Improvement on Entropy Weighting Model in Groundwater Quality Evaluation

Yan Feng[1, 2],Zhang Xinglei[1], Xiao Chengzhi[1], Tang Bei[1], Liu Chenglin[1,2]

1 School of Civil Engineering and Architecture, Nanchang University, Nanchang, 330031, China;
2 Key Laboratory of Poyang Lake Environment and Resource Utilization (Nanchang University), Ministry of Education, Nanchang, 330031, China
E-mail: yfmilan@163.com

Abstract. Entropy weighting model (EWM) is a widely used weighting method in groundwater quality assessment. EWM assigns weight according to the information amount principle. For the indicator with a higher dispersion degree, a larger weight parameter is assigned; and vice versa. However, through multiple practical applications, we find that the conventional EWM often neglects the indicator with high pollution degree, for it only takes the statistics characteristics of the observation data into consideration, but ignores their practical meaning in aquatic environment management. To solve this problem, this study makes an improvement on the conventional EWM through introducing a pollution degree principle. And then, the improved EWM is applied into the groundwater quality evaluation in Suihua city as an illustration. The results show that: (i) the conventional EWM theory neglects the indicator with high pollution degree, and makes the synthetic assessment result distorted and over-optimistic. (ii) Through introducing a pollution degree principle, the improved EWM highlights the heavily polluted indicators and excludes the “zero weight” phenomenon, which makes the final comprehensive groundwater quality assessment much stricter and more reasonable. (iii) The comprehensive groundwater quality of Suihua city is “Marginal”, and its crucial pollutants are iron, manganese and ammonia nitrogen.

1. Introduction

Entropy weighting model (EWM) is a widely used weighting method in groundwater quality assessment [1]. EWM is designed by sociologists based on statistics and information theory [2]. For the indicator with a higher dispersion degree, which indicates that a larger amount of information is contained, a larger weight parameter is assigned; and vice versa [3].

EWM was firstly introduced into groundwater quality evaluation by Lu in 2010 [4]. And the results showed that EWM assigned weights according to the observation data, and none subjective factor was contained, which avoided the difference in weights parameters introduced by subjective experiences [4]. Since then, EWM has got a widely application in the groundwater management [5]-[8].

However, through multiple practical applications, we find that the weights parameters assigned by the conventional EWM are not always reasonable, for it often neglects the importance of the indicator with high pollution degree, which makes the synthetic assessment result over-optimistic.

To solve this problem, we make an improvement on the current EWM theory in this study. And to test the effectiveness of the improved EWM, it is applied into the groundwater quality evaluation of Suihua city as an illustration.

IOP Conf. Series: Earth and Environmental Science **178** (2018) 012006 doi:10.1088/1755-1315/178/1/012006

2. Materials and Methods

2.1. Study Area

As is illustrated in Fig. 1, Suihua city is an important industrial city and transportation hub in northeast China [9]. The land area of Suihua city is 35211 km^2, and the population is 5.2 million [9]. Groundwater is the major water resource for Suihua city [9]. To afford the industrial, commercial, and residential water demand, approximately 15×10^4 m^3 groundwater is pumped every month [9].

