

## Research Article

# Effectiveness of Entropy Weight Method in Decision-Making

Yuxin Zhu,<sup>1,2</sup> Dazuo Tian,<sup>3</sup> and Feng Yan <sup>2</sup>

<sup>1</sup>School of Hydropower & Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

<sup>2</sup>School of Civil Engineering and Architecture, Nanchang University, Nanchang 330031, China

<sup>3</sup>Hunan Institute of Water Resources and Hydropower Research, Changsha 410007, China

Correspondence should be addressed to Feng Yan; yfmilan@163.com

Received 9 October 2019; Accepted 24 February 2020; Published 26 March 2020

Academic Editor: Maria C. Cunha

Copyright © 2020 Yuxin Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Entropy weight method (EWM) is a commonly used weighting method that measures value dispersion in decision-making. The greater the degree of dispersion, the greater the degree of differentiation, and more information can be derived. Meanwhile, higher weight should be given to the index, and vice versa. This study shows that the rationality of the EWM in decision-making is questionable. One example is water source site selection, which is generated by Monte Carlo Simulation. First, too many zero values result in the standardization result of the EWM being prone to distortion. Subsequently, this outcome will lead to immense index weight with low actual differentiation degree. Second, in multi-index decision-making involving classification, the classification degree can accurately reflect the information amount of the index. However, the EWM only considers the numerical discrimination degree of the index and ignores rank discrimination. These two shortcomings indicate that the EWM cannot correctly reflect the importance of the index weight, thus resulting in distorted decision-making results.

## 1. Introduction

The entropy weight method (EWM) is an important information weight model that has been extensively studied and practiced [1, 2]. Compared with various subjective weighting models, the biggest advantage of the EWM is the avoidance of the interference of human factors on the weight of indicators, thus enhancing the objectivity of the comprehensive evaluation results [3, 4]. Therefore, the EWM has been widely used in decision-making in recent years [5–7]. For example, Wu et al. [8] made a comprehensive assessment on lake water quality in Shahu Lake to provide valuable information about present lake water quality for decision-making [8]. Based on the EWM, Zhang and Wang [9] evaluated stress factors and the efficiency of water management measures in the Chongqing city of China [9]. Yu et al. [10] studied the water characteristics of Gucheng Lake, such as eutrophication, health, and spatial distribution by the EWM [10].

The EWM evaluates value by measuring the degree of differentiation. The higher the degree of dispersion of the measured value, the higher the degree of differentiation of

the index, and more information can be derived. Moreover, higher weight should be given to the index, and vice versa. According to the traditional literature, the results of the EWM are always reliable and effective [11, 12]. However, based on engineering practice, we have found that the weighted result of the EWM cannot always accurately reflect the information amount and importance of the index [13]. Subsequently, the decision-making result is distorted [14, 15].

In this study, we consider the site selection of water source as an example and discuss the multi-index decision-making process with the help of Monte Carlo Simulation. The simulation reveals the distortion phenomenon in the calculation of the EWM and its influence on decision-making.

## 2. Methods and Materials

**2.1. EWM.** In this method,  $m$  indicators and  $n$  samples are set in the evaluation, and the measured value of the  $i$ th indicator in the  $j$ th sample is recorded as  $x_{ij}$ .

The first step is the standardization of measured values [16, 17]. The standardized value of the  $i$ th index in the  $j$ th sample is denoted as  $p_{ij}$ , and its calculation method is as follows:

$$p_{ij} = \frac{x_{ij}}{\sum_{j=1}^n x_{ij}}. \quad (1)$$

In the EWM, the entropy value  $E_i$  of the  $i$ th index is defined as [18]

$$E_i = -\frac{\sum_{j=1}^n p_{ij} \cdot \ln p_{ij}}{\ln n}. \quad (2)$$

In the actual evaluation using the EWM,  $p_{ij} \cdot \ln p_{ij} = 0$  is generally set when  $p_{ij} = 0$  for the convenience of calculation.

