lambda a = [\lambda a^-, \lambda a^+], \text{ among them } \lambda \geq 0, \text{ if } \lambda = 0, \text{ then } \lambda a = 0$$

$$(4) \text{ If and only if } a^- = b^-, b^- = b^+, \text{ then } a = b$$

1.3 Interval number converted to contact number (Liu et al.2010)

For the interval number $[a^-, a^+]$ can be expressed as $x \pm \Delta x$, among them $x = \frac{a^- + a^+}{2}, \Delta x = \frac{a^+ - a^-}{2}$, x is the expected value of the interval number, and Δx is the maximum difference value of the interval number.

$$\text{Let } a = x, b = \Delta x, \text{ then the number of contacts } u = a + bi = \frac{a^- + a^+}{2} + \frac{a^+ - a^-}{2} i \quad (1)$$

1.4 Judgment criteria

According to the definition of the number of intervals, the standard value of Environmental Quality Standard for Surface Water (GB3838-2002) can be defined as the interval number. The definition method is as follows:

Let the standard value of factor j in the water quality standard of i ($i = 1, 2, 3, \dots, n$) be defined as the interval number x_{ij} , and the standard value of factor i of the original water quality standard is $c_{(i-1)j}$, then the standard value of class i is c_{ij} . For the benefit index (the bigger the better), such as DO, the interval number of the grade standard is $x_{ij} = [c_{ij}, c_{(i-1)j}]$, when $i = 1, x_{1j} = [c_{nj}, c_{(n-1)j}]$. For the cost indicator (the smaller the better), such as BOD, the interval number of the grade standard is $x_{ij} = [c_{(i-1)j}, c_{ij}]$, when $i = 1, x_{1j} = [0, c_{1j}]$. When the measured data defines the number of intervals, the minimum value is recorded as the lower limit of the number of intervals, and the maximum value is recorded as the upper limit of the number of intervals.

According to the above method, the five standard values in Environmental Quality Standard for Surface Water (GB3838-2002) are defined as the interval number x_1 (grade 1), x_2 (grade 2), x_3 (grade 3), x_4 (grade 4), and x_5 (grade 5). The multiple measured values of the water environment to be evaluated are recorded as the interval number x_k ($k = 1, 2, \dots, n$).

In the process of water quality evaluation, it is usually based on which grade of water quality monitoring value is closest to the assessment criteria, that is, the water quality is determined to be several grades of water quality. According to the principle of closeness, the Euclidean distance is used as the criterion for judging, that is, the Euclidean distance between the sample to be graded and the i -th grade water quality standard is the smallest, and the sample to be graded belongs to the i -th grade water quality.

Use formula (1) to convert the number of grade standard intervals x_i ($i=1,2,3,4,5$) and the number of sample intervals to be graded x_k ($k=1,2,\dots,n$) into interval numbers.

From the definition of Euclidean distance, the Euclidean distance formula for determining the k -th criterion to be graded and the j -th evaluation factor of the i -th rating criterion is as follows

$$d_{kj} = \sqrt{(u_{ij} - u_{kj})^2} = \sqrt{A^2 + (2AB + B^2)}i \quad (2)$$

Algorithm according to the number of contacts, $u_{ij} - u_{kj} = |a_{ij} - a_{kj}| + (b_{ij} + b_{kj})i = A + Bi$, so $|a_{ij} - a_{kj}| = A, b_{ij} + b_{kj} = B$

1.5 Water quality grading steps

1.5.1 Interval number normalization

Because the units of the evaluation indicators are different, which is not conducive to comparison, it is necessary to standardize the number of evaluation standard intervals and the actual monitoring value intervals. At present, there are many methods for data normalization, and the methods for normalizing different index types are different. For the water quality evaluation indicators, the following standardized treatment methods are adopted [12]:

$$x_{ij} = \left(\frac{c_{ij}^-}{\max x_{ij}}, \frac{c_{ij}^+}{\max x_{ij}} \right) (\text{Cost index}) \quad (3)$$

$$x_{ij} = \left(\frac{\min x_{ij}}{c_{ij}^-}, \frac{\min x_{ij}}{c_{ij}^+} \right) (\text{Benefit index}) \quad (4)$$

In the formula, $\max x_{ij}$ represents the maximum value of the same magnitude corresponding to each evaluation factor in the water quality evaluation standard, and $\min x_{ij}$ represents the minimum value of the same magnitude corresponding to each evaluation factor in the water quality evaluation standard.

