Hybrid model for water quality assessment in Bosten Lake of China

Yumin Wang1,*, Shuyuan Ran2
1Southeast University, 2#, Sipailou Street, Nanjing City, Jiangsu Province, China, 210096
2Electric Power Design Institute Co. Ltd. in Jiangsu Province, China, 211102

*Corresponding author e-mail: wangyumin@seu.edu.cn

Abstract. A hybrid method to comprehensively assess the water quality level was proposed by combining triangular fuzzy numbers, Monte Carlo approaches, and fuzzy matter element model, which can provide more meaningful information for the environmental protection agency. The model was applied to assessment of water quality in the Bosten Lake. The results implied that water quality level for the Bosten Lake is level III, and the water quality of sampling sites were S7 > S13 > S17 > S9 > S11 > S15 > S12 > S14 > S10 > S1 > S6 > S16 > S5 > S4 > S3 > S2 > S8. The methodology can be employed to deal with uncertainty and fuzzy data by stochastically simulating their distribution characteristics, and obtain a better understanding of water quality levels.

1. Introduction
Water resources in lakes with desirable quantity and quality are a prerequisite for sustainable development [1]. However, water quality issue of lakes has become a serious problem worldwide induced from both natural process (e.g., climate change, etc.) and anthropogenic activities (e.g., pollution, domestic drainage, etc.) [2-4].

In freshwater lakes, various solutes such as chemical oxygen demand COD, dissolved oxygen DO, total phosphorus TP, and total nitrogen TN affect lakes’ water quality. However, for specific lake, water quality levels judged by concentration of each solute and environmental quality standards (GB3838-2002) [5] were often inconsistent [6]. Therefore, synthetic evaluation by selecting important solutes as indicators became essential. In the past literatures, various synthetic methods have been applied to assess water quality. The water quality index (WQI) transform the water quality data into a value to interpret the water quality, which have been widely applied all around the world [7, 8]. Multivariate statistics analysis methods, such as cluster analysis CA, principal component analysis PCA, have been conducted to explore relationships among parameters [9], geochemical assessment of groundwater [10], and temporal and spatial trends [11]. To deal with uncertainty of water quality indicators, it is useful to combing fuzzy theory and entropy theory [12-14]. Matter element is also an effective method for comprehensive evaluation by using information scientifically and rationally [15]. However, large water quality data required by multivariate statistical analysis to obtain reasonable results were often difficult to achieve. Since fuzzy theory can deal with ambiguous water quality information, the fuzzy matter element (FME) model has been widely used for water quality assessment [16-18]. Since many uncertainties exist in scarce and imprecise monitoring dat, triangular fuzzy numbers (TFN) are employed to describe the input data for assessing the water quality [19]. The
Monte Carlo (MC) method can be applied to deal with the complex operation of TFN by converting the arithmetis operation to the real numbers operations based on variables’ probability distributions.

In this study, firstly, to deal with the data uncertainty, the hybrid model of TFN, FME, and MC method was developed by introducing the TFN and MC theory into the FME model. Secondly, the hybrid model was used. Finally, the fuzzy nearitude to certain water quality classification and comprehensive water quality evaluations were obtained for each sampling site, and could supply accordance for environmental management agency to make reasonable policies.

2. Method

2.1 Study area

The Bosten Lake (41°56′~42°14′N, 86°40′~87°56′E) is located in the south part of Tianshan Mountain and in Bohu County. The lake has an area of 968 km² at water level of 1047.5m, with about 55km long, 25km wide, and mean depth of 7.38m. The Bosten Lake has many effects on controlling floods, supplying water, and protecting environment and ecology [20].

![Figure 1. The distribution of sampling sites in Bosten Lake](image)

2.2 Hybrid model

The specific operation processes were described as follows:

1) Triangular fuzzy number method

Suppose a TFN $\tilde{A}$ is represented by $A_1, A_2,$ and $A_3$, referring to the lower, expected and upper values of $\tilde{A}$ with $A_1 < A_2 < A_3$. The membership function of the TFN $\tilde{A}$ is expressed as:

$$\mu_{\tilde{A}}(x) = \begin{cases} 0 & x \leq A_1 \\ \frac{(x-A_1)}{(A_2-A_1)} & A_1 < x \leq A_2 \\ \frac{(A_3-x)}{(A_3-A_2)} & A_2 < x \leq A_3 \\ 0 & x > A_3 \end{cases}$$  (1)

The parameters of $\tilde{A}$ are determined as follows:

$$A_i = \overline{A} - i \sigma, A_i = \underline{A} + i \sigma$$  (2)

