A cloud model based on hybrid similarity approach for water quality evaluation

Yin Huimin, Zhao Chunlan, Wang Bing, Wang Xiaobo, Li Yi and He Ting

July 16, 2019

A cloud model based on hybrid similarity approach for water quality evaluation

1st Huimin Yin
School of Science
Southwest Petroleum University
Chengdu, China
email address:1733605375@qq.com

2nd Chunlan Zhao*
School of Science & Institute for Artificial Intelligence
Southwest Petroleum University
Chengdu, China
email address:308303451@qq.com

3rd Bing Wang
School of Computer Science
Southwest Petroleum University
Chengdu, China
email address:15222476@qq.com

4th Xiaobo Wang
Comprehensive Room
Sichuan Ecological Environment Monitoring center
Chengdu, China
email address:wangxiaobo@scemc.cn

5th Yi Li
School of Science
Southwest Petroleum University
Chengdu, China
email address:506692799@qq.com

6th Ting He
School of Science
Southwest Petroleum University
Chengdu, China
email address:843697008@qq.com

***Abstract*—Water quality assessment is an essential way to research water environment. Considered uncertainty of randomness and fuzziness in water quality assessment, a cloud model based on hybrid similarity approach (CHS) for water quality assessment was proposed in this paper. Structure Equation Model(SEM) and Entropy methods based on monitored data could determine the prior weights, and then they were combined to obtain a comprehensive weight, SEM can find out the relationship between indicators, which other methods could not. At the same time, based on the advantages both distance and shape similarity could obtain hybrid similarity, then determine water quality level by maximum hybrid similarity value between standard cloud and comprehensive cloud. This approach is utilized to a part of Minjiang River in China, and compared to other three methods, which are Single Factor (SF) method, Comprehensive Pollution Index (CPI) method, and Grey Relation Analysis (GRA) model. The results show that CHS and GRA are in accordance with each other, and they are more reasonable than other two methods because of considering the uncertainty of water quality. This approach was effective to evaluate the water quality level, as a reference for water quality management and applications.**

***Index Terms*—cloud model, water quality evaluation, structure equation model, entropy, hybrid similarity**

I. Introduction

The continuing increase in global population has had a negative effect on the environment in recent decades, leading to pollution of rivers, estuaries and oceans, especially in developing countries[1–4]. Water quality degradation decreasing result in disruption the ecological balance of water bodies and threaten regional environmental, which has been regarded as one of the most serious environmental issues worldwide[1, 4]. Thus, water quality evaluation is a efficient way to prevent worse water pollution in advance.

Various methods have been proposed to evaluation the water quality. The common methods were the Single Factor(SF) method and multifactor method[5, 6]. The SF method was used to classify water quality because of its simple theory, using unique indicator which was the most polluted amongst all of the evaluation factors included in the water standard. i.e. reference [7] assessed water quality in Chinese Taihu lake basin. The Comprehensive Pollution Index(CPI) method was proposed to measure the level by the mean value of all indicators, which are more reasonable than SF. i.e. reference [8] applied CPI to evaluate the Ganges river system at five different sites during multiple seasons. CPI could make higher composite grade, though only one index is high and the others are low, that is, to unreasonable water quality evaluation.

In fact, the uncertainty of water evaluation was caused by complicity of water environment. There are two types of uncertainty that should be considered in the water quality evaluation: the first one is randomness, which is often exhibited in the monitoring and analysis of data; and the second one is fuzziness of classification standard, evaluation class, and degree of pollution [8]. Diverse study for methods based on the above uncertainties have been proposed to determine the water quality, which can be divided into four types:(1) methods based on various statistical and stochastic techniques for randomness[1, 4, 8–11], i.e. reference [9] evaluated the groundwater quality by cluster analysis, exploratory factor analysis and principal component analysis;(2) methods based on fuzzy membership function, fuzzy logic and fuzzy set theory, for instance, reference [12] reported the application of fuzzy set theory for decision-making in the assessment of physico-chemical quality of groundwater for drinking purposes. (3) methods based on machine learning and artificial intelligence for unknown patterns which were hardly captured in the assessment process, i.e. reference [13] settled out method based on combined neural network and genetic algorithm (GA); and (4) hybrid models based on two or more methods or techniques mentioned above[14], i.e. reference [15] proposed a probabilistic fuzzy hybrid model to assess river water quality.

As stated above, randomness and fuzziness are simultane-

Thankfully acknowledge support provided by the subject of the major projects of national science and technology.

ously considered in the water quality assessment, which would be prefered to decrease the error caused by both randomness and fuzziness. Luckily, a new model proposed by Li et al. [16] based on those reduced randomness and fuzziness in system evaluation. The model could quantify both randomness and fuzziness depending on three fixed parameters and presents more advantages than a single randomness or fuzziness model [1]. In 2014, reference [17] firstly applied this model in the field of water quality evaluation, obtaining satisfied results. However, water quality evaluation is a multi-criteria decision-making process, and its key point is to obtain appropriate index weights in the cloud model, thus to receive rational results.

At present, there are amount of studies on the weight, which can be divided into two clusters, one is the subjective weight, and the other one is objective weight. The former is strongly influenced by experts' knowledge, resulting in high biases[18], while the later does not consider differences among indices, and it ignores practical situations [19]. Therefore, in 2018, reference [1] proposed a comprehensive weight based on AHP-Entropy for water quality evaluation. Nevertheless, d-ifferent indicators are correlated with each other, and common weight methods can not solve the problem. In fact, Structural Equation Model(SEM) can deal with this issue due to its statistical advantages, allowing error caused by measurement. In 2019, reference [20] assessed the drilling risk based on SEM and Monte Carlo(MC) to determine the relationship between indicators and find out the key factors. In this study, we constructed a comprehensive weight based on SEM and entropy weight to determine reasonable water quality weight.