The range of entropy value  $E_i$  is [0, 1]. The larger the  $E_i$  is, the greater the differentiation degree of index  $i$  is, and more information can be derived. Hence, higher weight should be given to the index. Therefore, in the EWM, the calculation method of weight  $w_i$  is [1, 19]

$$w_i = \frac{1 - E_i}{\sum_{i=1}^m (1 - E_i)}. \quad (3)$$

**2.2. Water Quality Index.** In the decision-making of a water source location, several alternative water sources are selected according to the water quantity. Then, according to the evaluation results for the water quality of each water source, the region with the best water quality will be selected as the final water source.

Water quality index is the most common evaluation model for water quality in water resource management. The representative water quality indexes in the study area are assumed to be permanganate ( $\text{COD}_{\text{Mn}}$ ), ammonia nitrogen ( $\text{NH}_3$ ), and sulfide. According to China's Environmental Quality Standards for Surface Water, the environmental quality of each index is divided into five levels, and the corresponding threshold is illustrated in Table 1.

Given that  $\text{COD}_{\text{Mn}}$ ,  $\text{NH}_3$ , and sulfide are the indicators with smaller and better values than other indicators, their water quality index can be calculated as follows:

$$\text{WQI}_{ij} = \begin{cases} 80 + 20 \cdot \frac{x_{ij} - r_{i1}}{l_{i1} - r_{i1}}, & l_{i1} \leq x_{ij} \leq r_{i1}, \\ 60 + 20 \cdot \frac{x_{ij} - r_{i2}}{l_{i2} - r_{i2}}, & l_{i2} < x_{ij} \leq r_{i2}, \\ 40 + 20 \cdot \frac{x_{ij} - r_{i3}}{l_{i3} - r_{i3}}, & l_{i3} < x_{ij} \leq r_{i3}, \\ 20 + 20 \cdot \frac{x_{ij} - r_{i4}}{l_{i4} - r_{i4}}, & l_{i4} < x_{ij} \leq r_{i4}, \\ 20 \cdot \frac{x_{ij} - r_{i5}}{l_{i5} - r_{i5}}, & l_{i5} < x_{ij} \leq r_{i5}. \end{cases} \quad (4)$$

TABLE 1: Threshold for classification of pollutants (mg/L).

Evaluation index	Excellent	Good	Medium	Bad	Worst
$\text{COD}_{\text{Mn}}$	[0, 2]	(2, 4]	(4, 6]	(6, 10]	(10, 15]
$\text{NH}_3$	[0, 0.15]	(0.15, 0.5]	[0.5, 1]	(1, 1.5]	(1.5, 2]
Sulfide	[0, 0.05]	(0.05, 0.1]	(0.1, 0.2]	(0.2, 0.5]	(0.5, 1]

where  $\text{WQI}_{ij}$  is the water quality index of the  $i$ th pollutant in the  $j$ th water area.  $l_{ik}$  and  $r_{ik}$  are the lower and upper thresholds of the  $i$ th index in the  $k$ th evaluation grade, respectively.

The comprehensive water quality index  $\text{SWQI}_j$  of the  $j$ th water area is defined as [20]

$$\text{SWQI}_j = \sum_{i=1}^m (w_i \cdot \text{WQI}_{ij}). \quad (5)$$

The domain of  $\text{SWQI}_j$  is [0, 100]. When  $100 - 20 \cdot k \leq \text{SWQI}_j < 120 - 20 \cdot k$ , the comprehensive water environment quality of the  $j$ th water area is judged at the  $k$  level.

### 3. Results and Discussion

**3.1. Measured Values.** In this section, we present a set of typical data generated by Monte Carlo Simulation to discuss the inauthenticity of the EWM and its influence on the location of a water source. The measured values generated by Monte Carlo Simulation are shown in Table 2.

According to Table 2, sulfide was the least dispersed, which only concentrated in the narrow range of [0, 0.045 mg/L], and all the samples were excellent. The information content of  $\text{COD}_{\text{Mn}}$  and  $\text{NH}_3$  was higher than that of sulfide, regardless of the dispersion degree of the measured value or the different degree of the grade. Therefore, sulfide should be the least weighted index. The measured values of  $\text{NH}_3$  could divide the five groups of evaluation samples into five grades, whereas  $\text{COD}_{\text{Mn}}$  could only divide the groups into three grades. The information content of  $\text{NH}_3$  was higher than that of  $\text{COD}_{\text{Mn}}$  and should be given the highest weight. To sum up, the reasonable ranking of weights should be  $\text{NH}_3 > \text{COD}_{\text{Mn}} > \text{sulfide}$ .