2) Fuzzy matter element model

The fuzzy matter element $R_k$ is established by $m$ indicators $C_i$, $n$ water quality levels $G_j$, and the fuzzy values $\mu_{ij}$, and can be expressed as follows:
where $C_i$ refers to the $i$th indicator, $i = 1, 2, \ldots, m$; $G_j$ refers to the $j$th water quality classification, $j = 1, 2, \ldots, n$; $\mu_{ij}$ refers to the fuzzy membership degree, which is expressed as follows:

$$
\mu_{ij} = \exp\left(-\frac{(x_i - a_{ij})^2}{b_{ij}}\right)
$$

(4)

where $a_{ij}, b_{ij}$ are defined as follows:

$$
a_{ij} = x_p(x_i), \quad j = 1
$$

$$
a_{ij} = \frac{x_p + x_q}{2}, \quad j = 2,3,4
$$

$$
a_{ij} = x_q(x_i), \quad j = 5
$$

(5)

where $x_p$ and $x_q$ are the lower, upper boundaries, the variable in parentheses refers to the cost indicators, which means that the smaller is the better, i.e., Chl-a, COD, TN, and TP, otherwise, the variable refers to the efficiency indicators, i.e., SD, $b_{ij}$ is expressed as Eq. (6):

$$
b_{ij} = \begin{cases} 
\frac{x_q - x_p}{\sqrt{\ln 2}}, & j = 1,5 \\
\frac{x_q - x_p}{2\sqrt{\ln 2}}, & j = 2,3,4 
\end{cases}
$$

(6)

Table 1. Water quality classification based on the environmental quality standards for surface water, China (GB3838-2002)

| Parameters         | I     | II    | III   | IV    | V     |
|-------------------|-------|-------|-------|-------|-------|
| DO (mg/L) ≥       | 7.5   | 6     | 5     | 3     | 2     |
| COD$_{Mn}$ (mg/L) ≤ | 2     | 4     | 6     | 10    | 15    |
| TP (mg/L) ≤       | 0.01  | 0.025 | 0.05  | 0.1   | 0.2   |
| TN (mg/L) ≤       | 0.2   | 0.5   | 1     | 1.5   | 2     |
| SD (m) ≥          | 15    | 4     | 2.5   | 1.5   | 0.5   |

The correlation $r_{ij}$ is expressed as follows [16, 21]:

$$
r_{ij} = \begin{cases} 
\frac{2(x_i - x_p)}{x_p - x_q}, & x_i \leq \frac{x_p + x_q}{2} \\
\frac{2(x_i - x_q)}{x_p - x_q}, & x_i > \frac{x_p + x_q}{2} 
\end{cases}
$$

(7)

Let

$$
r_{ij} = \max_{j=1,2,\ldots,n} (r_{ij})
$$

(8)

For benefit indicators, let

$$
r_j = \begin{cases} 
(m - f_{j_{\max}}) \times (1 + r_{j_{\max}}), & r_{j_{\max}} \geq -0.5 \\
(m - f_{j_{\max}}) \times 0.5, & r_{j_{\max}} < -0.5 
\end{cases}
$$

(9)

Similarly, for cost indicators, let

$$
r_j = \begin{cases} 
f_{j_{\max}} \times (1 + r_{j_{\max}}), & r_{j_{\max}} \geq -0.5 \\
f_{j_{\max}} \times 0.5, & r_{j_{\max}} < -0.5 
\end{cases}
$$

(10)

Finally, the weight of indicators is determined by Eq. (11) as follows:

$$
w_i = \frac{r_i}{\sum r_i}
$$

(11)

The nearitude to the $j$th water quality classification is represented by Hamming nearitude ($\rho_H$) as follows:
\[
\rho H_j = 1 - \sum_{j=1}^{n} w_j |\mu_{ij} - \mu_{0j}|
\]

where \( \mu_{ij} \) is the normalized fuzzy matter matrix \( (\mu_{ij}) \), and

\[
\mu_{0j} = \begin{bmatrix}
\rho_{01} \\
\rho_{02} \\
\vdots \\
\rho_{0n}
\end{bmatrix}
\]

Suppose \( \rho H_j = \max\{\rho H_j\}, (j = 1, 2, \ldots, n) \), then the evaluated lake belonged to the \( j \)th level.

For comparing the evaluation objects with the same water quality classification, the non-integral feature value is generated by Eq. (14) as follows:

\[
J = \frac{\sum_{j=1}^{n} j \times \rho H_j}{\sum_{j=1}^{n} \rho H_j}
\]

When \( j < J < j + 1, j = 0, 1, 2, 3, \text{and } 4 \), the water quality levels are 1, 2, 3, 4, and 5, respectively. The comprehensive water quality levels were acquired by Eq. (15), which is showed as follows:

\[
L = \sum_{s=1}^{5} J_s P_s
\]

where \( L \) is the comprehensive water quality level, \( q \) is the possible number of lake’s water quality level, \( J_s \) is the possible level, \( P_s \) is the probabilities of each level, which is obtained by MC simulation of forecast variable \( J \).