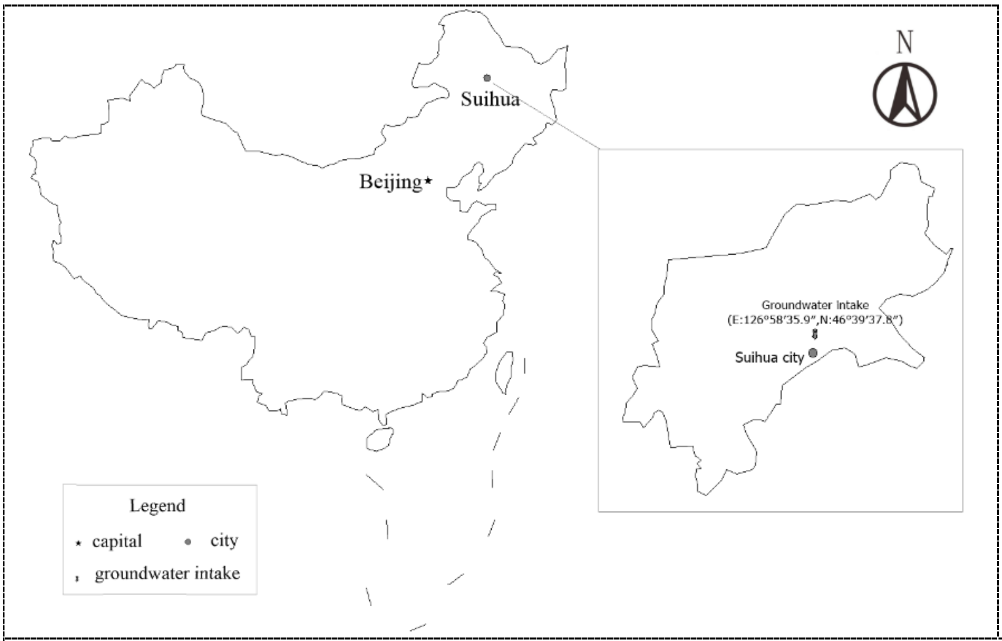

Figure 1. Figure with short caption (caption centred).

To protect the groundwater environment condition, the Ministry of Environmental Protection of China published the National Environmental Quality Standards of Groundwater in 1993. According to this standard, there are 22 indicators should be evaluated in the groundwater quality assessment, which are illustrated in Table 1[10].

Table 1. The assessment indicators and their acceptable domain

indicator	acceptable domain	indicator	acceptable domain
pH*	6.5-8.5	nitrate nitrogen	≤20
total hardness	≤450	nitrite nitrogen	≤0.02
sulfate	≤250	ammonia nitrogen	≤0.2
chloride	≤250	fluoride	≤1
manganese	≤0.1	total cyanide	≤0.05
iron	≤0.3	total mercury	≤0.001
copper	≤1	arsenic	≤0.05
zinc	≤1	selenium	≤0.01
volatile phenol	≤0.002	cadmium	≤0.01
anionic surfactant	≤0.3	chromium (VI)	≤0.05
permanganate index	≤3	lead	≤0.05

*pH is a dimensionless indicator; while the units of the other indicators are mg/L.

To investigate the groundwater quality condition, the Environmental Protection Bureau of Suihua City makes a detailed environmental monitoring at the groundwater intake every month, and publishes these data as water quality monitoring reports since 2017, May [11]. And in this study, all the data are cited from these monthly reports published by the Environmental Protection Bureau of Suihua City.

2.2. Synthetic Evaluation Model

In this study, we use the weighted Canadian Council of Ministers for the Environment Water Quality Index (weighted CCME WQI) model to evaluate the groundwater quality condition of Suihua City. Weighted CCME WQI is the improvement of the conventional CCME WQI model through introducing weights parameters [12]-[14]. Denote the number of indicators as m, the number of monitoring times as n, the threshold of the ith indicator as c_i , and the observation data of the ith indicator at the jth monitoring time as x_{ij}, respectively.

For the indicator the larger the better, if x_{ij} is less than c_i, it is determined to be failed, and the overproof degree e_{ij} is calculated by

$$e_{ij} = \frac{c_i}{x_{ij}} - 1. \tag{1}$$

For the indicator the smaller the better, if x_{ij} is larger than c_i, it is determined to be failed, and the overproof degree e_{ij} is calculated by

$$e_{ij} = \frac{x_{ij}}{c_i} - 1. \tag{2}$$

And then, weighted CCME WQI quantifies the synthetic water quality condition from three factors: scope (F_1), frequency (F_2) and amplitude (F_3), which represent the percentage of failed indicators, the percentage of failed tests, and the average overproof condition, respectively [12]-[14]. Use y_{ij} =1 to represent that x_{ij} is failed, and y_{ij} =0 represent that x_{ij} is failed, respectively. Scope (F_1) is generated by

$$F_1 = \sum_{i=1}^{m} \left(w_i \cdot \max_{j=1,\dots,n} \left\{ y_{ij} \right\} \right) \times 100. \tag{3}$$

In Eq. (3), w_i is the weight of the ith pollutant.