Of the five participating water sources, only Water 1 was rated excellent in all three criteria. In addition, by comparing  $\text{COD}_{\text{Mn}}$  and  $\text{NH}_3$ , which was of great importance in determining water quality, the pollution degree of Water 1 was the lowest. In the least-significant sulfide index comparison, the concentration of pollutants in Water 1 was higher than that of the other four samples but was still at excellent level. Therefore, the water quality of Water 1 was the best among all the participating samples and should be selected as the water source.

**3.2. Entropy Weight Results.** Table 3 shows the weights of the calculated indexes based on the EWM.

TABLE 2: Measured values generated by Monte Carlo Simulation and their grades.

Sample	Measured values (mg/L)			Degree		
	COD <sub>Mn</sub>	NH <sub>3</sub>	Sulfide	COD <sub>Mn</sub>	NH <sub>3</sub>	Sulfide
Water 1	0.3	0.1	0.045	Excellent	Excellent	Excellent
Water 2	0.35	1.75	0	Excellent	Worst	Excellent
Water 3	2.25	1.25	0	Good	Bad	Excellent
Water 4	3.75	0.75	0	Good	Medium	Excellent
Water 5	5.25	0.25	0	Medium	Good	Excellent

TABLE 3: The process of entropy weighting.

Sample 1	Normalized value					Entropy	Weight
	Water 1	Water 2	Water 3	Water 4	Water 5		
COD <sub>Mn</sub>	0.025	0.029	0.189	0.315	0.441	0.994	0.138
NH <sub>3</sub>	0.024	0.427	0.305	0.183	0.061	0.995	0.119
Sulfide	1.000	0.000	0.000	0.000	0.000	0.970	0.743

Conclusions were easily drawn based on Table 3. In the weighted result of the EWM, the weight of sulfide with the lowest dispersion degree of measured data and no grade discrimination was as high as 0.743, far more than any other pollutants. However, the weight of NH<sub>3</sub>, the smallest among all the indicators, had the highest degree of discrimination and grade discrimination, which was only 0.119. The weights in Table 3 were opposite of the reasonable weight ranking discussed above.

By comparing Tables 2 and 3, the weight distortion of the EWM was found to have come from two aspects:

- (1) The measured value of sulfide contained too many zero values. After standardization, the zero value in the measured value was converted into the zero value in the normalized value. In the calculation process of the EWM, when the normalized value is  $p_{ij} = 0$ , let  $p_{ij} \cdot \ln p_{ij} = 0$ . Thus, excessive zero values led to low entropy value and high weight of sulfide.
- (2) When the threshold of the evaluation index was divided differently, such as the NH<sub>3</sub> and permanganate index in this example, no necessary relationship emerged between the numerical and the grade differentiation degrees. However, the EWM only considered the numerical discrimination degree and ignored the rank discrimination degree of the index. However, in the multi-index decision-making involving classification, the classification degree more accurately reflected the information amount of the index.

**3.3. Comprehensive Evaluation Results.** The weight calculation results in Table 3 were substituted into the water quality index model. The water quality index of each evaluated water area was calculated, as shown in Table 4.

Combined with Table 4, the weight distortion led to the following problems in the comprehensive evaluation results:

- (1) As the EWM gave much weight to sulfide without grade differentiation and too little weight to NH<sub>3</sub>

with the highest grade differentiation, the comprehensive water quality of the five water areas was rated excellent without any difference. This result led to difficult choices.

- (2) Given that the EWM gave too much weight to sulfide with the lowest pollution degree and ignored NH<sub>3</sub> with the most serious pollution degree, the evaluation result of water quality was too optimistic. By comparing Tables 2 and 4, NH<sub>3</sub> pollution in Water 2 had the worst water level and Water 3 had a bad level. These areas had extremely high nutrient content and eutrophication risk. Given the low weight of NH<sub>3</sub> given by the EWM, both water sources were rated excellent.
- (3) According to the discussion in Section 3.1, all the indicators of Water 1 could reach excellent level. Thus, Water 1 had the best water quality and should be selected as the water source. However, given the distortion of the weight result of the EWM, Water 1 had the worst water quality sample in the final decision-making and should not be selected as the water source.

