Groundwater quality evaluation model based on multi-scale fuzzy comprehensive evaluation and big data analysis method

Hongxia Cheng and Zhang Minghui

ABSTRACT

The reasonable use of water resources has become an important issue for the sustainable development of humanity in the future. Many researches focus on groundwater quality inspection, but not groundwater quality assessment. This paper aims to study groundwater quality evaluation models based on multi-scale fuzzy comprehensive evaluation and big data analysis methods. We combine coarse-grained multi-scale fuzzy entropy and fuzzy comprehensive evaluation method to establish a groundwater quality evaluation model based on big data environment. The evaluation of groundwater samples from 327 test points in Huangpu District, Xuhui District, Hongkou District, and Putuo District of Shanghai was conducted. The results show that the overall condition of Shanghai groundwater is better, and more than 94% of samples qualified as drinking water sources. The method presented in this paper not only guarantees that the coarse-grained data on all scales are consistent with the length of the original data, but also avoids the phenomenon of data loss, which greatly improves the accuracy of subsequent algorithms.

Key words | big data analysis, groundwater quality evaluation model, multi-scale fuzzy comprehensive evaluation, water resources

Hongxia Cheng (corresponding author)
Zhang Minghui
School of Information Science and Technology,
Zhengzhou Normal University,
Zhengzhou 450044,
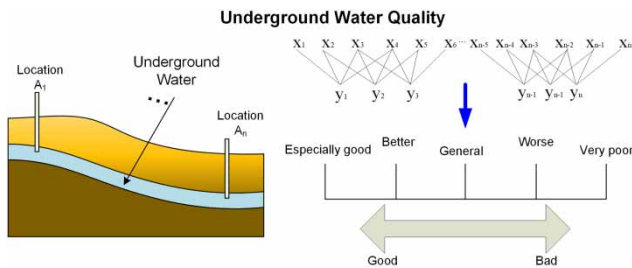China
E-mail: *chenhongxia_x@126.com*

HIGHLIGHTS

- This article first introduces the theory of the multi-scale fuzzy comprehensive evaluation method, then uses the scale analysis method to determine the weight and the fuzzy comprehensive evaluation method to calculate the current status of groundwater resources in Shanghai.
- Finally, it compares the conclusions obtained with the fuzzy comprehensive evaluation method and analyzes it.
- Similarities and differences in evaluation results and their causes.
- It can be seen from the comprehensive comparison that the evaluation using the multi-scale fuzzy comprehensive evaluation method can more intuitively compare the differences in water quality between different administrative regions.
- The evaluation system in this paper is more comprehensive, and its evaluation results are more comprehensive and reasonable than the fuzzy comprehensive evaluation method.

GRAPHICAL ABSTRACT

INTRODUCTION

Results from the 2018 World Health Organization and UNICEF Global Water Supply and Sanitation Evaluation show that the population of 43% in rural Africa, 56% in Latin America, and 67% in Asia do not have access to good quality drinking water (Wu & Sun 2016). The main source of water in many places in rural areas is shallow groundwater. Shallow holes and mechanical or artificially excavated wells are mainly used to provide drinking water (Abbasnia et al. 2018). However, due to the continuous detection of toxic organic chemicals and high concentrations of pathogenic microorganisms from these drinking waters, the safety of water supply in these areas has attracted global attention. Recent studies by the World Health Organization have confirmed that arsenic (>0.01 mg/L) and fluoride (>1.5 mg/L) are higher in shallow groundwater in Argentina, Bangladesh, Cambodia, China, Mongolia, and Tanzania. In addition, coupled with nitrate pollution from human activities and agricultural production, attention to groundwater in developing countries has been strengthened (Gautam et al. 2015).

A series of environmental geological problems caused by groundwater extraction is widespread and becoming increasingly serious. The sustainable development strategy emphasizes the coordination of environment and economy, and the pursuit of harmony between man and nature (Venkatramanan et al. 2015; Sridharan & Nathan 2017). The role of human survival and development on the environment is mainly transmitted through the intermediary link of the development and utilization of natural resources, and whether the utilization mode is reasonable and the

utilization efficiency is of critical significance. Therefore, changing the extensive form of traditional water resources development and utilization, strengthening hydrogeological research, conductimg a comprehensive and detailed investigation of groundwater resources, establishimg a scientific model to evaluate groundwater resources, and based on this, conducting reasonable development and utilization, will have a profound impact on all aspects of China's social and economic development (Mohamed et al. 2018). The serious environmental geological problems caused by the exploitation of groundwater resources have attracted great attention from the governments of many countries. Many countries not only have specialized water resources management institutions, but also have formulated various detailed assessment standards. At the same time, hydrogeological experts and environmental protection experts from various countries are revealing the mechanism of environmental geological problems and studying the corresponding control measures through various channels. However, in the various researches on groundwater resources, no matter which method is adopted, the most important and basic work is the investigation and correct understanding of regional hydrogeological conditions (Azimi et al. 2018).