Frequency (F_2) is calculated by

$$F_2 = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} \left(w_i \cdot y_{ij} \right)}{n} \times 100. \tag{4}$$

And amplitude (F_3) is generated by

$$F_3 = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} \left(w_i \cdot e_{ij} \cdot y_{ij} \right)}{\sum_{i=1}^{m} \sum_{j=1}^{n} \left(w_i \cdot e_{ij} \cdot y_{ij} \right) + n} \times 100. \tag{5}$$

All of scope (F_1), frequency (F_2) and amplitude (F_3) are the factors the smaller the better; and their values lie in the domain [0,100]. To make a comprehensive and intuitive representation of the water quality condition, CCME WQI uses the cube root principle to generate the final index [12]-[14]：

$$\text{CCME WQI} = 100 - \frac{\sqrt{F_1^2 + F_2^2 + F_3^2}}{1.732}. \tag{6}$$

According to the values of CCME WQI, the synthetic water quality condition is classified into five categories, which are illustrated in Table 2.

Table 2. Categories of CCME WQI

Category	CCME WQI	Description
Excellent	[95, 100]	The water quality is protected with a virtual absence of threat or impairment. And the aquatic environment condition is very close to pristine levels [12]-[14].
Good	[80, 95)	The water quality is protected with only a minor degree of threat or impairment. And the aquatic environment condition rarely departs from desirable levels [12]-[14].
Fair	[65, 80)	The water quality is usually protected but occasionally threatened or impaired. And the aquatic environment condition sometimes departs from desirable levels [12]-[14].

Marginal	[45, 65)	The water quality is frequently threatened or impaired. And the aquatic environment condition often departs from desirable levels [12]-[14].
Poor	[0, 45)	The water quality is almost always threatened or impaired. And the aquatic environment condition usually departs from desirable levels [12]-[14].

2.3. Entropy Weighting Model (EWM)

EWM is an application of information theory in management science [15], [16]. For the indicator with a higher dispersion degree, which indicates that a larger amount of information is contained, a larger weight is assigned; and vice versa [16]. In EWM, the entropy of the *i*th pollutant is defined as H_i:

$$H_i = -\frac{1}{\ln n}\cdot\sum_{j=1}^{n}\left(\frac{x_{ij}}{\sum_{j=1}^{n}x_{ij}}\ln\frac{x_{ij}}{\sum_{j=1}^{n}x_{ij}}\right). \quad (7)$$

H_i is a measurement of pollutant's information amount, its domain is [0,1]. The higher H_i is, the less its information amount is, and a smaller weight should be assigned. On the other hand, the less H_i is, the higher its information amount is, and a larger weight should be assigned [15], [16]. Based on this principle, the weights parameters are generated according to

$$w_i^* = \frac{1-H_i}{m-\sum_{i=1}^{m}H_i}. \quad (8)$$

The current EWM is established in social sciences. It only takes the statistical characteristics of the observation data into consideration, but neglects its practical meaning. For example, assume that all the monitoring data of chromium (VI) in each month are 1mg/L. Obviously, chromium (VI) is significantly overproof and needs special attention. However, based on the current EWM, its weight is 0, for none discrimination is contained in its observation data. As a result, the indicator with high pollution degree is ignored, which makes the synthetic assessment result over-optimistic.