**3.4. The Potential Solutions of the Distortions of EWM.** As is illustrated in Section 3.2, the EWM has the following two distortions:

- (i) When the measured data set contains too many zero values, its entropy value may be undervalued, which makes the weight overexaggerated
- (ii) When the threshold of the evaluation index is divided differently, the numerical discrimination degree cannot correctly reflect the grade differentiation degree

The first distortion is a technical problem instead of a theoretical background problem. As is discussed in Section 3.2, the proximate cause for the first distortion is that the zero values in the measured data set correspond to the zero normalized value. As a result, it may be solved by modifying

TABLE 4: Water quality evaluation results.

Sample	COD <sub>Mn</sub>	NH <sub>3</sub>	Sulfide	Water quality index	Water quality degree
Water 1	97.0	86.7	82.0	84.6	Excellent
Water 2	96.5	10.0	100.0	88.8	Excellent
Water 3	77.5	30.0	100.0	88.6	Excellent
Water 4	62.5	50.0	100.0	88.9	Excellent
Water 5	47.5	74.3	100.0	89.7	Excellent

the standardization method to avoid the zero values in the normalized data set. For example, a protentional substitution formula of equation (1) is

$$p_{ij} = \frac{(x_{ij} + C)}{\sum_{j=1}^n (x_{ij} + C)}, \quad (6)$$

where  $C$  is a constant which should at least satisfy

$$x_{ij} + C > 0. \quad (7)$$

Obviously, for the zero values in the measured data set, the larger the  $C$  is, the farther its normalized value exceeds zero, which reduces the influence of the first distortion. However, for the measured data set, the larger the  $C$  is, the less its discrimination degree is. Therefore, the concrete selecting method of the constant  $C$  still needs to be further studied.

The second distortion is a theoretical background problem rather than a technical problem because the EWM only considered the numerical discrimination degree and ignored the rank discrimination degree of the index. Considering that the classification degree more accurately reflected the information amount of the index, we thought that the theoretical basis of the EWM is partial in the multi-index decision-making problem. A protentional solution method is used to introduce new variables which represent rank discrimination degree into the weighting process. However, the concrete combination method between the numerical discrimination degree and the rank discrimination degree also needs to be further studied.

#### 4. Conclusion

The rationality of the EWM in decision-making is questionable. First, when too many zero values are in the measured values, the standardized results of the EWM are prone to distortion. Subsequently, this outcome will lead to the excessive weight of the index with low actual differentiation degree. Second, the classification degree can accurately reflect the information amount of the index in multi-index decision-making involving classification. However, the EWM only considers the numerical discrimination degree and ignores the rank discrimination degree of the index. These two shortcomings indicate that the EWM is unable to reflect the importance of the index weight correctly, thus resulting in distorted decision-making results.

The protentional solutions of these two distortions are modifying the standardization method and introducing new rank discrimination degree variables, respectively. However,

the concrete algorithms of these solutions still need to be further studied.

#### Data Availability

The data used to support the findings of this study are included within the article.

#### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

#### Acknowledgments

This work was supported by the National Natural Science Foundation of China under the contract no. 51709142.