Since each cloud can be represented by three fixed digital parameters, the similarity between different clouds according to their distribution and shape could be found, so that evaluation level based on the maximum similar cloud could be obtained. In 2004, Similar cloud and the measurement method was first mentioned in reference[21], which based on the distance of cloud droplets. This method was widely applicated in collaborative recommendations, similarity search and evaluation and so on. i.e.reference[21] calculated the collaborative recommendations similarity by the definition of cosine Angle in 2007; In 2011, reference [22] proposed the similar degree of cloud model based on the area of droplet expectation curve and maximum boundary curve. Nevertheless, reference[21] method has a little difference when the three parameters of cloud model are far from each other, and reference[22] method has had relatively complicated computation. Both of them could not measure the similarity of cloud overall. Fortunately, reference[23] claimed the hybrid similarity in 2018. This method was established from distance similarity and shape similarity, which were calculated from the three fixed parameters. So my study presented an improved hybrid similarity methodology, by calculating the arithmetic square root contribution rate of the product of two similarities, so that the convergent in distance similarity and shape similarity for different levels could be better.

In this paper, we proposed the cloud model based on SEM-Entropy weight and hybrid similarity. The structure of this paper is: Firstly, determining the indicators of the research based on monitored data, and calculating a comprehensive weight based on SEM and entropy weight methods. Secondly, constructing an evaluation system of the cloud model and calculating three fix parameters of standard cloud and comprehensive cloud based on combined weights. Thirdly, obtaining water quality level in the study areas according to the hybrid similarity between standard cloud and comprehensive cloud. Finally, applying this method to Minjiang River in China and validating it by comparing its results with other three methods.

II. Data study

The study focuses on the Minjiang river, which was the typical representation of polluted rivers in Sichuan province. Considered one part (Fig.1) of Minjiang river as an example for analysis, the main sources of pollutants in this fracture surface were the inflow from four tributaries Jinniu river, Simeng river, Tiequan river, the inflow of Dongfeng canal, the discharge of industrial wastewater, and the discharge of sewage treatment plant on the upstream. The employed values in this study consisted of six evaluation indices, including Chemical Oxygen Demand (COD), Biochemical Oxygen Demand (BOD), Index Manganese(IMn), Ammonium Nitrogen ($NH_3 - N$), Total Phosphorus (TP), Dissolved Oxygen (DO), and these data were monitored monthly from January 2011 to December 2016. The quality of these data satisfied the national standards, additionally, we used the z-score method to standardize these indices, thus avoiding inconsistency induced by measurement and calculation.

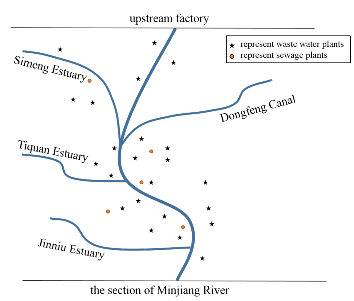

Fig. 1: The one part of Minjiang River.

III. Methodology

Realizing the status of water quality by consequence or uncertainty degrees, the cloud model could get the transformation between qualitative concept and quantitative data. The standard cloud, represented the cognition of qualitative concept water quality, to be obtained the similarity by comparison to comprehensive cloud, which weight was determined by means of SEM-Entropy weight methods, thus finally, obtaining the level according to the maximum similarity.

The cloud model based on SEM-Entropy weight hybrid similarity approach for water quality assessment could be illustrated in Fig.2, and summarized as below:

Step 1: elected suitable water quality indices and transfer these to standardized data;

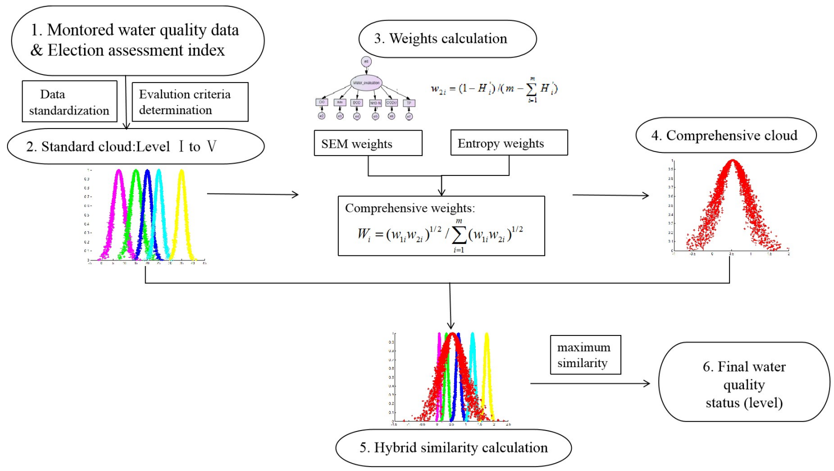

Fig. 2: Framework of the cloud model-based water quality assessment approach.

Step 2: determined water quality criteria (COD, BOD, IMn, NH3-N, TP, DO) and parameters (Ex, En, He) of each index, thus to obtain standard cloud model based on given levels (I,II,III,IV,V);

Step 3: calculated the indexs comprehensive weight based on SEM and Entropy weights;

Step 4: by the combination weight and each indexs parameters to determine the comprehensive cloud;

Step 5: substituted the parameters transformed from monitoring data into cloud models to obtain the distribution of cloud drop, and calculated the hybrid similarity considered distance and shape between comprehensive cloud and standard cloud;

Step 6: the water quality grade could be gained, based on the level with maximum hybrid similarity calculated from the cloud model.