Yehia et al. (2017) assessed the quality of groundwater for different uses by determining the chemical composition and natural radioactivity of a desert groundwater area. Jasrotia et al. (2018) team compared the physical and chemical parameters of groundwater sample analysis results to the standard guidance values recommended by the World

Health Organization for drinking water and public health standards. Thematic layers related to Ca^{2+}, Mg^{2+}, NO^{3-}, and total hardness (TH) were generated using a GIS platform. The water chemistry of groundwater used for drinking purposes was assessed by plotting cations and anions in Piper's trilinear diagram, but this method is not efficient. Vadiati et al. (2016) developed a new groundwater quality assessment method based on a fuzzy inference system. In their research, widely accepted indicators were also used and compared. Among them, mixed fuzzy indicators minimized uncertainty. This method is useful for groundwater quality assessment, but it is not efficient. Zheng et al. (2015) used triangular fuzzy numbers to represent the interval range of exposure parameters. By selecting acceptable risk management reliability levels, the interval range of exposure parameters is converted to interval estimates. A groundwater quality health risk model based on triangular fuzzy numbers is established. By using the above model, interval estimates of health risk rates of carcinogens and non-carcinogens caused by drinking water, skin contact, and respiratory pathways can be calculated in drinking water sources. Ren et al. (2018) developed an inexact interval-valued triangular fuzzy multiple attribute preference model (IVTF-MAPM) method to support the selection of groundwater remediation strategies. Introducing interval-valued triangular fuzzy parameters into attributes makes it possible to handle multiple uncertainties existing in multiple real worlds, taking into account more possible values and expressing decision information more accurately. Based on the evaluation of groundwater remediation technology, an attribute system consisting of 15 remediation schemes was established, and each remediation scheme corresponds to ten attributes. The pairwise comparison between the selected schemes is represented by the value preference model, and the attribute weights are used in the internal hierarchy analytical score.

Based on the actual situation in Shanghai, this paper divides the quality of groundwater resources in Shanghai into five grades: very poor, poor, fair, good, and very good. A reasonable determination was made. On this basis, the fuzzy mathematics evaluation method is used to divide the Shanghai groundwater quality into five levels of fuzzy evaluation. After data processing by MATLAB, the multi-scale principle is used to classify the index evaluation results.

PROPOSED METHOD

Fuzzy comprehensive evaluation

Basic overview of fuzzy comprehensive evaluation

Fuzzy evaluation refers to some concepts of fuzzy mathematics, and proposes some evaluation methods to solve the actual evaluation problems (Wang & An 2016). Specifically, the basis of fuzzy comprehensive evaluation is fuzzy mathematics. Some factors whose boundaries are unclear and insufficiently quantified are quantified based on fuzzy relations. It is a method that can comprehensively evaluate the membership status of the evaluation affairs from multiple factors (Dai & Zhao 2015). In the process of groundwater quality evaluation, there are many factors that affect groundwater quality, most of which cannot be completely determined, and it is difficult to describe them with mathematical language. Then, in the safety risk evaluation, different influencing factors will have different degrees of influence on it. This forces us to fully consider none of the factors when evaluating it as a whole, because this way we can obtain credible results. As a result, however, to solve this problem, fuzzy comprehensive evaluation method is a good choice.

The core of fuzzy comprehensive evaluation is to use $B = AOR$ for fuzzy transformation calculation. In the model formula: A can be considered as the set of weights of the evaluation factors, and the element $a1(0 \leq a_1 \leq 1)$ in the set is the weight value corresponding to the evaluation factors, which represents the single factor u_1. The magnitude of the effect of the assessment factors on the calculation of water quality also reflects the u_1 assessment level to a certain extent, and the a_1 value is the weight value of each secondary evaluation index obtained by the above analytic hierarchy process, that is, each secondary evaluation. The weight vector of the index to the criterion layer constitutes a weight matrix A1; and the membership degree matrix R1 is formed from the calculated membership degrees of the respective secondary evaluation index values corresponding to the respective evaluation levels. The evaluation result B1 indicates the result of comprehensive evaluation, and the membership of the evaluation level can be seen (Song 2017).

Process of fuzzy comprehensive evaluation method

1. *Establishment of fuzzy comprehensive evaluation index*: The premise of considering the next comprehensive evaluation is to establish an evaluation index. A reasonable evaluation index will be beneficial to the evaluation process, and an unreasonable evaluation index will cause a large deviation in the evaluation results (Fan *et al.* 2016). How to establish a scientific and reasonable evaluation index is very important. Generally, according to the nature of the research target and the accident cases that occurred in the past, it is considered in various aspects in combination with relevant norms (Li *et al.* 2016).

   For example, the establishment of a factor set: $U = \{u_1, u_2, u_3, \ldots, u_i, \ldots, u_m\}$

   The first level is $u_i(i = 1, 2, \ldots, m)$, where the i-th factor is the factor in the highest level, and the *n*-th factor in the second level determines it, that is:

$$u_i = \{u_{i1}, u_{i2}, u_{i3}, \ldots, u_{ij}, \ldots, u_{in}\}(j = 1, 2, \ldots, n) \qquad (1)$$

2. *Reasonable establishment of weight vectors*: The weight vector is established by analytic hierarchy process or expert scoring method. The weights introduced as described above need to meet its consistency test (Long & Chen 2016; Li *et al.* 2017; Zhang *et al.* 2018). The importance of each layer of factors is different, and its weights are not the same, so the set of weights at each level can be divided into:

   First layer of the sub-weight set:

$$A = \{a_1, a_2, \ldots, a_i, \ldots, a_m\} \qquad (2)$$

   $a_i(i = 1, 2, \ldots, m)$ is the weight of the i-th factor u_i in the first level.