2.4. Improvement of Entropy Weighting Model

As is introduced in Section 2.3, the current EMW only takes the amount of the pollutant's information into consideration, but neglects its practical significance in environment management. This principle makes the weight parameter cannot represent the pollutant's importance correctly. To solve this problem, this study improves the current EWM theory through introducing a pollution degree weighting principle. Firstly, calculate the pollution degree of each observation data based on Eq. (9)

$$p_{ij} = \begin{cases}(x_{ij}/c_i)^{1/2} & \text{For the indicator the smaller the better}\\ (c_i/x_{ij})^{1/2} & \text{For the indicator the larger the better}\end{cases}. \quad (9)$$

In Eq. (4), p_{ij} is the pollution degree of x_{ij}. $p_{ij}>1$ indicates that x_{ij} is failed; while $p_{ij}\le 1$ indicates that x_{ij} is acceptable. The larger p_{ij} is, the higher its pollution degree is; and vice versa. As a result, the pollution degree weighting principle can be defined as

$$w_i^{**} = \frac{\sum_{j=1}^{n}p_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n}p_{ij}}. \quad (10)$$

In most cases, the weights parameters generated by Eq. (8) and Eq. (10) are different. Therefore we take their average value as the comprehensive weight result.

$$w_i = \frac{w_i^* + w_i^{**}}{2}. \quad (11)$$

Obviously, the improved EWM model assigns weight based on not only the indicator's discriminative degree, but also the indicator's pollution condition. For the indicator with a high discriminative degree and a high pollution degree, a larger weight is assigned; and vice versa.

3. Result and Discussion

3.1. Observation Data

According to the monthly groundwater quality report published by the Environmental Protection Bureau of Suihua City [11], all the observation data are illustrated in Table 3.

Table 3. The observation data of each indicator

Indicator	May	June	July	August	Indicator	May	June	July	August
pH*	7.09	6.91	7.10	7.10	nitrate nitrogen	0.018	0.018	0.018	0.016
total hardness	183	176	186	189	nitrite nitrogen	0.009	0.009	0.009	0.016
sulfate	4.46	4.56	6.40	4.73	ammonia nitrogen	0.466	0.491	0.487	0.473
chloride	3.79	3.86	1.83	5.50	fluoride	0.190	0.245	0.229	0.424
manganese	1.40	1.41	1.34	1.57	total cyanide	0.004	0.004	0.004	0.004
iron	1.34	1.32	1.39	1.46	total mercury	0.00004	0.00004	0.00004	0.00004
copper	0.001	0.001	0.001	0.001	arsenic	0.0039	0.0054	0.0046	0.0053
zinc	0.001	0.001	0.001	0.001	selenium	0.0004	0.0004	0.0003	0.0004
volatile phenol	0.0003	0.0003	0.0003	0.0003	cadmium	0.0001	0.0001	0.0001	0.0001
anionic surfactant	0.05	0.05	0.05	0.05	chromium (VI)	0.004	0.004	0.004	0.004
permanganate index	1.31	1.68	1.56	1.62	lead	0.001	0.001	0.001	0.001

*pH is a dimensionless indicator; while the units of the other indicators are mg/L.

As is illustrated in Table 3, there are 3 failed indicators: iron, manganese, and ammonia nitrogen. All of the other indicators are controlled within the acceptable domain.

In these failed pollutants, manganese is the worst indicator. Its observation data vary from 1.32mg/L to 1.46 mg/L, which is about 13 times larger than its acceptable threshold. Iron is the indicator with the second highest pollution condition, its observation data lie in the domain [1.34mg/L, 1.57mg/L], which is about 4 times larger than its acceptable threshold. Ammonia nitrogen is better than iron, its observation data vary from 0.466mg/L to 0.491 mg/L, which is about 2 times larger than its acceptable threshold.

The high pollutions of manganese and iron are caused by the hydro geochemical environment of the study area. The groundwater of Suihua city is pumped from the Quaternary gravel and sand porous phreatic aquifer [9]. The primary minerals in this aquifer contain large amount iron and manganese [9]. However, the pollution of ammonia nitrogen is introduced by human factors. According to the research of Li, most of the livestock farms, poultry farms, and the toilets in countryside are lack of seepage control measures [17]. As a result, the large amount of the residents' and animals' excreta seeps into underground reservoirs in rainy season, which brings high concentration of ammonia nitrogen into the groundwater.