#### References

- [1] L. Liu, J. Zhou, X. An, Y. Zhang, and L. Yang, "Using fuzzy theory and information entropy for water quality assessment in Three Gorges region, China," *Expert Systems with Applications*, vol. 37, no. 3, pp. 2517–2521, 2010.
- [2] Z. Zhi-Hong, Y. Yi, and S. Jing-Nan, "Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment," *Journal of Environmental Sciences*, vol. 18, pp. 1020–1023, 2006.
- [3] X. W. Ding, X. Chong, Z. F. Bao, Y. Xue, and S. H. Zhang, "Fuzzy comprehensive assessment method based on the entropy weight method and its application in the water environmental safety evaluation of the heshangshan drinking water source area," *Three Gorges Reservoir Area*, vol. 9, p. 15, 2017.
- [4] M. Taheriyoun, M. Karamouz, and A. Baghvand, "Development OF an entropy-based fuzzy eutrophication index for reservoir water quality evaluation," *Iranian Journal of Environmental Health Science & Engineering*, vol. 7, pp. 1–14, 2010.
- [5] J. Wu, P. Li, H. Qian, and J. Chen, "On the sensitivity of entropy weight to sample statistics in assessing water quality: statistical analysis based on large stochastic samples," *Environmental Earth Sciences*, vol. 74, no. 3, pp. 2185–2195, 2015.
- [6] F. Yan, B. Qian, and X. Xiao, "Geo-accumulation vector model for evaluating the heavy metal pollution in the sediments of Western Dongting Lake," *Journal of Hydrology*, vol. 567, no. 7, pp. 112–124, 2019.
- [7] F. Yan, D. Y. Qiao, and B. Qian, "Improvement of CCME WQI using grey relational method," *Journal of Hydrology*, vol. 543, no. 2, pp. 316–323, 2019.
- [8] J. H. Wu, C. Y. Xue, R. Tian, and S. Wang, "Lake water quality assessment: a case study of Shahu Lake in the semiarid loess area of northwest China," *Environmental Earth Sciences*, vol. 76, p. 15, 2017.

- [9] J.-Y. Zhang and L.-C. Wang, "Assessment of water resource security in Chongqing City of China: what has been done and what remains to be done?" *Natural Hazards*, vol. 75, no. 3, pp. 2751–2772, 2015.
- [10] F. C. Yu, G. H. Fang, and X. W. Ru, "Eutrophication, health risk assessment and spatial analysis of water quality in Gucheng Lake, China," *Environmental Earth Sciences*, vol. 59, no. 8, pp. 1741–1748, 2010.
- [11] X. Lu, L. Y. Li, K. Lei, L. Wang, Y. Zhai, and M. Zhai, "Water quality assessment of Wei River, China using fuzzy synthetic evaluation," *Environmental Earth Sciences*, vol. 60, no. 8, pp. 1693–1699, 2010.
- [12] Y. Zhou, Q. Zhang, K. Li, and X. Chen, "Hydrological effects of water reservoirs on hydrological processes in the East River (China) basin: complexity evaluations based on the multi-scale entropy analysis," *Hydrological Processes*, vol. 26, no. 21, pp. 3253–3262, 2012.
- [13] Y. Cui, P. Feng, J. L. Jin, and L. Liu, "Water resources carrying capacity evaluation and diagnosis based on set pair analysis and improved the entropy weight method," *Entropy*, vol. 20, 2018.
- [14] D. Wang, V. P. Singh, Y.-S. Zhu, and J.-C. Wu, "Stochastic observation error and uncertainty in water quality evaluation," *Advances in Water Resources*, vol. 32, no. 10, pp. 1526–1534, 2009.
- [15] S. V. Weijjs, G. Schoups, and N. van de Giesen, "Why hydrological predictions should be evaluated using information theory," *Hydrology and Earth System Sciences*, vol. 14, no. 12, pp. 2545–2558, 2010.
- [16] A. D. Gorgij, O. Kisi, A. A. Moghaddam, and A. Taghipour, "Groundwater quality ranking for drinking purposes, using the entropy method and the spatial autocorrelation index," *Environmental Earth Sciences*, vol. 76, p. 9, 2017.
- [17] X. G. Li, X. Wei, and Q. Huang, "Comprehensive entropy weight observability-controllability risk analysis and its application to water resource decision-making," *Water SA*, vol. 38, pp. 573–579, 2012.
- [18] G. H. Dong, J. Q. Shen, Y. Z. Jia, and F. H. Sun, "Comprehensive evaluation of water resource security: case study from Luoyang City, China," *Water*, vol. 10, p. 19, 2018.
- [19] V. Amiri, M. Rezaei, and N. Sohrabi, "Groundwater quality assessment using entropy weighted water quality index (EWQI) in Lenjanat, Iran," *Environmental Earth Sciences*, vol. 72, no. 9, pp. 3479–3490, 2014.
- [20] J. Chen, Y. Zhang, Z. Chen, and Z. Nie, "Improving assessment of groundwater sustainability with analytic hierarchy process and information entropy method: a case study of the Hohhot Plain, China," *Environmental Earth Sciences*, vol. 73, no. 5, pp. 2353–2363, 2015.