



Application of Set Pair Analysis Method Based on Entropy Weight in Groundwater Quality Assessment - A Case Study in Dongsheng City, Northwest China

LI PEI-YUE, QIAN HUI and WU JIAN-HUA

School of Environmental Science and Engineering
Chang'an University, Xi'an, Shaanxi, 710054, China

lipy2@163.com

Received 5 July 2010; Accepted 2 September 2010

Abstract: Groundwater quality assessment is an essential study which plays important roles in the rational development and utilization of groundwater. Groundwater quality greatly influences the health of local people. However, most traditional water quality comprehensive assessment methods which have complicated formulas are difficult to apply in water quality assessment. In this paper, a novel method for groundwater quality assessment called set pair analysis was introduced and entropy weight was assigned to each index to improve the assessment model. The calculation steps are depicted in the paper and take groundwater quality assessment in Dongsheng City as a case study. The assessment results indicated that groundwater qualities in the study area were relatively good, Set Pair Analysis method, which was an optimal method for groundwater quality assessment and worth promoting, was easy to use and calculation processes which use almost all the relative information were simple, results were reasonable, reliable and intuitive.

Keywords: Set pair analysis, Groundwater quality assessment, Entropy weight, Groundwater quality

Introduction

Groundwater is vital to local people in arid and semi-arid area and nearly 70% of the water supply for human consumption is often sourced from groundwater in Northwest China. Whether groundwater is suitable for drinking is an essential question that needs discussion. Till now, many groundwater quality assessment methods have been proposed by many scholars such as water quality index (WQI) method¹, fuzzy mathematics method², osculating value method³, grey cluster method⁴ and artificial neural network method⁵. However, water quality assessment is a combination of certainty of evaluation criteria and uncertainty of analysis results. Due to the complexity and randomness of the evaluation factors and the

nonlinear relationship between water quality criteria and the evaluation indices, it is difficult to establish a unified evaluation model, which promotes the further study on the water quality assessment methods.

Set Pair Analysis (SPA) was proposed by Zhao⁶, a Chinese scholar, in 1989. It is a modified uncertainty theory considering both certainties and uncertainties as an integrated certain-uncertain system and depicting the certainty and uncertainty systematically from three aspects as identity, discrepancy and contrary⁶⁻⁸. In the SPA theory, two relative sets in an uncertainty system are constructed and connection degree of the set pair can be established according to the identity, discrepancy and contrary. Based on the connection degree formula, a series of SPA-based researches have been conducted. Su *et al.*⁹ employed SPA to assess the urban ecosystem health level and concluded that SPA can serve as an effective relative-measure to compare different ecosystem health levels of urban ecosystems. Zhou¹⁰ used a SPA based fuzzy assessment method in the real-time risk assessment. Gao and Chen¹¹ introduced SPA in risk ranking and found that the approach was very convenient to operate and the ranking result was more comprehensive. Wang *et al.*¹² introduced SPA into water resources system assessment and studied two cases. They concluded that the SPA was simple in concept, convenient to calculate and feasible for application.

As can be seen from above information, SPA has been introduced into many fields and shown its excellence both in concept and application. However, SPA is a newly proposed theory and needs further in-depth study. Since it was introduced into water system, it has drawn much attention. In the previous studies, the average connection degree of the set pair was the arithmetic mean of every indicator, which is not appropriate because of the different importance of the different indicators. In this paper, some improvements were made to the original SPA method by introducing information entropy theory to determine the weights of evaluation indexes. This will surely provide a new way of thinking and methods for the evaluation of groundwater quality.