3.2. Study Area

According to the conventional EWM, the weights parameters are generated and listed in Table 4.

Table 4. The weighting parameters generated by the conventional EWM theory

Indicator	Entropy	Weight	Indicator	Entropy	Weight
pH	1.0000	0.0004	nitrate nitrogen	0.9991	0.0067
total hardness	0.9997	0.0019	nitrite nitrogen	0.9737	0.1949
sulfate	0.9915	0.0630	ammonia nitrogen	0.9998	0.0012
chloride	0.9535	0.3451	fluoride	0.9635	0.2709
manganese	0.9987	0.0094	total cyanide	1.0000	0.0000

iron	0.9994	0.0041	total mercury	1.0000	0.0000
copper	1.0000	0.0000	arsenic	0.9943	0.0422
zinc	1.0000	0.0000	selenium	0.9949	0.0375
volatile phenol	1.0000	0.0000	cadmium	1.0000	0.0000
anionic surfactant	1.0000	0.0000	chromium (VI)	1.0000	0.0000
permanganate index	0.9969	0.0229	lead	1.0000	0.0000

As is illustrated in Table 4, according to the conventional EWM, the weights of chloride, fluoride and nitrite nitrogen are much larger than the other 19 indicators, for their dispersion degrees in observation data are significantly higher than the other indicators. The sum weight value of chloride, fluoride and nitrite nitrogen is 0.8109, suggesting that these three indicators have absolute advantages in the comprehensive groundwater quality assessment.

On the other hand, as is introduced in Section 3.1, iron, manganese and ammonia nitrogen are the indicators with the highest pollution conditions. However, their weights parameters are 0.0094, 0.0041 and 0.0012, respectively. In other words, the failed indicators only account for about 1.5% proportions in the synthetic evaluation, which is almost negligible.

Furthermore, all the weights parameters of copper, zinc, volatile phenol, anionic surfactant, total cyanide, total mercury, cadmium, chromium (VI) and lead are zero, for they have none dispersion degree in observation data at all. In other words, these nine pollutants are completely ignored in the comprehensive groundwater quality assessment.

Based on the improved EWM established in Section 2.4, the weights parameters are generated and listed in Table 5.

Table 5. The weighting parameters generated by the improved EWM

Indicator	**Sub-weight 1***	**Sub-weight 2***	**Comprehensive Weight**	**Indicator**	**Sub-weight 1***	**Sub-weight 2***	**Comprehensive Weight**
pH	0.0004	0.0500	0.0252	nitrate nitrogen	0.0067	0.0022	0.0045
total hardness	0.0019	0.0478	0.0248	nitrite nitrogen	0.1949	0.0544	0.1247
sulfate	0.0630	0.0106	0.0368	ammonia nitrogen	0.0012	0.1158	0.0585
chloride	0.3451	0.0090	0.1770	fluoride	0.2709	0.0385	0.1547
manganese	0.0094	0.1633	0.0863	total cyanide	0.0000	0.0212	0.0106
iron	0.0041	0.2776	0.1409	total mercury	0.0000	0.0150	0.0075
copper	0.0000	0.0024	0.0012	arsenic	0.0422	0.0231	0.0326
zinc	0.0000	0.0024	0.0012	selenium	0.0375	0.0145	0.0260
volatile phenol	0.0000	0.0290	0.0145	cadmium	0.0000	0.0075	0.0037
anionic surfactant	0.0000	0.0305	0.0153	chromium (VI)	0.0000	0.0212	0.0106
permanganate index	0.0229	0.0536	0.0383	lead	0.0000	0.0106	0.0053

*Sub-weight 1 is generated by the information amount principle; while sub-weight 2 is generated by the pollution degree principle.