   Second layer of the sub-weight set:

$$A_i = \{a_{i1}, a_{i2}, \ldots, a_{ij}, \ldots, a_{in}\}(j = 1, 2, \ldots, n) \qquad (3)$$

   a_{ij} is the weight of the determinant u_{ij} in the second level.

3. *The establishment of the judging set*: Assuming p as the number of total judgments, then this judgment set can be established as:

$$V = \{v_1, v_2, \ldots, v_3\} \qquad (4)$$

Multivariate multiscale fuzzy entropy

Improve coarse graining

Determine the sequence $\{x_{kj}\}_{i-1}^N$, $(k = 1, 2, 3, \ldots, p)$ of p variables, and coarse-grain the original sequence at each scale. The coarse-grained sequence is expressed as:

$$y_{kj}^{\in} = \frac{1}{\in}\sum_{j=1}^{\in}\sum_{i=j}^{j+\in} x_{ki}\left(1 \le j \le \frac{N}{\in}, k = 1, 2, 3, \ldots p\right) \qquad (5)$$

Take scale 4 as an example, the difference between improved coarse graining and traditional coarse graining is shown in this case. The specific process is shown in Figure 1.

As shown in Figure 1, the traditional coarse graining is to sequentially compress the original sequence according to the scale factor. When the length of the original sequence data is limited, as the scale increases, the length of the coarsely grained sequence decreases continuously. When the length of the original data is not an integer multiple of the scale factor, some data will be lost. The above factors will inevitably affect the calculation accuracy of subsequent algorithms. The improved coarse-grained algorithm uses a moving average method to coarse grain the original time series on all scales, which not only ensures that the coarse-grained sequences on each scale are the same length as the original sequence, but also avoids data loss. This greatly improves the accuracy of subsequent algorithms. The improved coarse-grained algorithm is shown in Figure 2.
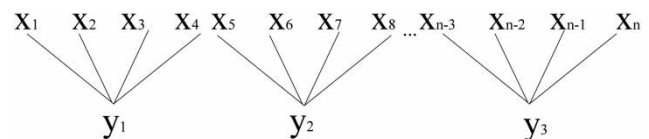

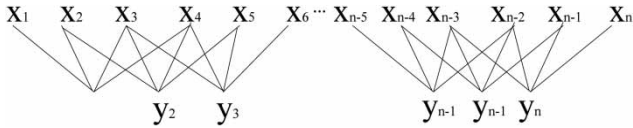**Figure 1** │ Traditional coarse-graining process.

Figure 2 | Improved coarse-graining process.

Multivariate embedding reconstruction is performed on the p-variable sequence $\{x_{kj}\}_{i-1}^{N}$, $(k = 1, 2, 3, ..., p)$ obtained after improved coarsening to obtain N-n composite delay vectors. The specific embedding method is given as,

$$
Y_m(i) = \begin{bmatrix} y_{1,j} & y_{1,i+\tau1} & \cdots & y_{1,i+(m_1-1)\tau_1} \\ y_{2,i} & y_{2,i+\tau_2} & \cdots & y_{2,i+(m_2-1)\tau_2} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ y_{p,i} & y_{p,i+\tau_z} & \cdots & y_{p,i+(m_p-1)\tau_p} \end{bmatrix} \tag{6}
$$

where, $M_m(i) = [m_1, m_2, ..., m_p] \in R^p$ is embedding vector, $\tau = [\tau_1, \tau_2, ..., \tau_p]$ is time delay vector, and $Ym(i) \in R^m$ ($m = \sum_{k=1}^{p} m_k$).

Big data technology architecture and key technologies

Big data technology system

The situational awareness big data technology system mainly includes four aspects: data collection and preprocessing, data storage and management, data analysis, and data display.

1. *Data collection and preprocessing*: Situational awareness big data are complex and the data sources are diverse. Big data processing first collects data from data sources and performs pre-processing operations. For the collected data, proper pre-processing can lay a good foundation for subsequent data analysis. Due to the inevitable existence of noise and interference terms during the data acquisition and transmission process, data errors, even omissions, may occur. Especially when there are different data sources for big data, it is easy for similar, duplicate, or inconsistent data to appear. Therefore, it is necessary to denoise the data and recover the lost data, that is, perform data cleaning.

Filtering is the most common noise reduction method, such as Wiener filtering and Kalman filtering. Interpolation techniques can often effectively recover lost data.

2. *Data storage and management*: Collected data need to be stored in the database in a way that facilitates data processing. Traditional relational databases cannot store unstructured data, have poor scalability, and have difficulty handling massive data. Big data storage and management technology must not only ensure the reliability and readability of files, meet the real-time and effectiveness of data processing, but also minimize costs and improve economics. Big data storage and management technologies are mainly divided into two modes: stream processing and batch processing. When conducting business with high real-time requirements such as online monitoring, the stream processing mode is suitable. At this time, the data are regarded as a stream. When the data stream arrives, it is directly analyzed and processed and the result is returned. In other cases, batch processing mode is used for data storage to provide support for subsequent analysis and processing. In the big data environment, storage and management technologies tend to be distributed. Typical big data storage management technologies include distributed databases based on massively parallel processing (MPP), distributed file storage systems (such as GFS, HDFS, NoSQL), distributed data processing systems (such as Big Table, H Base, Mongodb), etc.