A. Relevant weight method

1) Structural Equation Model: As a multivariate statistical analysis method, Structural Equation Model (SEM) was widely used in recent years, which used the covariance matrix of variables to analyze the relationship about variables and especially considered the errors among variables. An SEM was composed by measurement model and structural model, all of which were estimated simultaneously [24]. Fig.3 was a generic SEM schematic diagram used to illustrate the basic concepts in SEM. In Fig.3, X and Y represented measured variables, η represented endogenous latent variable, ξ represented exogenous latent variable; w_i were the coefficient between endogenous latent variable and measured variable, e_i were the residual errors.

An SEM measurement model was used to specify latent variables as linear functions of other variables in the system. When these other variables were observed, they took on the role of "indicators" of the latent constructs. In this way, SEM measurement models were similar to factor analysis, but there was a basic difference—all elements of the matrix defining the latent variables (factors) in terms of linear combinations of the observed variables take on non-zero values [24].

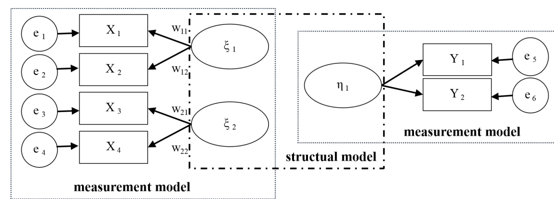

Fig. 3: A generic SEM schematic diagram.

Given indicator system in this paper, we just estimated the indicator weight of water quality by means of measurement model. Equation (1) was one measurement model in Fig.3, w_{1i} were the weights of latent variable ξ_1 , which was the water quality evaluation in this paper, X_i were the six indices actual monitoring data.

$$X_i = w_{1i}\xi_i + e_i \tag{1}$$

2) Entropy: SEM has been widely applied in the weight determination of risk assessment in various industries. It could avoid the influence of human subjective factors of AHP expert scoring. Meanwhile, due to its unique advantages, it could allow certain errors in the weight determination and found the relationship between factors. However, water quality assessment had much uncertainty factors, which could lead the complexity and instability. Regarded as a measurement of disorder or uncertainty of a system, the notion of "entropy" taken from the theoretical foundation of the modern information theory [25] has been introduced in hydrology and water quality, particularly in uncertainty analyses [26, 27].

In water quality evaluation, Entropy of the observed data under the ith criterion could be calculated by "(2)":

$$H_i = -\sum_{k=1}^{n} p_k ln p_k \tag{2}$$

where H_i represented the uncertainty of observed water quality data of one criterion with n potential intervals or statements; p_k was the frequency of the kth statement and if $p_k = 0$ then $0\ ln0 = 0$. Now the entropy-based weight of the ith criterion w_{2i} could be attained by H_i [28], as "(3)":

$$w_{2i} = (1 - H'_i)/(m - \sum_{i=1}^{m} H'_i) \tag{3}$$

where $H'_i = H_i/lnn$, m is the number of criteria.

For the feasibility of SEM weight in the water quality assessment, unlike the AHP-based weight, did not reflect the subjective importance of criteria, but indicated the internal convergence of data, and used entropy-based to demonstrate the relative severity of "competition" of each criterion. i.e. if the observed data had a central tendency which corresponded to a low entropy, the "competitiveness" of the criterion controlling the decision-making was less compared to others. In this manner, low entropy contributed to low weights [4].

3) Combination SEM Weight and Entropy Weight: Here, a comprehensive weight calculating algorithm coupled with entropy was proposed, which was expected to balance the potential uncertainty of the SEM approach. A hybrid SEM-Entropy weights algorithm of each indicator, obtained by using arithmetic square root contribution rate of the product of two weights, was given as "(4)":

$$W_i = (w_{1i}w_{2i})^{\frac{1}{2}} / \sum_{i=1}^{m} (w_{1i}w_{2i})^{\frac{1}{2}} \tag{4}$$

Where w_{1i} was the SEM weight of the i indices, w_{2i} was Entropy weight of the i indices, W_i was the final comprehensive weight.

B. Cloud Model

The cloud model, first proposed by Li et al. [15], was a type of transformation model that synthetically described the randomness and fuzziness of concepts and could implement the uncertain transformation between a qualitative concept and its quantitative instantiations.

1) Model parameters: The cloud model could effectively describe the overall quantitative property of a concept by the four numerical characteristics as follows:

Ex (Expectation) represented the mathematical expectation that the cloud drops belonged to a concept in the universe. It could be regarded as the most representative and typical sample of the qualitative concept.

En (Entropy) represented the uncertainty measurement of a qualitative concept. It was determined by both randomness and the fuzziness of the concept. In one aspect, as the measurement of randomness, En reflected the dispersing extent of the cloud drops and in the other aspect, it was also the measurement of fuzziness, representing the scope of the universe that could be accepted by the concept.

He (Hyper entropy) represented the uncertain degree of entropy En.

N (Number) represented the number of the repeat simulations. Note that in Fig.4, the x-axis represented the values of the water quality index, and the y-axis represented the certainty degree of a water quality grade. N was set to 2000 in this study to balance accuracy, robustness, and computational expense [1].