As is illustrated in Table 5, according to the improved EWM, the sum weight value of chloride, fluoride and nitrite nitrogen is 0.4564, which is much smaller than that generated by the conventional

EWM. On the other hand, the sum weight value of iron, manganese and ammonia nitrogen is 0.2857, which is about 20 times larger than that generated by the conventional EWM.

The reason for this phenomenon is that the improved EWM takes the indicator's pollution condition into consideration. As a result, the importance of clean indicators is decreased, while the significance of heavily polluted indicators is highlighted.

Furthermore, through introducing the pollution degree principle, the "zero weight" phenomenon is excluded, too. As is shown in Table 5, the sum weight value of copper, zinc, volatile phenol, anionic surfactant, total cyanide, total mercury, cadmium, chromium (VI) and lead is 0.0698, which indicates that these pollutants cannot be ignored in the comprehensive groundwater quality assessment, though they have none dispersion degree in observation data at all.

Obviously, the major difference between the conventional EWM and the improved EWM is that, the conventional EWM only takes the information amount principle into consideration, while the improved EWM assigns weight based on not only the information amount principle, but also the pollution degree principle.

From the perspective of water resources management, the weight generated by the improved EWM is more reasonable than that generated by the conventional EWM.

(i) The failed indicators, especially the heavily polluted indicators, are the items deserve special attentions in the water resources management. These pollutants are neglected in the conventional EWM, for their dispersion degrees in observation data are small. However, through introducing an auxiliary pollution degree principle, the improved EWM highlights the importance of these indicators.

(ii) The improved EWM solves the "zero weight" phenomenon in the conventional EWM theory. As is discussed above, based on the conventional EWM theory, all the weights parameters of copper, zinc, volatile phenol, anionic surfactant, total cyanide, total mercury, cadmium, chromium (VI) and lead are zero, for they have none dispersion degree in observation data at all. However, these indicators are important toxicological indices in the water resources management; especially total cyanide, cadmium and chromium (VI) have been proved to be carcinogenic. Therefore, these indicators should not be completely neglected in evaluation, though they have none dispersion degree in observation data.

3.3. Comprehensive Evaluation Results

Based on the weight CCME WQI model introduced in Section 2.2, the comprehensive groundwater quality of Suihua city is evaluated and illustrated in Table 6. To make the expression more concise, we use CCME WQI (CEWM) and CCME WQI (IEWM) to represent evaluation results with the weights parameters generated by conventional EWM and improved EWM, respectively.

Table 6. The comprehensive groundwater quality of Suihua city

Evaluation Model	F_1	F_2	F_3	**CCME WQI**	**Category**
CCME WQI (CEWM)	1.46	1.46	8.19	95.12	Excellent
CCME WQI (IEWM)	28.57	28.57	68.81	53.93	Marginal

As is illustrated in Table 6, there is significantly difference between the evaluation results of CCME WQI (CEWM) and CCME WQI (IEWM). According to CCME WQI (CEWM), the comprehensive groundwater quality is "Excellent", indicating that water quality is protected with a virtual absence of threat or impairment However, based on CCME WQI (IEWM), the comprehensive groundwater quality is "Marginal", indicating that the aquatic environment is frequently threatened or impaired.

Obviously, the evaluation result of CCME WQI (CEWM) is over-optimistic. As is illustrated in Table 3, the concentrations of manganese are about 13 times larger than its acceptable threshold; the concentrations of iron are approximately 4 times as large as its acceptable threshold; and the concentrations of ammonia nitrogen are significantly overproof, too. These phenomena prove that the groundwater of Suihua city suffers from frequently and significantly pollutions of manganese, iron and ammonia nitrogen; and its environment often departs from desirable levels. Therefore, compared with CCME WQI (CEWM), the evaluation result of CCME WQI (IEWM) is much more reasonable.

