The fuzzy comprehensive evaluation (FCE) and the principal component analysis (PCA) model simulation and its applications in water quality assessment of Nansi Lake Basin, China

Shiguo Xu[1*], **Yixiao Cui**[1*], **Chuanxi Yang**[1*], **Shujing Wei**[2], **Wenping Dong**[3], **Lihui Huang**[4], **Changqing Liu**[5], **Zongming Ren**[1†], **Weiliang Wang**[1†]

[1]*Institute of Environment and Ecology, Shandong Normal University, Jinan 250014, P. R. China*
[2]*Jinan Engineering Vocational Technical College, Jinan 250200, P. R. China*
[3]*Shandong Academy of Environmental Science, Jinan 250013, P. R. China*
[4]*Shandong Provincial Key Laboratory of Water Pollution Control and Resource Reuse, School of Environmental Science and Engineering, Shandong University, Jinan, 250100, P. R. China*
[5]*Qingdao Technological University, Qingdao 266033, P. R. China*
These authors contributed equally to this work.

ABSTRACT

The Fuzzy Comprehensive Evaluation (FCE) and the Principal Component Analysis (PCA) were simulated to assess water quality of the Nansi Lake Basin, China. The membership functions were established via the Nor-Half Sinusoidal Distribution Method, and the weight was calculated via the Exceeding Standard Multiple Method. To enhance the efficiency of extracting principal pollutant, the eigenequation was solved through the Jacobi Method, and the principal components were extracted based on eigenvalue, contribution ratio, accumulating contribution ratio, principal component loading and score. Water quality classification based on "National Surface Water Environmental Quality Standards of China (GB3838-2002) was used to assess the water quality. Considering the difference of the temporal and spatial distribution in average, water quality of Level I was 28.9%, 28.1%, 25.1%, 25.6%, respectively in spring, summer, autumn, and winter, which suggested that water quality in spring and summer was better than in autumn and winter. The order of water quality was Zhaoyang Lake (Level I) > Nanyang Lake (Level I) > Dushan Lake (Level III) > Weishan Lake (Level III and IV). There were four extracted principal components that can replace the fourteen pollutant indexes for assessing water quality. According to the annual mean data of the 1st principal components, the most important pollutions were heavy metals, including As (0.933), Hg (0.931), Cd (0.929), Cr(VI) (0.926), Pb (0.925), and Cu (0.534). It is proved that the combined FCE-PCA model could provide valuable information in the water quality assessment for the Nansi Lake Basin.

Keywords: Accumulating contribution ratio, Fuzzy comprehensive evaluation, Nansi Lake Basin, Principal components, Principal component analysis, Water quality assessment

1. Introduction

Rivers and lakes play a unique role for societal development through provision (e.g. products and food), support (e.g. wastewater processing and supply of clean water) and enrichment (e.g. aesthetic, recreational and cultural) services. The growing demand for freshwater resources to sustain human activities, coupled with the adverse effects of the same activity, such as the discharge of industrial wastewater and domestic sewage, are likely to cause a crisis in the near future if water resources are not appropriately managed [1, 2]. The shortage of water resources and water pollution is a very serious problem in the arid and semi-arid areas of Northwestern China.

Water quality management involves water quality monitoring,

analysis, assessment, and reporting [3, 4]. Water quality assessment is a challenge within the study of the environment. Proper identification of water quality status in a river or lake system based on limited observations is essential for meeting the goals of environmental management [5]. Numerous methods have been proposed in literature to assess water quality, including expert assessment, index assessment, neural networks, and grey clustering [6-8]. Considering that pollution degree of water is a vague concept with inherent imprecision and unclear of classification criteria and the boundaries between different classes of water quality, difficulties of water quality classification and assessment always exist in conventional assessment methodologies such as Water Quality Index (WQI) when describing integrated water quality [9]. Conventional methods for water quality assessment do not consider the uncertainties involved either in measurement of water quality parameters or in the limits provided by regulatory bodies, meanwhile, they also use crisp set and concentration values, which are close to or far from the limits, which are included in the same class [10]. This unclearness has led some environmental researchers to look for advanced assessment methods based on the Fuzzy Comprehensive Evaluation (FCE) and the Principal Component Analysis (PCA).

FCE is designed to interpret the uncertainties of water quality assessment [11]. It comprehensively evaluates the contributions of various pollutants according to predetermined weights, and decreases the fuzziness by using membership functions [12]. Therefore, the sensitivity is higher than other index evaluation techniques [13]. A well designed FCE may be capable of covering the uncertainties in the sampling and analysis process, comparing sampling results to quality standards for each parameter, and summarizing individual parameter values [14]. FCE has been widely used in environmental quality assessment, and has been proven effective in solving problems of fuzzy boundaries and in controlling the effect of monitoring errors [15].