3) Monto Carlo method performed Crystal Ball software

The TFN variables were the independent variables, and \( \rho H_j \) and \( J \) were the predictive variables. The frequency distribution of results became convergent when \( N \) is large enough.

3. Results and discussions

After simulating 60,000 times, the stochastic simulations of each sampling site became convergent. Figure 2 showed the fuzzy nearitude values to water quality classifications of 17 sampling sites in the Bosten Lake.

![Figure 2. Fuzzy nearitude values to classifications I, II, III, and IV](image)

The fuzzy nearitudes to level III were the greatest in all the sampling sites, which meant general level for Bosten Lake is III. In addition, we can also find that the water quality in S4, S7, S13, S14, S16, and S17 were worse than the water quality in other sampling sites. Among them, the water quality in S7 is the worst since the fuzzy nearitude to level IV was the greatest in S7.

The comprehensive water quality scores of sampling sites were shown in Table 2.
Table 2. Comprehensive water quality scores

| Region  | Sites | Comprehensive water quality |
|---------|-------|----------------------------|
| Northwest | S7    | 2.9997                     |
|          | S9    | 2.9955                     |
|          | S11   | 2.986                      |
| Southwest | S12   | 2.984                      |
|          | S13   | 2.9986                     |
|          | S14   | 2.966                      |
| East     | S16   | 2.957                      |
|          | S17   | 2.9966                     |
| Middle   | S1    | 2.962                      |
|          | S2    | 2.776                      |
|          | S3    | 2.842                      |
|          | S4    | 2.887                      |
|          | S5    | 2.89                       |
|          | S6    | 2.937                      |
|          | S8    | 2.014                      |
|          | S10   | 2.964                      |
|          | S15   | 2.985                      |

The water quality levels were higher in the northwestern, southwestern, and eastern region than in the middle region, which is consistent with the other relevant studies [23] that proved the correctness. The reason is that the northwestern is the inlet of the Bosten Lake from the Huangshui Ditch, receiving the return water of agricultural irrigation ditches with plenty of TN and TP, and the southwestern is the inlet of the Bosten Lake from the Qingshui River with lots of COD drained from upstream industrial enterprises. However, the reason why the water level of eastern of the Bosten Lake is higher than the middle may be due to the effect of wind and bump stations, which formed the lake flow transferring pollutants from northern and southwestern to the eastern.

By employing the proposed hybrid model, the water quality level, as well as the possible levels and corresponding probabilities were obtained, showed in Table 5. The developed methodology can provide more valuable information for decision makers by identifying the sampling sites with a worsening tendency. The comprehensive water quality level were S7 > S13 > S17 > S9 > S11 > S15 > S12 > S14 > S10 > S1 > S6 > S16 > S5 > S4 > S3 > S2 > S8.

4. Conclusion

By employing the hybrid method, the fuzzy nearitudes of water quality of the evaluation objects to a certain classification were obtained which can decide the water level of the sampling sites in the Bosten Lake. In addition, the non-integral ranks and comprehensive water quality were also obtained by the hybrid method, which can distinguish the difference among sampling sites in the Bosten Lake. The fuzzy nearitudes of all the sampling sites to water quality level III were the biggest, which mean that the general water quality is level III in the Bosten Lake. The non-integral values of all the sampling sites were between 2 and 3, which also prove that the water quality level was III. The comprehensive water quality level showed that the water quality level were S7 > S13 > S17 > S9 > S11 > S15 > S12 > S14 > S10 > S1 > S6 > S16 > S5 > S4 > S3 > S2 > S8. In addition, the water quality level of zones in the lake decreased in the order of Northwest > Southwest > East > Middle.

References

[1] Wu, Z.; Wang, X.; Chen, Y.; Cai, Y.; Deng, J., Assessing river water quality using water quality index in Lake Taihu Basin, China. The Science of the total environment 2018, 612, 914-922.
[2] Wu, Z.; Zhang, D.; Cai, Y.; Wang, X.; Zhang, L.; Chen, Y., Water quality assessment based on the water quality index method in Lake Poyang: The largest freshwater lake in China. Scientific reports 2017, 7, (1), 17999

[3] Todd, A. S.; Manning, A. H.; Verplanck, P. L.; Crouch, C.; McKnight, D. M.; Dunham, R., Climate-change-driven deterioration of water quality in a mineralized watershed. Environmental science & technology 2012, 46, (17), 9324-32.

[4] Ali, E. M.; Khairy, H. M., Environmental assessment of drainage water impacts on water quality and eutrophication level of Lake Idku, Egypt. Environmental pollution 2016, 216, 437-49.