3. *Data analysis*: Data analysis technology is the core technology of big data technology. With this technology, people can discover the value of the data, extract the hidden laws and results, and make more scientific decisions. Due to the randomness and uncertainty of the load of the information security system and the attack, the difficulty of data analysis is greatly increased. Special research is needed to adopt reliable data analysis techniques. Data mining, mathematical statistics, and machine learning are common data analysis techniques. Among them, data mining technology, as a typical data analysis method, can extract potentially useful information and knowledge from a large amount of incomplete and fuzzy data. It involves statistics, artificial intelligence, and database technologies. Various

algorithms such as class analysis, association analysis, classification analysis, sequence analysis, deviation detection, predictive analysis, pattern similarity mining, and regression analysis are implemented. In order to meet the needs of high-speed analysis and processing of big data, big data analysis technology mostly adopts the idea of parallelism, and greatly reduces the calculation time through distributed parallel algorithms. Cloud computing technology distributes big data on a large number of computers, realizes the virtualization of computing resources and physical resources, makes the use of big data possible, is the core principle of big data analysis technology, and provides a platform support for big data analysis. Among them, the distributed data storage technology with GFS and HDFS as the mainstream, the programming model with Map Reduce as the mainstream, large-scale data management technology mainly using Big Table or HBase, virtualization technology, and cloud computing platform management technology are the five core technologies of cloud computing.

4. *Data display*: In order to help users understand the data analysis results more simply and intuitively, the data need to be reasonably displayed to users. The presentation of big data results is more focused on interactivity and visualization than traditional text forms, and visualization techniques have been introduced. Visualization technology is based on computer graphics and image processing technology. It converts data into graphics or images for display on the screen and interacts with users. At present, visualization technology is widely used in intelligent perception and cognition of information security systems. Situational awareness data are complex and massive, difficult to visualize, and are still under development. The results of situational awareness are presented in real-time situational maps, historical situational maps, and situational forecast charts at a given time period in the future. Combined with different visualization graphics, the data observability is improved, so that users can quickly and accurately understand the system operational situation, thereby assisting users to make accurate decision-making.

Application advantages of big data in image processing

First, data technology can realize the reproduction of images, improve the sharpness of images, and not reduce the sharpness of images due to image copying and transmission. Second, in the application of big data technology, the accuracy of image processing can be guaranteed, and the image can be simulated by using two-bit data sets. With modern means, modern scanning technology enables the pixels of an image to be guaranteed. Third, the scope of application of image processing is wide. With the support of big data technology, images have different sources and can truly reflect the size of things. In aerial image processing and electron microscope image processing, the nature of things can be truly reflected through digital coding. Fourth, the flexibility of image processing is very high. In the application of big data technology, image processing can be achieved by means of linear operations and non-linear processing, and digital images can be processed by means of logical relationships. Fifth, image processing under big data technology has great compression potential. In image processing, each pixel is not independent, and the relationship between pixels is very close. The gray-scale similarity between image pixels is large, which promotes image compression.

EXPERIMENTS

Data source

This paper takes the groundwater resources of Shanghai as the research object. Since December 2018, the team has collected groundwater samples from 327 detection points in Shanghai's Huangpu District, Xuhui District, Hongkou District, and Putuo District. It is extracted every day, night and day for one year.

Groundwater samples from the study area were collected in December 2019 and a representative group of 1,000 water samples was selected. All the samples were sent to a professional water quality testing laboratory to get the water quality data. The more uniform sampling distribution can basically represent the water quality of the groundwater in the study area. The sampling location is also the location of the spring water distribution in the study area.

Construction method of impact factor evaluation index system

Establish a hierarchical hierarchical structure

A hierarchical hierarchical structure is established to decompose a complex problem into the components of the index, and then continue to decompose until it can be analyzed intuitively. Finally, a hierarchical hierarchy is formed that has a dominating relationship.

Establishing the grid

The grid acquisition method is a thinking model of human judgment in structural theory. Elements and attributes together form a grid, and linear scales are used to express element attributes. Generally, a scale of 1–5 scales is used to indicate the five grades of the evaluation index, namely: particularly good V, relatively good IV, average III, poor II, and, extremely poor I, as shown in Figure 3. The water quality grade is classified according to the standard of Chinese National Environmental Quality Standard for Surface Water (GB 3838-2002). In this standard, the water grade can be qualified by the quality parameters such as temperature change, pH value, oxygen content, and heavy metal content.

Analyze the grid elements and judge the weights under a single criterion

Different experts are selected to evaluate the weights of the first-level indicators of the impact factor evaluation. Adopting the expert consultation method, the questionnaire to the experts is used to ask the experts to score and combine the results to get the final results. As for the evaluation index, assuming m experts score it, the expert score table
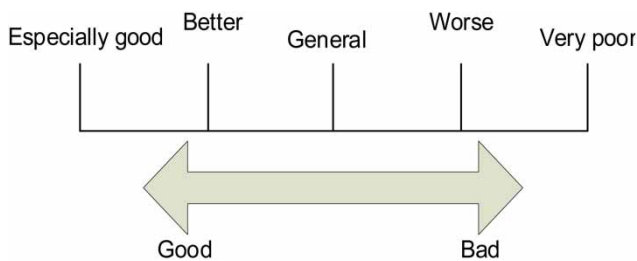

Figure 3 | Importance table.

is shown in Table 1, and the fuzzy Borda method is used to analyze the raster data.