Standardize the cost indicator using formula (3), and use the formula (4) for standardization.

1.5.2 Interval number converted to contact number

Converting the number of intervals after normalization into the corresponding a+bi type relation number expression according to equation (1)

1.5.3 Classification method

According to the comparison rule of the a+bi type of contact number, the equation (2) is used to calculate the Euclidean distance d_{kj} of the j-th evaluation factor of the k-th sample to be classified and the i-th rating standard, and the k-th sample to be classified and the first The Euclidean distance calculation formula for the i-grade rating standard is as follows:

$$d_{ki} = \sum d_{kj} \quad (5)$$

By comparing the magnitude of the Euclidean distance of each grade, the category corresponding to the minimum value of the Euclidean distance is the evaluated water environment quality grade, so $d_k = \min[d_{k1}, d_{k2}, d_{k3}, d_{k4}, d_{k5}]$ ($k=1, 2, \dots, 5$)

Where d_{k1} , d_{k2} , d_{k3} , d_{k4} , and d_{k5} respectively represent the Euclidean distance between the kth sample to be graded and the standard values of the first grade, the second grade, the third grade, the fourth grade, and the fifth grade.

2. Case Study

In this paper, Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), Hydroxybenzene(HOB), Cyanide (CN-), Mercury (Hg), Arsenic (As), Chromium (Cr^{6+}) were selected as the evaluation index, and surface water quality standard (GB 3838-2002) is adopted to divide the water environment into grade 1 (excellent), grade 2 (good), grade 3 (light pollution), grade 4 (moderate pollution) and grade 5 (heavy pollution). The Classification of Environmental Quality Standard for Surface Water (GB3838-2002) are shown in Table 1.

To examine the feasibility and effectiveness of the model, the suggested method has been applied to water quality assessment. In this paper, Fenhe water quality monitoring data (sampling section I, sampling section

II, and sampling section III), the minimum observation and maximum of three cross sections were used. The observations were used as evaluation samples, and the monitoring data are shown in Table 2.

Table 1 Surface Water Environmental Quality Standards

Indices	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
DO (mg/L)	[9,10]	[6,9]	[5,6]	[3,5]	[0,3]
BOD (mg/L)	[0,3]	[3,4]	[4,5]	[5,6]	[6,10]
HOB (mg/L)	[0,0.001]	[0.001,0.002]	[0.002,0.005]	[0.005,0.01]	[0.01,0.1]
CN ⁻ (mg/L)	[0,0.005]	[0.005,0.05]	[0.05,0.2]	[0.05,0.2]	[0.05,0.2]
Hg (mg/L)	[0,0.00005]	[0,0.00005]	[0.00005,0.0001]	[0.0001,0.001]	[0.0001,0.001]
As (mg/L)	[0,0.05]	[0,0.05]	[0,0.05]	[0.05,0.1]	[0.05,0.1]
Cr ⁶⁺ (mg/L)	[0,0.01]	[0.01,0.05]	[0.01,0.05]	[0.01,0.05]	[0.05,0.1]