PCA is a multivariate statistical technique used to identify important components or factors that explain most of the variances of a system. It is designed to reduce the number of variables to a small number of indexes while attempting to preserve the relationships present in the original data [16]. In recent years, the PCA model has been applied to a variety of environmental issues, including evaluation of ground water monitoring wells, interpretation of groundwater hydrographs, examination of spatial and temporal patterns of heavy metals contamination and identification of herbicide species related to hydrological conditions [17]. Winter et al. applied the PCA method to investigate the areal distribution of various types of water level fluctuation patterns and to determine if fewer wells could be measured while still achieving effective long-term monitoring goals at four small research sites in the USA, the results showed that the PCA technique was very useful in summarizing information from large data sets to select long-term monitoring wells [18]. Mustapha et al. [19] used the PCA method to assess water quality of the Jakara River Basin; these authors made the contributions to the existing knowledge on the spatial variations of surface water quality and are believed to serve as a baseline data for further studies. Additional, the PCA method has been used to estimate spatial and temporal patterns of heavy metal and organic pollutant, and these results have provided good

examples of the effective applications of PCA. However, there are few documented examples of the evaluation of the high dynamic and complex water quality monitoring in river, lake or drainage basin [20]. Therefore, the veracity of water quality assessment for Nansi Lake Basin could be enhanced via the combination between FCE and PCA [21].

The aims of this study are to demonstrate the application of FCE-PCA model to assess the water quality and to extract the principal pollutants of the Nansi Lake Basin, and to evaluate the importance of various water quality parameters. The specific objectives are: (1) to present detailed procedures of FCE-PCA model, (2) to assess the water quality results of the Nansi Lake Basin, and (3) to extract the principal pollutants that are most important in assessing variations in the Nansi Lake Basin water quality.

2. Experimental Design

2.1. Background of the Nansi Lake Basin

The South-to-North Water Diversion Project was put forward by the government of China. Nansi Lake is a buffer lake of the South-to-North Water Diversion Project (east route), so water quality of the Nansi Lake Basin (34°27′N-35°20′N, 116°34′E-117°21′E) has to be better than that of Class III in accordance with the "National Surface Water Environmental Quality Standards of China (Chinese Environmental Protection Agency, GB3838-2002)." The main pollution sources of the Nansi Lake Basin are industrial pollution sources, agricultural pollution sources and domestic pollution sources from three cities residing within the Shandong Province: Zaozhuang, Jining, and Heze [22]. It is important to manage and evaluate the water quality of the Nansi Lake Basin, based on the water quality requirement of the South-to-North Water Diversion Project, because if we do not then the quality of life of the surrounding area and those within its reach will surely diminish. The Nansi Lake Basin is shown in Fig. 1. As the largest shallow freshwater lake in Shandong Province, China, it consists of four sublakes from north to south, Nanyang Lake, Dushan Lake, Zhaoyang Lake, and Weishan Lake, with a whole catchment area of 30.4×10^3 km². The total surface area of the lake is $1..2 \times 10^3$ km² with 126 km in length from north to south and 5-25 km in width from east to west, and the average depth is only 1.46 m [23].

There are 3.17×10^4 square kilometers in the Nansi Lake Basin, mainly including Zaozhuang, Jining, and Heze sity. in order to evaluate the water quality of the Nansi Lake Basin, the five monitoring sections (Fig. 1) and the fourteen monitoring indexes were ensured with hydrological and hydraulic, basin environment, pollution sources and land use, and the indexes were estimated from the earlier studies on the point source pollution and water quality variation trends in the Nansi Lake Basin from 2002 to 2012 [24]. The Models were improved for scientific simulation at the same time.