Calculate the weight of the evaluation index system of influencing factors

In the comparison of related influencing factors, in order to improve the accuracy of pairwise comparison of influencing factor indicators at different levels and different levels, an analytic hierarchy process is used to calculate the weighting of influencing factor indicators at each level on the basis of converting qualitative problems into quantitative problems. The quantitative scaling method used in the paper is a 1–9 proportional scaling method to establish a comparison matrix of influencing factor indicators (Table 2).

Expert judgment is made on the pairwise comparison matrix established by Delphi method, and the comprehensive

Table 1 | Expert score sheet

		B_i				
		B_1	B_2	B_3	...	B_Y
P_m	P_1	T_{11}	T_{12}	T_{13}	...	T_{1Y}
	P_2	T_{21}	T_{22}	T_{23}	...	T_{2Y}
	P_3	T_{31}	T_{32}	T_{33}	...	T_{3Y}
	...	...	...	...	...	...
	P_X	T_{X1}	T_{X2}	T_{X3}	...	T_{XY}

Table 2 | Comparison matrix of influencing factor indicators

Scaling	Meaning
1	Two factors are of equal importance compared
3	Factor i compared to j, one of which is slightly more important than the other
5	Factor i compared to j, one of which is more important than the other
7	Factor i compared to j, one of which is more important than the other
9	Factor i compared to j, one of which is more important than the other
2468	The middle number of the above two adjacent judgments
Reciprocal	The reciprocal of the comparison of the above two factors

judgment matrix of the expert group is obtained by using the weighted arithmetic average method, and then single order measurement and consistency inspection are performed on each level of the judgment matrix, and the factors affecting the quality of groundwater are obtained. For the weight of the evaluation index: when the consistency ratio CR >0.1, the estimated consistency of the elements in the judgment matrix is too poor, and it should be re-estimated; when the consistency ratio CR <0.1, the estimates of the elements in the judgment matrix are basically consistent. It has satisfactory consistency and has passed the consistency test. AHR software was used for analysis and calculation. The specific content is as follows.

Constructing the overall objective

A first-level indicator of the groundwater quality judgment matrix, performing weight calculation and consistency ratio test, the calculation process is as follows.

First, the maximum feature λ_{max} and the feature vector $W = [w_1, w_2, ..., w_n]^T$ of the judgment matrix are calculated so that both satisfy $X_w = \lambda_{max} W$. The feature vector $W = [w_1, w_2, ..., w_n]^T$ obtained after normalization of W is used as the ranking weight of the upper index $X_1, X_2, ..., X_n$ of this level of index.

The approximate calculation method is used to calculate λ_{max} and W. The specific steps are:

1. The elements in the judgment matrix X are multiplied by rows, that is:

$$M_i = \prod_{j=1}^{n} X_{ij} (i = 1, 2, ..., n) \tag{7}$$

2. Calculation of $\overline{w_i}$:

$$\overline{w_i} = \sqrt[n]{m_i} \tag{8}$$

3. Normalize $\overline{w_i}$ to get w_i:

$$w_i = \frac{\overline{w_i}}{\sum_{i=1}^{n} \overline{w_i}} \tag{9}$$

where, $W = [w_1, w_2, ..., w_n]^T$ is the required feature vector.

4. Calculate the maximum feature root λ_{max}:

$$\lambda_{max} = \sum_{i=1}^{n} \frac{(Xw)_i}{nw_i} \tag{10}$$

where $(X_w)_i$ represents the i-th element representing X_w.

RESULTS AND ANALYSIS

Multi-scale fuzzy entropy analysis

First, normalize the water quality data of each water intake to ensure that the amplitude and length of each water quality data are within the range of 0–1; the impact of the large value points on the whole is reduced. After that, coarse graining is performed according to the scale factor. The coarse graining uses a sliding window method. After coarse graining, the length of the data on each scale is the original sequence 1 (the scale factor). The coarse-grained time series uses traditional multiscale fuzzy entropy for feature extraction. The parameters selected for multiscale fuzzy entropy are: scale factor $\omega = 1$–10, embedding dimension $m = 2$, delay vector tau $= 1$, similarity tolerance $r = 0.2 \times$ std (std represents the normalized standard deviation). Multi-scale fuzzy entropy feature extraction results are shown in Figure 4.

As shown in Figure 4, on scales 1–2, normal water quality data and abnormal water quality data are far away from each
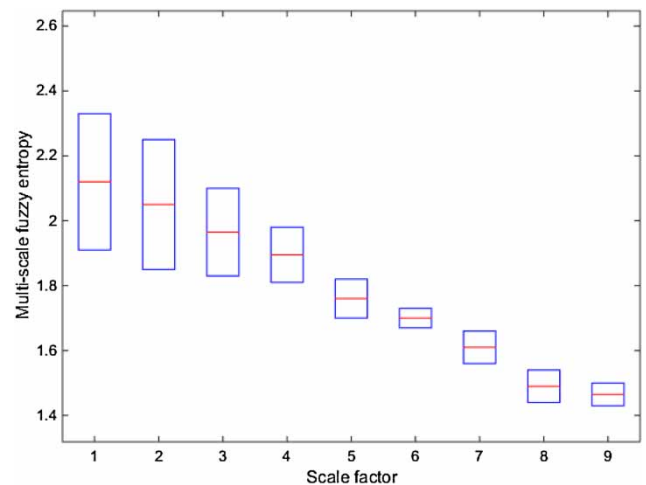

Figure 4 │ Multi-scale fuzzy entropy feature extraction results.

other, and the variance curves have no overlapping parts. Starting from scale 3, as the scale increases, the two data set variance curves. The overlapping components gradually increase, and the variance curves of the two data sets have completely overlapped on scales 6 and 7. On the scales after scale 8, the complexity of the two data sets cannot be clearly distinguished. As a whole, the mean curve of the mean of the two data sets fluctuates greatly at each scale. With the increase of the scale, the complexity of the abnormal gait gradually decreases, while the complexity of the normal gait slowly decreases and gradually tends to smooth.