The fundamental cause for the distortion in CCME WQI (CEWM) is that the conventional EWM only takes the statistics dispersion degree of the observation data into consideration, but neglects their

practical meaning. Therefore, the indicators whose monitoring values concentrate in the high pollution domains are easily to be ignored. As is shown in Table 6, for the importance of iron, manganese and ammonia nitrogen is neglected, both of the scope (F_1) and frequency (F_2) in CCME WQI (CEWM) are 1.46, indicating that percentages of failed indicators are failed tests are almost negligible.
However, through introducing an auxiliary pollution degree principle, the improved EWM highlights the importance of heavily polluted indicators. Therefore, both of the scope (F_1) and frequency (F_2) in CCME WQI (IEWM) are 28.57, which are about 20 times larger than these in CCME WQI (CEWM). And as a result, the comprehensive groundwater quality assessment conclusion of CCME WQI (IEWM) is much stricter and more reasonable than CCME WQI (CEWM).

4. Conclusions

(i) The conventional EMW only takes the statistics dispersion degree of the indicator's observation data into consideration, but neglects its practical significance in aquatic environment management. Therefore, the "zero weight" phenomenon is prone to existing, and the indicators with high pollution degrees are easily to be ignored. As a result, the conventional EMW often makes the synthetic groundwater quality evaluation result distorted and over-optimistic.
(ii) The improved EWM assigns weights according to both of the information amount principle and the pollution degree principle. In the improved EWM, the "zero weight" phenomenon is excluded and the importance of heavily polluted indicators is highlighted, which make the final comprehensive groundwater quality assessment much stricter and more reasonable
(iii) The comprehensive groundwater quality of Suihua city is "Marginal", indicating that the aquatic environment is frequently threatened or impaired. The crucial pollutants in the groundwater of Suihua city are iron, manganese and ammonia nitrogen. To improve the groundwater environment, the water resources managers should take more methods to decrease the natural pollution of iron and manganese; and strengthen the seepage control of livestock farms, poultry farms and the toilets in countryside.

5. References

[1] Amiri V, Rezaei M and Sohrabi N 2014 *Environmental Earth Sciences* 72(9): 3479-90
[2] Amiri V, Sohrabi N and Dadgar M 2015 *Environmental Earth Sciences* 74(7): 6163-76
[3] Li P, Wu J and Qian H 2016 *Arabian Journal of Geosciences* 9(1): 1-17
[4] Lu X, Li Y, Lei K, Wang L, Zhai Y and Zhai M 2010 *Environmental Earth Sciences* 60(8): 1693-99
[5] Li P Y, Qian H and Wu J H 2010 *Journal of Chemistry* 2010, 7(S1), 209-16
[6] Wu J, Li P and Qian H 2014 *Journal of Chemistry* 8(2), 787-93
[7] Wu J, Li P and Qian H 2013 *Asian Journal of Chemistry* 25(17): 9795-99
[8] Li P, Wu J and Qian H 2014 *Environmental Geochemistry & Health* 36(4): 693-712
[9] Tang R, Wen ZH, Shu LC, Liu B and Li S 2013 *Water Resources Protection* 29(4): 19-25
[10] Ministry of Environmental Protection of China (http://kjs.mep.gov.cn)
[11] Environmental Protection Bureau of Suihua City (http://www.shhbj.gov.cn).
[12] Yan F, Qiao DY, Qian B, Ma L, Xing XG, Zhang Y and Wang XG 2016 *Journal of Hydrology* 543: 316-23
[13] Abtahi M, Golchinpour N, Yaghmaeian K, Rafiee M, Jahangiri-Rad M and Keyani A 2015 *Ecological Indicators* 53: 283-91
[14] Hurley T, Sadiq R and Mazumder A 2012 *Water Research* 46:3544-52
[15] Huang S, Chang J, Leng G and Huang Q 2015 *Journal of Hydrology* 527: 608-18
[16] Chang J, Li Y, Wang Y and Yuan M 2016 *Journal of Hydrology* 540: 824-34
[17] Li GD 2013 *Heilongjiang Science and Technology of Water Conservancy* 6(41): 15-18

Acknowledgments

This work is supported by the National Natural Science Foundation of China under the contract No. 51709142.