Table 2 Water Quality Monitoring Data of Fenhe River

Indices	Sampling section I	Sampling section II	Sampling section III
DO (mg/L)	[9.1,9.5]	[8.2,8.8]	[6.4,6.7]
BOD (mg/L)	[1.0,1.5]	[3.9,4.45]	[7.0,7.5]
HOB (mg/L)	[0.0008,0.003]	[0.006,0.008]	[0.01,0.2]
CN ⁻ (mg/L)	[0.0009,0.0015]	[0.002,0.004]	[0.008,0.009]
Hg (mg/L)	[0.0007,0.00089]	[0.0005,0.0007]	[0.00045,0.00055]
As (mg/L)	[0.002,0.004]	[0.068,0.076]	[0.038,0.045]
Cr ⁶⁺ (mg/L)	[0.0055,0.0068]	[0.01,0.02]	[0.00065,0.00078]

The specific data processing and calculation steps are as follows:

(1) Normalize the number of water quality assessment standard intervals according to formula (3), where DO is the benefit type index. Before the calculation, the number of DO intervals is reciprocated and then normalized using equation (3). The other six evaluation factors are cost indicators directly normalized using equation (3). The number of grading standard intervals after normalization is shown in Table 3. The actual monitoring interval after normalization is shown in Table 4.

(2) Convert the number of intervals after normalization into the number of connection using the formula (1), and the number of grading standards is shown in Table 5 while the monitored values are shown in Table 6.

(3) Calculate the Euclidean distance between the monitoring values of each monitoring section and each evaluation using equations (2) and (5). When calculating i take 0.5, the calculated results were shown in Table 7.

Table 3 The Interval Number After Normalization Processing of Evaluation Standards

Indices	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
DO (mg/L)	[0,0.3]	[0.3,0.34]	[0.34,0.51]	[0.51,0.61]	[0.61,1]
BOD (mg/L)	[0,0.3]	[0.3,0.4]	[0.4,0.5]	[0.5,0.6]	[0.6,1]
HOB (mg/L)	[0,0.005]	[0.005,0.01]	[0.01,0.025]	[0.025,0.05]	[0.05,0.5]
CN ⁻ (mg/L)	[0,0.025]	[0.025,0.25]	[0.25,1]	[0.25,1]	[0.25,1]
Hg (mg/L)	[0,0.05]	[0,0.05]	[0.05,0.1]	[0.1,1]	[0.1,1]
As (mg/L)	[0,0.5]	[0,0.5]	[0,0.5]	[0.5,1]	[0.5,1]
Cr ⁶⁺ (mg/L)	[0,0.1]	[0.1,0.5]	[0.1,0.5]	[0.1,0.5]	[0.5,1]

Table 4 The Interval Number After Normalization of Measured Values

Indices	Sampling section I	Sampling section II	Sampling section III
DO (mg/L)	[0.32,0.33]	[0.34,0.37]	[0.45,0.47]
BOD (mg/L)	[0.1,0.15]	[0.39,0.445]	[0.7,0.75]
HOB (mg/L)	[0.004,0.015]	[0.03,0.04]	[0.05,1]
CN ⁻ (mg/L)	[0.0045,0.0075]	[0.01,0.02]	[0.04,0.045]
Hg (mg/L)	[0.7,0.89]	[0.5,0.7]	[0.45,0.55]
As (mg/L)	[0.02,0.04]	[0.68,0.76]	[0.38,0.45]
Cr ⁶⁺ (mg/L)	[0.055,0.068]	[0.1,0.2]	[0.0065,0.0078]

Table 5 Connection Numbers of Evaluation Criteria

Indices	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
DO (mg/L)	0.15+0.15i	0.32+0.02i	0.425+0.085i	0.56+0.05i	0.805+0.195i