Results of fuzzy evaluation

Time difference

In this paper, a weighted average fuzzy mathematical model is used, and the obtained membership matrix R and weight matrix A are multiplied according to Matlab software. Table 3 and Figure 5 show the water quality categories of the method in this paper before and after July 2019.

According to Figure 5, it is known that among the 1,000 water samples before July 2019, 504 groups were Grade I water, accounting for 50.4%, 312 groups were Grade II water, accounting for 31.2%, and 115 groups of water samples were Grade III water, accounting for 11.5%, 26 groups of water samples were Grade IV water, accounting for 2.6%, and 43 groups of water samples were Grade V water, accounting for 4.3%. However, there are obvious differences in the evaluation results before and after July. After July, 432 of the 1,000 water samples were Grade I water, accounting for 43.2%, and 229 were Grade II water, accounting for 22.9%, and 156 groups were Grade III water, accounting for 15.6%, and the 103 groups of water samples are Grade IV, accounting for 10.3%, and
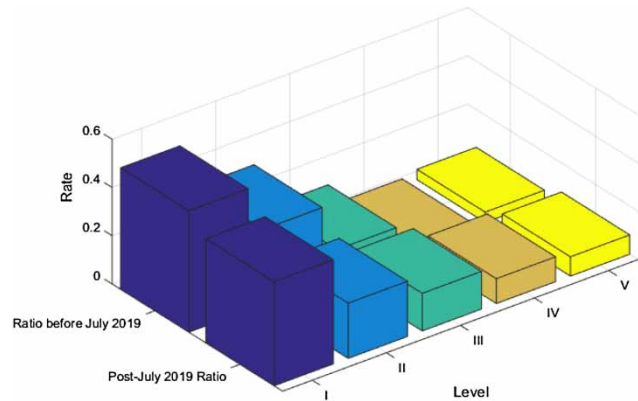

Figure 5 | Results of fuzzy evaluation.

the 43 groups reached Grade V water, accounting for 4.3%. It can be seen that the overall quality of groundwater in Shanghai has improved significantly. After analysis, this may be related to the waste sorting implemented by Shanghai in July. The waste was reasonably sorted and recovered, which reduced the pollution to soil and water quality.

Comparative analysis of water sources

The classification ratio of comprehensive water quality evaluation results are given in Table 4. Figure 6(a) shows the comprehensive evaluation and grading ratio of groundwater quality in 52 administrative districts in Shanghai. Grade V water quality accounts for 24% of the total number of evaluation objects. Grade IV water quality accounts for 34% of the total number of evaluation objects. Grade III water quality accounts for 26% of the total number of evaluation objects. Grade II water quality accounts for 16% of the total number of evaluation objects. Figure 6(b) exhibits the comprehensive evaluation and grading ratio of water quality of Shanghai's 327 water intakes. 50 of them belong to Grade V water quality, accounting for

Table 3 | Results of fuzzy evaluation

No.	Ratio before July 2019	Post-July 2019 ratio
I	0.504	0.432
II	0.312	0.229
III	0.115	0.156
IV	0.026	0.103
V	0.043	0.08

Table 4 | Classification ratio of comprehensive water quality evaluation results

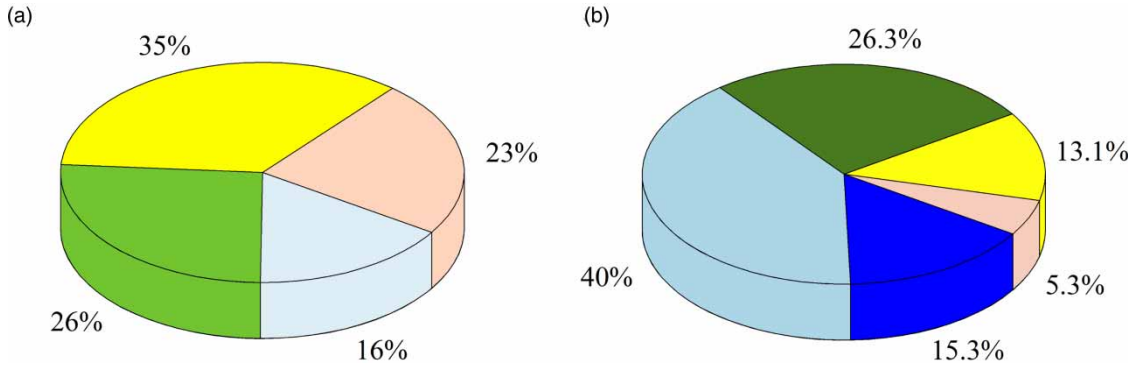|	Water source	Experimental water intake
V	0	15.3%
IV	16%	40%
III	26%	26.3%
II	35%	13.1%
I	23%	5.3%

Figure 6 │ Graded proportion of comprehensive water quality assessment results.