The principle of SPA

The basis of SPA is set pair and its key is connection degree. Suppose there are two sets, A and B. Put them together to form a set pair H (A, B). To investigate the relationship of H (A, B) and judge its good or bad degree, the connection degree is defined as

$$\mu = \frac{S}{N} + \frac{F}{N}i + \frac{P}{N}j = a + bi + cj \quad (1)$$

Where S is the number of the identical terms of characteristic; F is the number of the discrepant terms of characteristic; P is the number of contradictory terms of characteristic. Here, $a=S/N$, $b=F/N$ and $c=P/N$, respectively, represent the identity degree, discrepancy degree and contrary degree of these two sets and we can know that in the formula $a+b+c=1$. In the formula, the term i is the uncertainty coefficient of discrepancy which has different values ranging in $[-1, 1]$ in different conditions; j is the uncertainty coefficient of contradictory which is specified as -1 .

Equation (1) can not only reflect the relationship of overall structure of sets A and B, but also a , b and c can reflect the relationship of internal subtle structure of sets A and B. The connection degree has overcome the drawbacks of such traditional forms of relationship as correlation coefficient, subordinate degree or grey correlation degree, in which there is a

single index only¹². The SPA can not only show the relationship structure clearly, but also can reveal quantitatively three or more characteristics in complex relationship as well as can give the changeable value of comprehensive relationship, which may depend on required different standards or properly selected i . Due to its advantages as a simple mathematic depiction which has clear physical meaning, the SPA has been introduced into various fields.

SPA in groundwater quality assessment

The essence of assessment is the comparison of similarity between two objects in characteristics. If they are similar, they will be classified into one class, if not, they will be classified into different classes. Water quality comprehensive assessment is made based on similarity between monitoring sample and water quality standards. In order to make water quality assessment based on SPA method, a set pair must be made by putting the concentration of every index and water quality standards together first. Then, the connection degree between every index in every monitoring sample and standards must be calculated. Then the average connection degree of every monitoring sample can be obtained. Water quality rank will be decided according to the average connection degree. According to the Groundwater Quality Standards of China, groundwater quality can be classified into 5 ranks. Rank I (excellent quality water), rank II (good quality water), rank III (Medium or average quality water), rank IV (poor quality water) and rank V (extremely poor water). The groundwater quality ranking standards are shown in Table 1.

Table 1. Groundwater quality ranking standards

Indexes	I	II	III	IV	V
Cl ⁻	≤50	≤150	≤250	≤350	>350
SO ₄ ²⁻	≤50	≤150	≤250	≤350	>350
TH	≤150	≤300	≤450	≤550	>550
TDS	≤300	≤500	≤1000	≤2000	>2000

Note: Unites in the column are mg·L⁻¹

Connection degree plays a key role in water quality comprehensive assessment based on SPA, which can be calculated by following formulas.

(1) Connection degree between index j and standard of Rank I

$$\mu_{j1} = \begin{cases} 1 & x \in [0, S_{j(1)}] \\ 1 + \frac{2(x - S_{j(1)})}{S_{j(1)} - S_{j(2)}} & x \in [S_{j(1)}, S_{j(2)}] \\ -1 & x \in [S_{j(2)}, +\infty] \end{cases} \quad (2)$$

(2) Connection degree between index j and standard of Rank II

$$\mu_{j2} = \begin{cases} 1 + \frac{2(x - S_{j(1)})}{S_{j(1)} - 0} & x \in [0, S_{j(1)}] \\ 1 & x \in [S_{j(1)}, S_{j(2)}] \\ 1 + \frac{2(x - S_{j(2)})}{S_{j(2)} - S_{j(3)}} & x \in [S_{j(2)}, S_{j(3)}] \\ -1 & x \in [S_{j(3)}, +\infty] \end{cases} \quad (3)$$

(3) Connection degree between index j and standard of Rank III

$$\mu_{j3} = \begin{cases} 1 + \frac{2(x - S_{j(2)})}{S_{j(2)} - S_{j(1)}} & x \in [S_{j(1)}, S_{j(2)}] \\ 1 & x \in [S_{j(2)}, S_{j(3)}] \\ 1 + \frac{2(x - S_{j(3)})}{S_{j(3)} - S_{j(4)}} & x \in [S_{j(3)}, S_{j(4)}] \\ -1 & x \in [S_{j(4)}, +\infty] \text{ or } [0, S_{j(1)}] \end{cases} \quad (4)$$