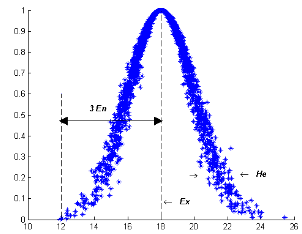

Fig. 4: The numerical characteristics of Cloud Model(N=2,000).

2) Standard cloud model: It was necessary to transform the surface water pollution level into a standard cloud model to assess water quality. For an indicator with bilateral constraints, the numerical characteristics of the standard cloud model could be calculated as "(5—7)":

$$Ex = (B_{max} + B_{min})/2 \tag{5}$$

$$En = (B_{max} - B_{min})/6 \tag{6}$$

$$He = k \times En \tag{7}$$

where B_{max} and B_{min} were the maximum and minimum of the concentration range of an water quality indicator, respectively, and k was a constant that changeed according to the randomness and fuzziness of different indicators. A larger He, as mentioned above, indicated greater randomness of assessment indicators; a smaller He suggested less randomness of assessment indicators and randomness that was more easily lost [29]. In this paper, k was assumed to be 0.1.

In this case, combined with standard water quality criteria (TABLE I), the three perimeters (Ex, En, and He) of each evaluation index were calculated according to "(5—7)". The modified equations were applicable for fixed intervals, noting that B_{max} for grade I of criterion COD, BOD was missing. Here, we referenced article [1] to solve this problem, results were shown in TABLE II.

TABLE II: Standard cloud model parameters of water quality grades of all criteria

Grade	**COD**			**BOD**			**IMn**		
	Ex	***En***	***He***	***Ex***	***En***	***He***	***Ex***	***En***	***He***
I	7.50	2.50	0.25	1.50	0.50	0.05	1.00	0.33	0.03
II	15.00	2.50	0.25	3.00	0.50	0.05	3.00	0.33	0.03
III	17.50	0.83	0.08	3.50	0.17	0.02	5.00	0.33	0.03
IV	25.00	1.67	0.17	5.00	0.33	0.03	8.00	0.67	0.07
V	35.00	1.67	0.17	8.00	0.67	0.07	12.50	0.83	0.08
Grade	$NH_3 - N$			**TP**			**DO**		
	Ex	***En***	***He***	***Ex***	***En***	***He***	***Ex***	***En***	***He***
I	0.08	0.03	0.00	0.01	0.00	0.00	11.25	1.25	0.13
II	0.33	0.06	0.01	0.06	0.01	0.00	6.75	0.25	0.03
III	0.75	0.08	0.01	0.15	0.15	0.00	5.50	0.17	0.03
IV	1.25	0.08	0.01	0.25	0.25	0.00	4.00	0.33	0.02
V	1.75	0.08	0.01	0.35	0.35	0.00	2.50	0.17	0.02

TABLE I: Quantitative boundaries of water quality grades of all criteria.

Evaluation indices	**COD** (mg/L)	**BOD** (mg/L)	**IMn** (mg/L)	NH_3-N (mg/L)	**TP** (mg/L)	**DO** (mg/L)
I	≤ 15	≤ 3	≤ 2	≤ 0.15	≤ 0.02	≥ 7.5
II	≤ 15	≤ 3	≤ 4	≤ 0.5	≤ 0.1	≥ 6
III	≤ 20	≤ 4	≤ 6	≤ 1	≤ 0.2	≥ 5
IV	≤ 30	≤ 6	≤ 10	≤ 1.5	≤ 0.3	≥ 0.3
V	≤ 40	≤ 10	≤ 15	≤ 2	≤ 0.4	≥ 2

3) Comprehensive Cloud Model: The comprehensive cloud [30] was based on the uncertainty of latent variable concepts. It extracted and combined the concepts of the indicator layer into the comprehensive concepts of the upper level to improve the abstraction of the concepts. This extraction process made the latent variables contained all the information in the concept of indicator layer, and had different intersections with the standard clouds of different levels.

According the numerical characteristics of six indicators, respectively combined with the comprehensive weight of water quality evaluation indices, to determine the comprehensive cloud numerical character $C_z(Ex_z,En_z,He_z)$ as "(8—10)":

$$Ex_z = \frac{W_1Ex_1En_1 + W_2Ex_2En_2 + \cdots + W_mEx_mEn_m}{W_1En_1 + W_2En_2 + \cdots + W_mEn_m} \quad (8)$$

$$En_z = W_1En_1 + W_2En_2 + \cdots + W_mEn_m \quad (9)$$

$$He_z = \frac{W_1En_1He_1 + W_2En_2He_2 + \cdots + W_mEn_mHe_m}{W_1En_1 + W_2En_2 + \cdots + W_mEn_m} \quad (10)$$

4) Similarity between Comprehensive Cloud Model and standard cloud model: When the cloud model of an water quality indicator assessment was achieved according to the given cloud drops, it was necessary to calculate the similarity to determine the level of the specific cloud belongs to. Fig.5 was the position relationship of standard cloud and comprehensive cloud, $y_z(x)$ represented comprehensive cloud, $y_1(x)$ and $y_2(x)$ represented the standard cloud of Grede I and Grade II,respectively. The expected area curve method was commonly used to judge which standard cloud was more similar to the comprehensive cloud. For the Fig.5, the overlapping area between $y_1(x)$ and $y_z(x)$ were nearly as the area between $y_2(x)$ and $y_z(x)$, but $y_1(x)$ and $y_2(x)$ had much difference in shape.

Therefore, considering the difference of cloud model‘s shape and distances in this paper, which to determine the similarity between the specific comprehensive cloud model and a standard comprehensive cloud model for judging water quality level.