15.3% of the total number of evaluation objects. 131 belong to Grade IV water quality, accounting for 40% of the total number of evaluation objects. 86 of them belong to Grade III water quality, accounting for 26.3% of the total number of evaluation objects. 43 belong to Grade II water quality. 5.3% belong to Grade V water quality. The general condition of groundwater is better, and more than 94% are qualified drinking water sources.

DISCUSSION

This paper uses four different evaluation methods for comparison. The evaluation results of the four different evaluation methods are different. The comparison of specific evaluation results is shown in Figure 7. It can be seen from Figure 7 that the results of the single factor evaluation method and the Nemerow index method used in the evaluation of the poor–very poor groundwater quality accounted for 53.4% and 62.1%, respectively; the fuzzy comprehensive evaluation results showed that the total proportion of groundwater quality category of Grade IV and V water is 19.8%, and the multi-scale comprehensive fuzzy assessment method in this paper shows that the total proportion of groundwater quality category of Grade IV and V water in the study area is 6.9%.

According to Figure 7, the evaluation principle of the single factor evaluation method is similar to that of the Nemerow index, which results in the characteristics of the evaluation results of the two methods being similar. The single factor method determines the evaluation result according to the highest category of a single evaluation factor
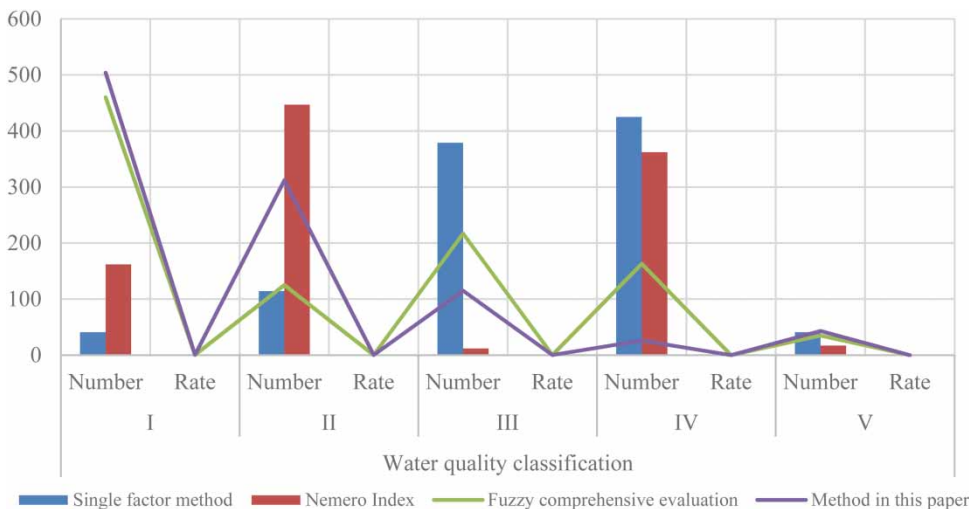

Figure 7 │ Comparison of different methods.

indicator; another method considers the impact of a single over-standard evaluation factor indicator on the overall water sample, even when other evaluation factor indicators in the water sample meet the standard requirements of the evaluation level. However, as long as one of the evaluation factor indicators exceeds the standard, the evaluation results of the water sample may be poor, so the evaluation results of the two methods are close, and the characteristics of the reflected groundwater quality are similar.

Both the fuzzy comprehensive evaluation method and the Nemerow index method emphasize the impact of over-standard evaluation factors on the evaluation results, but there are certain differences. The fuzzy comprehensive evaluation method considers the test results of all evaluation factors participating in the evaluation, and forms a row vector according to the weight allocation of each evaluation factor to finally form a corresponding weight matrix, which reflects the overall impact of the factors involved in the evaluation on the groundwater in the sample. The distribution of Grade I water in the comprehensive fuzzy evaluation results in this paper is relatively concentrated, with less than Grade II water accounting for 12.5%, and water samples exceeding Grade II water accounting for 24%. The main emphasis is on the significant impact of a small number of serious excess factors on water quality evaluation, and the more the number and multiple of evaluation factor excesses, the greater the impact on the final evaluation results, and the lower the evaluation factor is. The degree of influence of the evaluation results is average. This feature is reflected in the evaluation results in this article, that is, the proportion of Grade II water in the evaluation results of this method is 31.2% and that of Grade V water is 4.3%, which is obviously greater than the proportion of fuzzy comprehensive evaluation at the same level. Also, the proportion of Grade II and IV water increased in the evaluation results of the Nemerow index, while the proportion of Grade I and II water decreased.

CONCLUSIONS

This paper combines the coarse-grained multi-scale fuzzy entropy and fuzzy comprehensive evaluation method to establish a groundwater quality evaluation model in a big data environment. Our work introduces the theory of the multi-scale fuzzy comprehensive evaluation method, then uses the scale analysis method to determine the weight and the fuzzy comprehensive evaluation method to calculate the current status of groundwater resources in Shanghai. The evaluation of groundwater samples from 327 test points in Huangpu District, Xuhui District, Hongkou District, and Putuo District of Shanghai were assessed.