(4) Connection degree between index j and standard of Rank IV

$$\mu_{j4} = \begin{cases} 1 + \frac{2(x - S_{j(3)})}{S_{j(3)} - S_{j(2)}} & x \in [S_{j(2)}, S_{j(3)}] \\ 1 & x \in [S_{j(3)}, S_{j(4)}] \\ 1 + \frac{2(x - S_{j(4)})}{S_{j(4)} - S_{j(5)}} & x \in [S_{j(4)}, S_{j(5)}] \\ -1 & x \in [S_{j(4)}, +\infty] \text{ or } [0, S_{j(2)}] \end{cases} \quad (5)$$

(5) Connection degree between index j and standard of Rank V

$$\mu_{j5} = \begin{cases} -1 & x \in [0, S_{j(3)}] \\ 1 + \frac{2(x - S_{j(4)})}{S_{j(4)} - S_{j(3)}} & x \in [S_{j(3)}, S_{j(4)}] \\ 1 & x \in [S_{j(4)}, S_{j(5)}] \end{cases} \quad (6)$$

Where $S_{j(1)}$, $S_{j(2)}$, $S_{j(3)}$, $S_{j(4)}$ and $S_{j(5)}$ are the ranking standards of every rank, x is the sample index value; μ_{j1} , μ_{j2} , μ_{j3} , μ_{j4} and μ_{j5} are the connection degrees between every index and every ranking standard.

When the connection degrees between every index and every ranking standard are calculated, the average connection degree of every sample can be calculated by the following formula.

$$\mu_k = \sum_{j=1}^n \omega_j \times \mu_{jk} \quad (7)$$

Where μ_k is the average connection degree, ω_j is the weight of every index, μ_{jk} the connection degree between every index and every ranking standard. Thus, according to average connection degree, water quality rank can be decided by below formula.

$$\delta_i = \max\{\mu_k\} \quad (8)$$

Where δ_i is the rank which sample i belong to.

Entropy weight

The various indexes have different effects in the assessment system, some are notable and others are weak, therefore, a different weight should be given to each index respectively. The weight can be determined by various methods such as Analytic hierarchy process (AHP), Delphi method and information entropy method.

The concept of information entropy was first proposed by Shannon¹³ in 1948 and it was regarded as the uncertainty of a stochastic event or metric of information content. The steps for calculating entropy weight are described as follows:

Suppose there are m water samples taken to evaluate the water quality ($i=1,2,\dots,m$). Each sample has n evaluated parameters ($j=1,2,\dots,n$). According to real data, eigenvalue matrix X can be constructed:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad (9)$$

In order to eliminate the influence caused by the difference of different units of characteristic indices and different quantity grades, data pretreatment must be put into force¹³.

According to attribution of every index, the feature indexes may be divided into four types: efficiency type, cost type, fixed type and interval type¹³. For the efficiency type, the construction function of normalization is:

$$y_{ij} = \frac{x_{ij} - (x_{ij})_{\min}}{(x_{ij})_{\max} - (x_{ij})_{\min}} \quad (10)$$

While for the cost type, the construction function of normalization is:

$$y_{ij} = \frac{(x_{ij})_{\max} - x_{ij}}{(x_{ij})_{\max} - (x_{ij})_{\min}} \quad (11)$$

After transform, the standard-grade matrix Y can be obtained and shown below:

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix} \quad (12)$$

Then the ratio of index value of the j index in i sample is:

$$P_{ij} = y_{ij} / \sum_{i=1}^m y_{ij} \quad (13)$$

The information entropy is expressed by the formula below:

$$e_j = -\frac{1}{\ln m} \sum_{i=1}^m P_{ij} \ln P_{ij} \quad (14)$$

The smaller the value of e_j is, the bigger the effect of j index. Then the entropy weight can be calculated with the below formula:

$$\omega_j = \frac{1 - e_j}{\sum_{j=1}^n (1 - e_j)} \quad (15)$$

In the formula, ω_j is defined as the entropy weight of j parameter.