- Firstly, the shape similarity, which measured the similarity of uncertainty between the specific comprehensive cloud model and a standard comprehensive cloud model, was in connection with En and He but wasn‘t influenced by Ex, equation (11) as follow:

$$Sim_s(C_i,C_j) = \frac{min(\sqrt{En_i^2+He_i^2})),(\sqrt{En_j^2+He_j^2})}{max(\sqrt{En_i^2+He_i^2}),(\sqrt{En_j^2+He_j^2})} \quad (11)$$

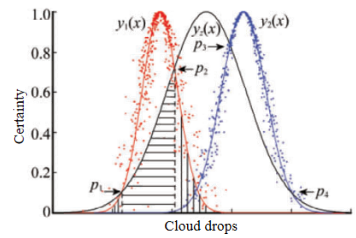

Fig. 5: The relationship between standard cloud and comprehensive cloud.

- Secondly, the distances similarity was only connected with Ex, the overlapping area of two expectation curve were $S(d)$, the distance similarity of indicator as "(12)":

$$Sim_d(C_i,C_j) = S(d)/S(0) \quad (12)$$

Where $d = |Ex_1 - Ex_2|$, $S(0)$ was the overlapping area of $d = 0$. $Sim_d(C_i,C_j)$ was a fitted curve $y = a \times exp(-((x-b)/c)2)$, by means of Gaussian curve to match, which came from Wangs article, the parameters of fitted curve y can be determine by seeking the table of fitted result, thus calculating the distance similarity.

- the improved hybrid similarity was calculated by using arithmetic square root contribution rate of the product of two similarities, so that the similarity in distance and shape could be better convergent, as "(13)":

$$Sim_c(C_i,C_j) = \frac{\sqrt{(Sim_s(C_i,C_j)) \times (Sim_d(C_i,C_j)}}{\sum_{i=1}^{5}\sqrt{(Sim_s(C_i,C_j)) \times (Sim_d(C_i,C_j)}} \quad (13)$$

For the index cloud, i was the level of standard cloud, j was the specific cloud s, for the comprehensive cloud, i was the level of standard comprehensive cloud and j was the comprehensive specific cloud z.

IV. Result and Discussion

A. Water Quality Indicator Assessment Results

Consider the water quality evaluation indicator of COD as an example. Its quantitative numerical characteristics were $(13.53, 3.5, 1.19)$ and its qualitative concept was shown in Fig.6 A, it mostly fell between Grade I ,Grade II and Grade III in the standard cloud models.

The similarity between the cloud model of COD and its corresponding standard cloud models, which could be calculated via "(11—13)" to the specific level belonged to, the TABLE III showed that $Sim_d(C_1,C_s) = 0.276$, $Sim_d(C_2,C_s) = 0.630$

and $Sim_d(C_3, C_s) = 0.157$, and others very small. The water quality evaluation indicator of COD belonged to Grade II in the standard cloud model according to the maximum similarity principle. That's to say, the quantitative assessment result of water quality evaluation of COD was Grade II. The qualitative assessment result of COD was between Grade I, Grade II and Grade III, and the quantitative assessment result of COD was Grade II. Therefore, combining qualitative and quantitative assessment results, the water quality evaluation level of COD was Grade II on the one part of Minjiang River.

From the TABLE III, the assessment results of water quality evaluation indicators of BOD, IMn and DO were also Grade II and $NH_3 - N$ was Grade III, TP was Grade IV in the a part of Minjiang river. The water quality evaluation in 2011 to 2016 were main affected by $NH_3 - N$ and TP, thus many remediation measures might be taken to improve water quality.

TABLE III: Similarity of water quality indicators between specific cloud and standard cloud

Index	I	II	III	IV	V	Index Grade
COD	0.276	**0.630**	0.157	0.011	0.000	**II**
BOD	0.294	**0.479**	0.120	0.025	0.000	**II**
IMn	0.034	**0.315**	0.156	0.003	0.000	**II**
$NH_3 - N$	0.054	0.141	**0.192**	0.063	0.005	**III**
TP	0.002	0.023	0.142	**0.199**	0.127	**IV**
DO	0.018	**0.195**	0.155	0.133	0.004	**II**

B. Water Quality Comprehensive Cloud Assessment Results

When the cloud model of an water quality indicator assessment was achieved according to the given cloud drops, it was necessary to calculate the similarity to determine the water quality level of specific comprehensive cloud model belongs to. As mentioned before, we obtained the comprehensive weight at the begining.

1) Analysis of the Weight: As described above, this study utilized SEM and Entropy two weight methods to calculate the comprehensive weight.

For the determine of SEM weight, it‘s essential to test the significance of the parameters, thus to validate the rationality of indicators weight in measurement model. To constantly amendment fitting index by AMOS software, the final test results of various fitness indexes of SEM established were shown in TABLE IV. The results showed that the multi-item fitting index of the model meeted the requirements, and the modified model had a higher fitting degree than original model, so the modified model was adopted, the results as Fig.7. From Fig.7 could be seen, the observation index BOD had a certain negative reaction to $NH_3 - N$, which could not be obtained from other methods such as AHP. This effect also passed the significance test, to may control $NH_3 - N$ with more BOD for water quality study. Then, the Entropy weight was determine by means of the "(2—3)", thus the comprehensive weight was calculated by "(4)" combining SEM weight and Entropy weight, and the all results were shown in TABLE V.