BOD (mg/L)	0.15+0.15i	0.35+0.05i	0.45+0.05i	0.55+0.05i	0.8+0.2i
HOB (mg/L)	0.0025+0.0025i	0.0075+0.0025i	0.0175+0.0075i	0.0375+0.0125i	0.275+0.225i
CN ⁻ (mg/L)	0.0125+0.0125i	0.1375+0.1125i	0.625+0.375i	0.625+0.375i	0.625+0.375i
Hg (mg/L)	0.025+0.025i	0.025+0.025i	0.075+0.025i	0.55+0.45i	0.55+0.45i
As (mg/L)	0.25+0.25i	0.25+0.25i	0.25+0.25i	0.75+0.25i	0.75+0.25i
Cr ⁶⁺ (mg/L)	0.05+0.05i	0.3+0.2i	0.3+0.2i	0.3+0.2i	0.75+0.25i

Table 6 Contact Number of Fenhe Monitoring Data

Indices	Sampling section I	Sampling section II	Sampling section III
DO (mg/L)	0.325+0.005i	0.355+0.015i	0.46+0.01i
BOD (mg/L)	0.125+0.025i	0.4175+0.0275i	0.725+0.025i
HOB (mg/L)	0.0095+0.0055i	0.035+0.005i	0.525+0.475i
CN ⁻ (mg/L)	0.006+0.0015i	0.015+0.005i	0.0425+0.0025i
Hg (mg/L)	0.795+0.095i	0.6+0.1i	0.5+0.005i
As (mg/L)	0.03+0.01i	0.72+0.04i	0.415+0.035i
Cr ⁶⁺ (mg/L)	0.0615+0.0065i	0.15+0.05i	0.00715+0.00065i

Table 7 Fenhe Water Quality Evaluation Results

Sampling section	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade	The interval clustering method ^[12]
I	0.46	0.65	0.86	1.11	1.80	1	1
II	0.75	0.57	0.77	0.54	1.10	4	2
III	1.12	0.77	0.93	0.93	1.27	2	2

In order to determine whether the value of i has any influence on the evaluation result, based on the monitoring values of the three sampling segments, the values of i are respectively -0.5, 0, and 0.5, and the results are compared. The evaluation results are shown in Fig. 1.

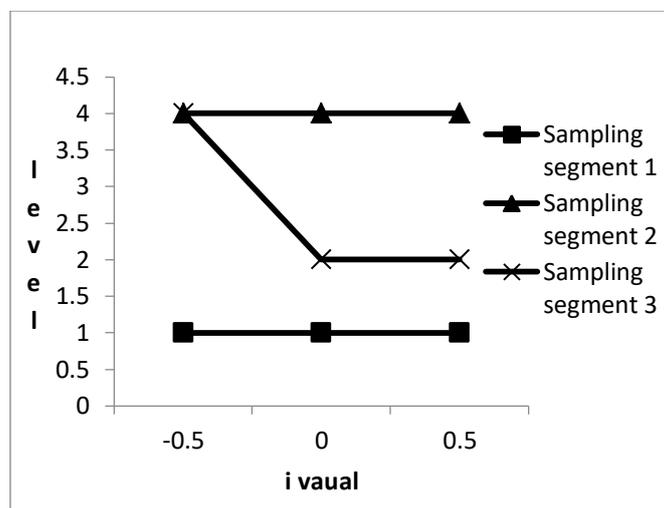


Fig.1 The Effect of i Value on the Evaluation Results

From the comparison between Table 7 and the interval clustering method [12], it is found that the results of the evaluation model proposed in this paper are basically consistent with the results of the interval clustering method except for the sampling segment 2. For the sampling segment 2, the interval clustering method evaluates to grade 2, and the evaluation result proposed in this paper is grade 4. Comparing the original data of water quality with the evaluation criteria, it is found that the DO content of dissolved oxygen in sample 2 is at grade 2, BOD is at grade 2~3, HOB is at grade 4, CN- is at grade 1, Hg is at grade 4~5, As is at 4 ~5 grade, and Cr6+ is in 2~4 grade. Among the seven evaluation indexes, 6 indicators are higher than 2, and the interval clustering method evaluates the water quality of the sampling section II to grade 2, which is obviously inconsistent with the actual situation. The comparison of the evaluation results of the two water quality assessment methods illustrates the feasibility and superiority of the proposed modle in the paper. The interval number model of the set pair analysis is reasonable, objective and has good use value.