Case study

Dongsheng city, a typical semi-arid area, is situated in the south of Erdos Basin, Inner Mongolia, China. In this area, groundwater in the main drinking water resources and groundwater quality is greatly influencing the local people's health and local economy¹⁴. Total 15 groundwater samples were collected in August and September 2007. Samples were collected in pre-cleaned plastic polyethylene bottles for physicochemical analysis. Prior to sampling, all the sampling containers were washed and rinsed thoroughly with the groundwater to be taken for analysis. Each of the groundwater samples was analyzed in the laboratory of

Xi'an Institute of Geology and Minerals Resources using standard procedures recommended by Chinese Ministry of Water Resources. In this paper, Only 4 indices having significant influences on the water quality including chloride, sulphate, total hardness (TH) and total dissolved solid (TDS) were selected for the groundwater quality comprehensive assessment. Sample analysis results are shown in Table 2.

Table 2. Water sample analysis results

Sample ID	Cl ⁻	SO ₄ ²⁻	TH	TDS
W116	46.5	134	132	446
W209	39.6	100	352	672
W117	18.9	52.6	152	333
W120	20.7	47.8	206	398
W211	87.8	148	387	810
W119	48.2	100	327	643
W114	22.4	90.8	267	606
W203	29.3	81.2	218	484
W115	8.6	38.2	173	306
W104	41.3	81.2	323	611
W108	49.9	57.4	178	561
W204	29.3	105	388	746
W107	29.3	81.2	274	530
W201	13.8	57.4	269	538
D1014	60.2	86	189	524

Note: Unites in the column are mg·L⁻¹

Sample D1014 was taken as a computing example. Connection degree between every index and every ranking standard can be calculated from equation 2-6. The results are shown in Table 3.

Table 3. Connection degree between every index and ranking standard in sample D1014

Indexes	Rank				
	I	II	III	IV	V
Cl ⁻	0.796	1	-0.796	-1	-1
SO ₄ ²⁻	0.280	1	-0.280	-1	-1
TH	0.480	1	-0.480	-1	-1
TDS	-1	0.904	1	-0.904	-1

According to formula (9) to formula (15), information entropy and entropy weight of every index can be calculated and the calculated results are listed below in Table 4.

Table 4. Information entropy and entropy weight of parameters

Items	Cl ⁻	SO ₄ ²⁻	TH	TDS
Information entropy e_j	0.93	0.97	0.98	0.99
Entropy weight ω_j	0.53	0.20	0.16	0.11

According to equation 7, average connection degree of sample D1014 can be obtained. The calculated results are 0.45, 0.99, -0.45, -0.99 and -1. As can be seen, the connection degree between sample D 1014 and rank II is 0.99 which is the largest. Hence, the rank of sample D1014 belongs to rank II. The ranks of other samples can be calculated in the same way and the calculated results are listed in Table 5.

Table 5. Results of water quality comprehensive assessment

Sample ID	Connection degree					Rank
	I	II	III	IV	V	
W116	0.51	0.89	-0.51	-1	-1	II
W209	0.27	0.59	-0.27	-0.81	-1	II
W117	0.95	0.34	-0.95	-1	-1	I
W120	0.77	0.36	-0.77	-1	-1	I
W211	-0.33	0.68	0.33	-0.68	-1	II
W119	0.27	0.84	-0.27	-0.88	-1	II
W114	0.37	0.36	-0.37	-0.95	-1	I
W203	0.53	0.56	-0.53	-1	-1	II
W115	0.94	0.02	-0.94	-1	-1	I
W104	0.34	0.72	-0.34	-0.90	-1	II
W108	0.69	0.97	-0.69	-0.97	-1	II
W204	0.25	0.26	-0.25	-0.71	-1	II
W107	0.40	0.54	-0.40	-0.99	-1	II
W201	0.50	0.21	-0.50	-0.98	-1	I
D1014	0.45	0.99	-0.45	-0.99	-1	II