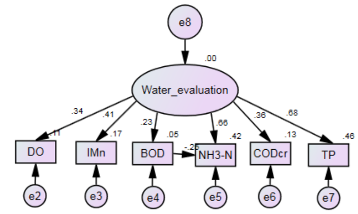

Fig. 7: the weight results of modified model in SEM.

TABLE IV: Test the significance of the measurement model of water quality evaluation

Model	CMIN	DF	P	CMIN/DF	RMR
origin	7.840	8	.062	.980	.053
modified	10.033	8	.263	1.254	.031
Model	**GFI**	**AGFI**	**NFI**	**CFI**	**RMSEA**
origin	.912	.844	.752	.896	.085
modified	.962	.900	.809	.946	.056

We could find that the SEM weight showed that $NH_3 - N$ and TP were the two crucial indicators among six factors, while BOD was considered the least crucial. For the Entropy weight, $NH_3 - N$ and BOD were regarded as the most important, and DO as the least crucial indexes. Similar to the SEM weight, the comprehensive weight showed that $NH_3 - N$ and TP were the two crucial indicators, followed by IMn, which was in accordance with the actual situation, indicating that the water quality evaluation of this part of Minjiang River was mainly affected by $NH_3 - N$ and TP. Since the Entropy weight not considered the differences of practical situations of each index, causing dissatisfied results that deviated from the decision makers‘ subjective cognition. While SEM transformed actual monitoring data into five grades that based on the water quality criteria,not only considered the measured results, but also considered the actual classification situation, and was not even affected by the subjective factors of expert scoring.

After investigating the actual polluted situation in the study area, it was learned that the monitoring part was located downtown, surrounded by five sewage plants and some waste water plants from the upstream, thus contributing to relatively serious levels of pollution. Although BOD should be larger in our subjective cognition because it was the most intuitive embodiment of organic pollutants, the value of 0.086 in SEM weight very far from actual conditions, but the Entropy weight could balance that. Therefore, both SEM weight and Entropy weight were complementary, the determine of comprehensive weight with the advantages of two weights, should be rational to resolve the relative assessment problem.

2) Analysis of Water Quality and Comparison with Other Methods: The water quality was classified by five grades in this study, we assumed that the level, no more than Grade III, was good, which lower grade means better water quality. By combining with the comprehensive weight and "(8—10)" to determine the standard comprehensive numerical character-

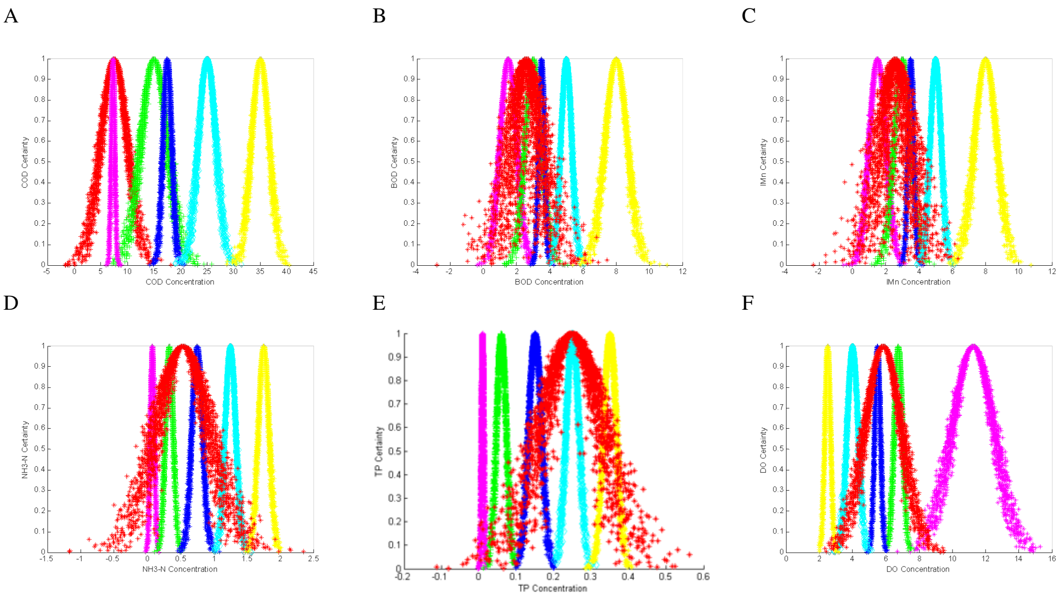

Fig. 6: Water quality evaluation indicator cloud models and corresponding standard cloud models. Pink, green, blue, cyan and yellow standard cloud models indicate Grade I, Grade II , Grade III, Grade IV and Grade V, respectively. All water quality evaluation indicator cloud models marked in red.

TABLE V: Index weight obtained by the three methods

Index	Entropy Weight	SEM Weight	Comprehensive Weight
COD	0.090	0.134	0.118
BOD	**0.148**	0.086	0.121
IMn	0.085	0.153	0.122
NH_3-N	**0.531**	**0.246**	**0.387**
TP	0.107	**0.254**	**0.177**
DO	0.039	0.127	0.075

istics $C_z(Ex_z, En_z, He_z)$, and then calculate the similarity between specific comprehensive cloud and standard comprehensive cloud. The validity of the proposed cloud model based on SEM-Entropy weight and hybrid similarity approach was assessed by comparison with other three methods: the Single Factor(SF) method, the Comprehensive Pollution Index (CPI) method and Grey Relational Analysis(GRA). Results of the various methods were shown in Fig.8.