It can be seen from Fig. 1 that when i takes -0.5, 0, and 0.5 respectively, there is no influence on the evaluation results of the sampling segment 1 and the sampling segment 2. For the sampling segment 2, when the value of i is -0.5, the evaluation result is different from the value of i, which is 0, 0.5, proving that it is reasonable to take $i=0.5$ when the set pair analysis operation is performed.

3. Conclusion

(1) Since the data of water quality monitoring is usually an interval number, the method of averaging is mostly adopted in the traditional evaluation methods in the application, which leads to the loss of much effective spatio-temporal information, bringing uncertainty to water quality evaluation. In this paper, an interval number evaluation model based on the set pair analysis number was proposed, and the water quality grade was evaluated by calculating the Euclidean distance between the number of connection to be evaluated and the water environment quality standard. The application of the case shows that the proposed evaluation method is simple in calculation, high in accuracy, and can reflect the pollution actual status of water environment quality. It is a practical evaluation method.

(2) By evaluating the water quality of the three sampling sections of the Fenhe River, it is found that the value of i has no effect on the evaluation results when $[0,1]$, indicating that it is feasible that the value of i is 0.5 when water quality is evaluated in set pair analysis.

References

- [1] Zhang X J, Xu Z G, Song X Y, et al.(2013) Application of several water environmental quality assessment methods in Qinghai Lake into lakes. *Environmental Engineering* 31(01):117-121.
- [2] Li R Z (2005) Research Progress and Trend Analysis of Theoretical Models of Water Quality Evaluation. *Journal of Hefei University of Technology (Natural Science)* 28(04):369-373.
- [3] Zhou L F, Xu S G, Sun W G(2007) Comprehensive Evaluation of Water Environment Quality of Zhalong Wetland Based on Grey Clustering Method. *Journal of Dalian University of Technology* 47(02):240-245.
- [4] Liu L, Zhou J K, An X L, et al. (2009) Using fuzzy theory and information entropy for water quality assessment in Three Gorges region, China. *Experts Systems with Applications* 37(3):2517-2521.
- [5] Xie C, Li Z J, Wang B, et al.(2018)Application of Improved Comprehensive Indexing Index Method in Groundwater—Taking Zhaochuan Basin as an Example. *Water Saving Irrigation* (09):56-61.
- [6] Hao Y, Liang X J, Meng F N, et al.(2016) Fuzzy Comprehensive Evaluation of Groundwater Quality Based on AHP-Entropy Weight Method—Taking Baicheng City as an Example. *Water Saving Irrigation* (06):51-54+58.
- [7] Luo L G, Pan J, An Z W, et al.(2010) Study on water environment quality assessment method based on interval number ordering. *Environmental Science and Management* 35(02):176-179.
- [8] Zhou R X, Shan X T, Yang J ,et al.(2013) Application of interval multi-attribute decision making method based on entropy weight in lake water quality assessment. *Journal of environmental science* 33(03):910-917.
- [9] Lu Y, Wang X(2009)The Operation of Binary Connection Number and Its Application .*Journal of Bohai University(Natural Science Edition)* 30(04):343-345.
- [10] Jiang J, Li L X, Feng Y G(2015) A Comprehensive Classification Method of Highway Tunnel Surrounding Rock Based on Set Pair Analysis Number and Its Application. *China Foreign Highway* 35(04):235-239.
- [11] Liu X M, Zhao K Q(2010) Interval Number Multiple Attribute Decision Making Based on Uncertainty Analysis of Connection Numbers. *Fuzzy Systems and Mathematics* 24(05):141-148.
- [12] Wong H, Hu B Q(2013) Application of interval clustering approach to water quality evaluation. *Journal of Hydrology*, 491:1-12.