As can be seen from Table 5 that sample W117, W120, W114, W115 and W201 belong to excellent quality water (rank I), while the other 10 samples belong to good quality water (rank II), which indicates that groundwater in the study area can be applied for consumption without any pretreatment. A comparison was made between the assessment results based on SPA method and assessment results based on fuzzy mathematics². The comparison showed that the assessment results based on the two different methods are consistent. The assessment results are also in accordance with the field investigation facts. There are few residents and there is no major industry in and around the study area, which implies that there is no pollution. Seen from the calculation process, SPA can be regarded as a simple method which is easy to use.

Conclusion

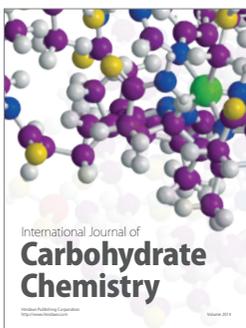
In the paper, a novel method for groundwater quality comprehensive assessment called Set Pair Analysis (SPA) was introduced. A weight based on information entropy was assigned to each index. The steps were depicted in detail and water quality assessment in Dongsheng City was taken as a computing example. The calculated results show that groundwater quality in Dongsheng City belongs to excellent quality (rank I) and good quality (rank II) that is suitable for consumption. SPA method, which is an optimal method for groundwater quality assessment and worth promoting, has simple formulas and is easy to use. The calculation processes use almost all the relative information and entropy weight is assigned to each index, which makes the assessment results reasonable, reliable and intuitive.

Acknowledgment

This research project was supported by the National Natural Science Foundation of China (Grant No. 40772160). Authors also would like to thank the editor and anonymous reviewers for their valuable comments that have greatly improved the quality of the article.

References

- 1 Nemati Varnosfaderany M, Mirghaffary N, Ebrahimi E and Soffianian A, *Water Sci Technol.*, 2009, **60(9)**, 2319-2327.
- 2 Li P Y and Wu S K. *Water Conservancy Science and Technology and Economy*, 2009, **15(9)**, 789-791. (in Chinese)
- 3 Lu B F and Han W P, *Asian Agricultural Research*, 2009, **1(2)**, 1-4.
- 4 Hatvani I G, Kovács J, Kovács I S, Jakusch P and Korponai J, *Ecological Engineering*, 2011, **37(4)**, 629-635
- 5 Singh K P, Basant A, Malik A and Jain G, *Ecological Modelling*, 2009, **220(6)**, 888-895.
- 6 Zhao K Q and Xuan A L, *System Engineering*, 1996, **14(1)**, 18-23, 72.
- 7 Su M R, Yang Z F and Chen B, *Communications in Nonlinear Science and Numerical Simulation*, 2009, **14(14)**, 1773-1780.
- 8 Zhao K Q, *Systems Engineering*, 1994, **12(4)**, 67-72 (in Chinese).
- 9 Su M R, Yang Z F, Chen B and Ulgiati S, *Ecological Modeling*, 2009, **220(18)**, 2341-2348.
- 10 Zhou J F, *Safety Sci.*, 2010, **48(6)**, 819-822.
- 11 Gao F and Chen Y W, *Engineering Sci.*, 2006, **4(1)**, 89-93.
- 12 Wang W S, Jin J L, Ding J and Li Y Q, *Science in China Series E: Technological Sci.*, 2009, **52(10)**, 3017-3023.
- 13 Ding S F and Shi Z Z, *J Information Sci*, 2005, **31(6)**, 497-502.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