[5] SEPA, Environmental quality standards for surface water (GB3838-2002). 2002, (in Chinese).

[6] Liu, D.; Zou, Z., Water quality evaluation based on improved fuzzy matter-element method. Journal of Environmental Sciences 2012, 24, (7), 1210-1216.

[7] Bora, M.; Goswami, D. C., Water quality assessment in terms of water quality index (WQI): case study of the Kolong River, Assam, India. Applied Water Science 2016, 7, (6), 3125-3135.

[8] Ponsadailakshmi, S.; Sankari, S. G.; Prasanna, S. M.; Madhuramal, G., Evaluation of water quality suitability for drinking using drinking water quality index in Nagapattinam district, Tamil Nadu in Southern India. Groundwater for Sustainable Development 2018, 6, 43-49.

[9] Noori, R.; Sabahi, M. S.; Karbassi, A. R.; Baghvand, A.; Taati Zadeh, H., Multivariate statistical analysis of surface water quality based on correlations and variations in the data set. Desalination 2010, 260, (1-3), 129-136.

[10] Singh, C. K.; Kumar, A.; Shashtri, S.; Kumar, A.; Kumar, P.; Mallick, J., Multivariate statistical analysis and geochemical modeling for geochemical assessment of groundwater of Delhi, India. Journal of Geochemical Exploration 2017, 175, 59-71.

[11] Ogwueleka, T. C., Use of multivariate statistical techniques for the evaluation of temporal and spatial variations in water quality of the Kaduna River, Nigeria. Environmental monitoring and assessment 2015, 187, (3), 137.

[12] Lobato, T. C.; Hauser-Davis, R. A.; Oliveira, T. F.; Silveira, A. M.; Silva, H. A. N.; Tavares, M. R. M.; Saraiva, A. C. F., Construction of a novel water quality index and quality indicator for reservoir water quality evaluation: A case study in the Amazon region. Journal of Hydrology 2015, 522, 674-683.

[13] Li, R.; Zou, Z.; An, Y., Water quality assessment in Qu River based on fuzzy water pollution index method. J Environ Sci (China) 2016, 30, 87-92.

[14] Liu, L.; Zhou, J.; An, X.; Zhang, Y.; Yang, L., Using fuzzy theory and information entropy for water quality assessment in Three Gorges region, China. Expert Systems with Applications 2010, 37, (3), 2517-2521.

[15] He, Y.; Dai, A.; Zhu, J.; He, H.; Li, F., Risk assessment of urban network planning in china based on the matter-element model and extension analysis. International Journal of Electrical Power & Energy Systems 2011, 33, (3), 775-782.

[16] Li, B.; Yang, G.; Wan, R.; Hormann, G., Dynamic water quality evaluation based on fuzzy matter-element model and functional data analysis, a case study in Poyang Lake. Environmental science and pollution research international 2017.

[17] Deng, X.; Xu, Y.; Han, L.; Yu, Z.; Yang, M.; Pan, G., Assessment of river health based on an improved entropy-based fuzzy matter-element model in the Taihu Plain, China. Ecological Indicators 2015, 57, 85-95.

[18] Zhao, Z.; Guo, Y.; Wei, H.; Ran, Q.; Gu, W., Predictions of the Potential Geographical Distribution and Quality of a Gynostemma pentaphyllum Base on the Fuzzy Matter Element Model in China. Sustainability 2017, 9, (7), 1114.

[19] Zhi, G.; Chen, Y.; Liao, Z.; Walther, M.; Yuan, X., Comprehensive assessment of eutrophication status based on Monte Carlo–triangular fuzzy numbers model: site study of Dongting Lake, Mid-South China. Environmental Earth Sciences 2016, 75, (12).

[20] Liu, Y.; Mu, S.; Bao, A.; Zhang, D.; Pan, X., Effects of salinity and (an)ions on arsenic behavior in sediment of Bosten Lake, Northwest China. Environmental Earth Sciences 2014, 73, (8), 4707-4716.
[21] Wong, H.; Hu, B. Q., Application of improved extension evaluation method to water quality evaluation. Journal of Hydrology 2014, 509, 539-548.
[22] Andersen, N. J. H.; Brandstrup, J., Monte Carlo Simulation in Crystal Ball 7.3. AARHUS university 2008.
[23] He, Y.; Jiang, L.; Zhao, J., Temporal and spatial variations of total nitrogen and total phosphorus nutrients in the Bosten lake. Journal of Shihezi University: natural science 2016, 34, (2), 260-264 (in Chinese).
[24] Cai, W. 1999 Extension theory and its application. Chinese Science Bulletin 44 (17), 1538–1548.