The results show that the overall condition of Shanghai groundwater is better, and more than 94% are qualified drinking water sources. Finally, it compares the conclusions obtained with the fuzzy comprehensive evaluation method and analyzes it. It can be seen from the comprehensive comparison that the evaluation using the multi-scale fuzzy comprehensive evaluation method can more intuitively compare the differences in water quality between different administrative regions. The evaluation system in this paper is more comprehensive, and its evaluation results are more comprehensive and reasonable than the fuzzy comprehensive evaluation method.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

Abbasnia, A., Alimohammadi, M., Mahvi, A. H., Nabizadeh, R., Yousefi, M., Mohammadi, A. A., Pasalari, H. & Mirzabeigi, M. 2018 Assessment of groundwater quality and evaluation of scaling and corrosiveness potential of drinking water samples in villages of Chabahr city, Sistan and Baluchistan province in Iran. *Data in Brief* **16**, 182–192.

Azimi, S., Moghaddam, M. A. & Monfared, S. A. H. 2018 Spatial assessment of the potential of groundwater quality using fuzzy AHP in GIS. *Arabian Journal of Geosciences* **11** (7), 142–164.

Dai, C. & Zhao, Z. H. 2015 Fuzzy comprehensive evaluation model for construction risk analysis in urban subway.

International Journal of Modeling Simulation & Scientific Computing **6** (3), 1–18.

Fan, G., Zhong, D., Yan, F. & Yue, P. 2016 A hybrid fuzzy evaluation method for curtain grouting efficiency assessment based on an AHP method extended by D numbers. *Expert Systems with Applications* **44** (C), 289–303.

Gautam, S. K., Maharana, C., Sharma, D., Singh, A. K. & Singh, S. K. 2015 Evaluation of groundwater quality in the Chotanagpur Plateau region of the Subarnarekha River Basin, Jharkhand State, India. *Sustainability of Water Quality and Ecology* **6**, 57–74.

Jasrotia, A. S., Taloor, A. K., Andotra, U. & Bhagat, B. D. 2018 Geoinformatics based groundwater quality assessment for domestic and irrigation uses of the Western Doon valley, Uttarakhand, India. *Groundwater for Sustainable Development* **6**, 200–212.

Li, J., Zhang, H., Han, Y. & Wang, B. 2016 Study on failure of third-party damage for urban gas pipeline based on fuzzy comprehensive evaluation. *Plos One* **11** (11), e0166472.

Li, Q., Xie, L. Y. & Li, H. Y. 2017 Optimization priority analysis of precision gear manufacturing process based on AHP fuzzy comprehensive evaluation method. *Acta Armamentarii* **38** (4), 750–757.

Long, F. & Chen, L. 2016 Fuzzy theory-based athletes professional skills comprehensive evaluation research. *Bio Technology* **10** (4), 8265–8272.

Mohamed, A. K., Liu, D., Mohamed, M. A. A. & Song, K. 2018 Groundwater quality assessment of the quaternary unconsolidated sedimentary basin near the Pi river using fuzzy evaluation technique. *Applied Water Science* **8** (2), 65–67.

Ren, L., Lu, H., Zhao, H. & Xia, J. 2018 An interval-valued triangular fuzzy modified multi-attribute preference model for prioritization of groundwater resources management. *Journal of Hydrology* **562**, 335–345.

Song, R. 2017 Fuzzy comprehensive evaluation for two kinds of suspension controllers. *Applied Mechanics & Materials* **865**, 561–564.

Sridharan, M. & Nathan, D. S. 2017 Groundwater quality assessment for domestic and agriculture purposes in Puducherry region. *Applied Water Science* **7** (5), 1–17.

Vadiati, M., Asghari-Moghaddam, A., Nakhaei, M., Adamowski, J. & Akbarzadeh, A. H. 2016 A fuzzy-logic based decision-making approach for identification of groundwater quality based on groundwater quality indices. *Journal of Environmental Management* **184** (2), 255–270.

Venkatramanan, S., Chung, S. Y., Rajesh, R., Lee, S. Y., Ramkumar, T. & Prasanna, M. V. 2015 Comprehensive studies of hydrogeochemical processes and quality status of groundwater with tools of cluster, grouping analysis, and fuzzy set method using GIS platform: a case study of Dalcheon in Ulsan City, Korea. *Environmental Science and Pollution Research* **22** (15), 11209–11223.

Wang, S. & An, X. 2016 The logistics storage location model of commodity logistics based on fuzzy comprehensive evaluation method. *Journal of Computational & Theoretical Nanoscience* **13** (12), 9664–9668.

Wu, J. & Sun, Z. 2016 Evaluation of shallow groundwater contamination and associated human health risk in an alluvial plain impacted by agricultural and industrial activities, mid-west China. *Exposure & Health* **8** (3), 311–329.

Yehia, M., Baghdady, A., Howari, F. M., Awad, S. & Gad, A. 2017 Natural radioactivity and groundwater quality assessment in the northern area of the Western Desert of Egypt. *Journal of Hydrology Regional Studies* **12** (C), 331–344.

Zhang, X., Li, J. & Liu, J. 2018 Stability evaluation of soft rock supporting system of inclined shaft based on fuzzy comprehensive evaluation method. *Journal of Liaoning Technical University (Natural Science Edition)* **37** (3), 476–481.

Zheng, D., Zeng, F., Shi, C., Sun, C. Z., Zang, Z. & Lin, S. U. 2015 Health risk assessment of groundwater quality in source of drinking water based on the uncertain parameters. *Scientia Geographica Sinica* **18** (4), 322–329.