During 2011 to 2016, it could be seen from the corresponding evaluation results of each month in Fig.8, there were 44 months (61.11 %) in CHS compared to 38 months (52.78%) in GRA at grade II, and 19 months (26.39%) in CHS compared to 24 months (33.33%) in GRA at grade III, that's to say, most of the grades between CHS and GRA were in accordance with each other and they were belonging to grade I to grade III, meaning Minjiang River had a good water quality, to some extend, a total of 62 months (86.11%) were in the same grade. There also existed different results in three months (Mar 2011, Apr 2011, Apr 2013) because the GRA results were only evaluated by fuzziness, while CHS contained the information of both fuzziness and randomness, so that some data randomness of indicators was missing using method GRA.

Methods	Time	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
SF	2011	III	IV	V	IV	IV	III	III	III	III	III	III	III
	2012	IV	III	IV	IV	V	IV	III	III	III	III	III	III
	2013	V	V	V	V	V	IV	V	III	III	III	III	III
	2014	IV	III	III	III	V	V	III	IV	III	III	V	V
	2015	V	V	V	V	V	V	III	III	IV	IV	IV	IV
	2016	IV	IV	IV	V	IV	IV	IV	IV	IV	IV	IV	IV
CPI	2011	IV	IV	IV	IV	III	III	III	III	III	III	III	IV
	2012	III	IV	IV	IV	III	III	III	III	III	III	III	III
	2013	IV	III	V	IV	III	III	III	III	III	III	III	III
	2014	III	III	IV	IV	III	III	III	III	III	III	III	III
	2015	III	III	III	III	III	III	III	III	II	III	III	III
	2016	IV	IV	IV	IV	III	III	III	III	III	III	III	IV
CHS	2011	II	II	III	III	IV	III	II	II	II	II	II	II
	2012	II	II	III	II	IV	I	II	III	III	II	I	III
	2013	II	II	III	II	IV	I	II	I	II	II	II	II
	2014	II	II	II	II	III	II	II	II	I	II	III	III
	2015	II	II	II	II	II	III	II	I	II	II	III	II
	2016	III	III	II	II	III	II	II	II	III	II	III	III
GRA	2011	III	III	IV	IV	IV	III	III	II	I	II	II	III
	2012	II	III	III	III	IV	II	II	II	II	II	II	III
	2013	III	III	II	IV	IV	III	II	III	II	II	II	III
	2014	II	III	III	III	III	II	III	II	II	II	II	II
	2015	II	III	II	III	II	II	II	II	II	II	II	II
	2016	III	III	III	III	III	II	II	II	II	II	II	III

Fig. 8: Comparison of water quality grades using various assessment methods.

Fig.8 also showed that the other two methods results of SF and CPI were even worse than the GRA, which were determined by the methods themself and not considered the uncertainty of water quality. In general, SF results were major determined by the worst level of all indices so that the water quality trait more poor grades, and the CPI based on the SF to determine the mean value of all polluted indices, thus judging the water quality level. CPI could make higher composite grade, though only one index is high and the others are low, leading to unreasonable water quality evaluation in final.

Generally, the cloud model based on SEM-Entropy weight and hybrid similarity could not only obtain more feasible water quality result by the advantages of the randomness and fuzziness using the similarity but also provide more reasonable idea for the water quality levels.

V. Conclusions

In this study, to begin with, we combined SEM and Entropy weight methods to determine a comprehensive weight of water quality evaluation; then constructed a hybrid similarity by the comprehensive weight to determine level of water quality; followed by comparing it with three common methods using monitored data from Minjiang River in China. The final results showed that proposed method was feasible and reasonable. Here were the mean reasons:

- The observation index BOD had a certain negative reaction to $NH_3 - N$ in SEM weight, which could not be obtained from other methods such as AHP. This effect also passed the significance test, to may control $NH_3 - N$ with more BOD for water quality study.
- The comprehensive weight based on SEM-Entropy weight showed that $NH_3 - N$ and TP were the two crucial indicators, while DO was considered the least crucial. Combining the advantages of two weights, the comprehensive weight should be more rational to resolve relative problem.
- During 2011 to 2016, most of the grades between CHS and GRA are in accordance with each other and they were belonging to grade I to grade III, meaning Minjiang River had a good water quality. There also existed different results because the GRA results were only evaluated by fuzziness, while CHS contained the information of both fuzziness and randomness, so that some data randomness of indicators was missing using method GRA. And the other two methods SF and CPI were even worse than the GRA, because of not considering the uncertainty of water quality.

Acknowledgment

The authors thankfully acknowledge support provided by the subject of the Major Projects of National Science & Technology (2016ZX05020-006).

References

[1] B. Yang, C. Lai, X. Chen, X. Wu, and Y. He, "Surface water quality evaluation based on a game theory-based cloud model," Water, vol. 10, no. 4, p. 510, 2018.

[2] A. R. Jackson and J. M. Jackson, " Environmental science: The natural environment and human impact," Pearson Education, 2000.

[3] A. S. Goudie, " The Human Impact on the Natural Environment: Past, Present, and Future," John Wiley & Sons, 2018.

[4] D. Wang, D. Liu, H. Ding, V. P. Singh, Y. Wang, X. Zeng, J. Wu, and L. Wang, "A cloud model-based approach for water quality assessment," Environmental research, vol. 148, pp. 24–35, 2016.

[5] Y. Kong, Z. Pang, C. Wu, and M. Jing, "A comparative study of single factor and multivariate statistical methods for surface water quality assessment," Water Policy, vol. 16, no. 1, pp. 157–167, 2014.

[6] H.l. YIN and Z.-x. XU, "Comparative study on typical river comprehensive water quality assessment methods," Resources and Environment in the Yangtze Basin, vol. 5, no. 011, 2008.

[7] H. Zhu and X. Lu, "An integrative model to assess water quality in china's lake taihu: Comparing single-factor and multifactor assessments," Integrated environmental assessment and management, vol. 15, 2019.

[8] G. Matta, P. Naik, A. Kumar, L. Gjyli, A. Tiwari, and J. Machell, "Comparative study on seasonal variation in hydro-chemical parameters of ganga river water using comprehensive pollution index (CPI) at rishikesh (uttarakhand) India," Desalination and Water Treatment, vol. 118, pp. 87–95, 2018.

[9] Y. Qian, K. W. Migliaccio, Y. Wan, and Y. Li, "Surface water quality evaluation using multivariate methods and a new water quality index in the indian river lagoon, florida," Water Resources Research, 2007.

[10] I. P. Kotti, G. K. Sylaios, and V. A. Tsihrintzis, "Fuzzy logic models for bod removal prediction in free-water surface constructed wetlands," Ecological engineering, vol. 51, pp. 66–74, 2013.

[11] I. Primpas, G. Tsirtsis, M. Karydis, and G. D. Kokkoris, "Principal component analysis: Development of a multivariate index for assessing eutrophication according to the european water framework directive," Ecological Indicators, vol. 10, no. 2, pp. 178–183, 2010.

[12] S. Dahiya, B. Singh, S. Gaur, V. K. Garg and H. S. "Kushwaha, Analysis of groundwater quality using fuzzy synthetic evaluation," Journal of Hazardous Materials, vol. 147,no. 3, pp. 938–946, 2007.

[13] J.T. Kuo, Y.Y. Wang, and W.S. Lung, "A hybrid neural–genetic algorithm for reservoir water quality management," Water research, vol. 40, no. 7, pp. 1367–1376, 2006.

[14] C.L. Zhao, X.J. Cao, X.J. Xie, X.H. Yuan and B.Wang, "Environmental Aspects of Drilling Operation Risks: A Probability Prediction based on Fuzzy Time Series," Ekoloji, vol. 27, no.106, pp. 1493–1502, 2018.

[15] W. Ocampo-Duque, C. Osorio, C. Piamba, M. Schuhmacher, and J. L. Domingo, "Water quality analysis in rivers with non-parametric probability distributions and fuzzy inference systems: Application to the cauca river, colombia," Environment international, vol. 52, pp. 17–28, 2013.

[16] D.Y. Li, C. Liu, and W. Gan, "A new cognitive model: Cloud model," International Journal of Intelligent Systems, vol. 24, pp. 357–375, 2009.

[17] J.Q. Wang, L. Peng, H.Y. Zhang, and X.H. Chen, "Method of multi-criteria group decision-making based on cloud aggregation operators with linguistic information," Information Sciences, pp. 177–191, 2014.

[18] Q. Zou, J. Zhou, C. Zhou, L. Song, and J. Guo, "Comprehensive flood risk assessment based on set pair analysis-variable fuzzy sets model and fuzzy ahp," Stochastic Environmental Research and Risk Assessment, vol. 27, no. 2, pp. 525–546, 2013.

[19] F. Jin, L. Pei, H. Chen, and L. Zhou, "Interval-valued intuitionistic fuzzy continuous weighted entropy and its application to multi-criteria fuzzy group decision making," Knowledge-Based Systems, pp. 132–141, 2014.

[20] C.L. Zhao, H.M. Yin, B. Wang, X.Y. Fan, and H. Wu, "Risk assessment of drilling site based on Structural Equation and Monte Carlo method,"Natural gas industry, 85-94.

[21] Y. Zhang, D. Zhao, D.Y. Li, "The Similar cloud and the measurement method," Information and Control, vol. 33, pp. 129–132, 2004.

[22] H.L. Li, C.H. Guo, and W.R. Qiu, "Similarity measurement between normal cloud models," Dianzi Xuebao(Acta Electronica Sinica), vol. 39, no. 11, pp. 2561–2567, 2011.

[23] J. Wang, J.J. Zhu, X.D. Liu, "An integrated similarity measure method for normal cloud model based on shape and distance," Systems Engineering-Theory & Practice, vol. 37, pp. 742–751, 2017.

[24] T. F. Golob, "Structural equation modeling for travel behavior research," Transportation Research Part B: Methodological, pp. 1–25, 2003.

[25] C. E. Shannon, "A mathematical theory of communication," Bell system technical journal, vol. 27, no. 3, pp. 379–423, 1948.

[26] V. P. Singh, H. Cui, and A. R. Byrd, "Derivation of rating curve by the tsallis entropy," Journal of hydrology, vol. 513, pp. 342–352, 2014.

[27] V. P. Singh, "Entropy theory in hydrologic science and engineering," McGraw-Hill Education New York, 2015.

[28] J.G. Zhang, V.P.Singh, "Entropy-Theory and Applications," China Water & Power Press, Beijing, pp. 79C-80, 2012.

[29] D. Liu, D. Wang, J. Wu, Y. Wang, L. Wang, X. Zou, Y. Chen, and X. Chen, "A risk assessment method based on rbf artificial neural network-cloud model for urban water hazard," Journal of Intelligent & Fuzzy Systems, vol. 27, no. 5, pp. 2409–2416, 2014.

[30] X. Cha, S.H. Ni, C. Xie, P. Zhang, "Indirect computation approach of cloud model similarity based on conception skipping," Systems Engineering and Electronics, vol. 37, pp. 1676–1682, 2015.