Based on distribution characteristics of industrial enterprise for the Nansi Lake Basin, the fourteen monitoring indexes were selected as assessment parameters to form an assessment factor set U, $U=\{$TN, Oils, TP, BOD, NH₃-N, V-phenol, COD$_{Cr}$, Cr(VI), Hg, Pb, Cd, As, CN, Cu$\}$. An assessment criteria set V was also established
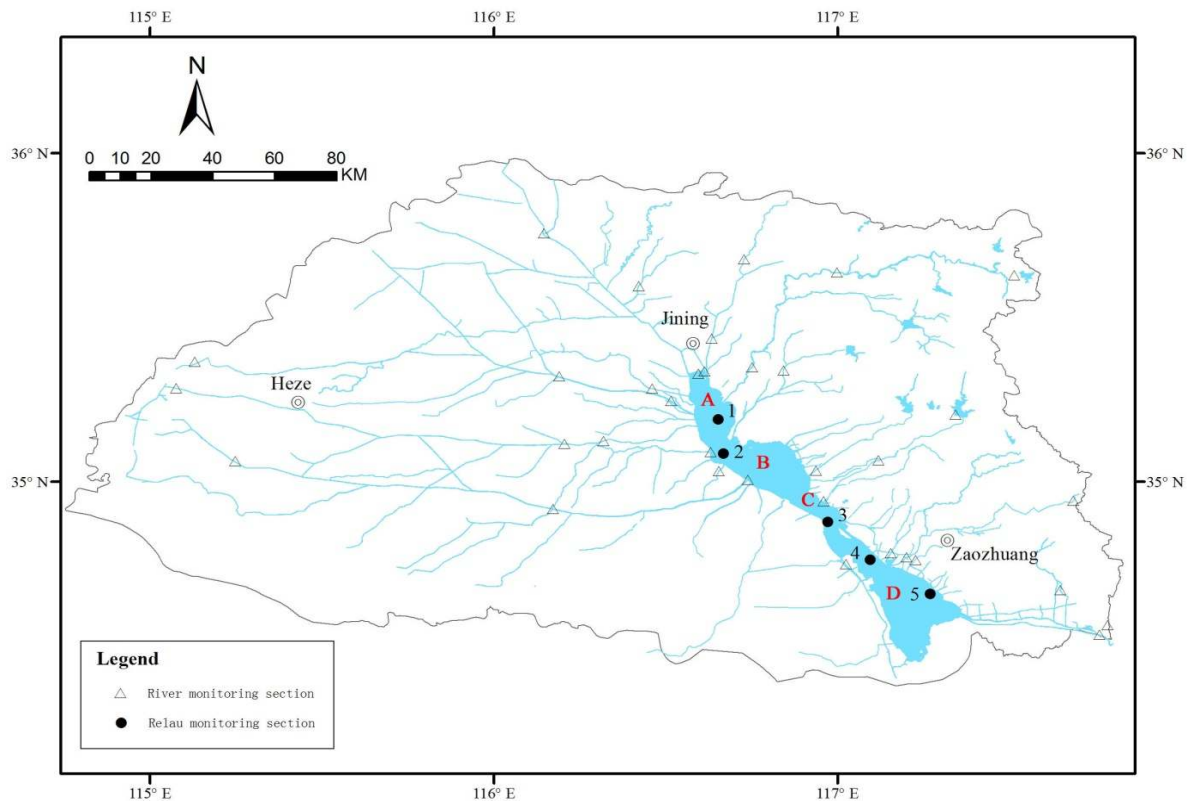
Fig. 1. Nansi Lake Basin and Monitoring Sections. △ represent river monitoring section and ● represent relau monitoring section. A: Nanyang sublake; B: Dushan sublake; C: Zhaoyang sublake; D: Weishan sublake.

**Table 1.** Assessment Factor Set and Assessment Criteria Set

Assessment parameters	I	II	III	IV	V
TN (mg/L)	0.2	0.5	1.0	1.5	2.0
Oils (mg/L)	0.05	0.05	0.05	0.5	1.0
TP (mg/L)	0.01	0.025	0.05	0.1	0.2
BOD (mg/L)	3	3	4	6	10
NH3-N (mg/L)	0.15	0.5	1.0	1.5	2.0
V-phenol (mg/L)	0.002	0.002	0.005	0.01	0.1
CODCr (mg/L)	15	15	20	30	40
Cr(VI) (mg/L)	0.01	0.05	0.05	0.05	0.1
Hg (mg/L)	0.00005	0.00005	0.0001	0.001	0.001
Pb (mg/L)	0.01	0.01	0.05	0.05	0.1
Cd (mg/L)	0.001	0.005	0.005	0.005	0.01
As (mg/L)	0.05	0.05	0.05	0.1	0.1
CN- (mg/L)	0.005	0.05	0.2	0.2	0.2
Cu (mg/L)	0.01	1.0	1.0	1.0	1.0

according to the National Surface Water Environmental Quality Standards of China (Chinese Environmental Protection Agency, GB3838-2002; Table 1).

2.2. The Design of AHP-FCE Model

In the design of the new model for Nansi Lake Basin, the advantages of multi-objective, multi-standard, and structurelessness of AHP were able to avoid one-sidedness and inconsistency resulting from environmental risk identification and evaluation index system. Due to the complexity of environmental risk, fuzzy demarcation, and difficulty in describing accurate scales, secondary-level FCE was designed to achieve comprehensive dynamic grading evaluation of environment risk. Further, the comprehensive dynamic risk classification evaluation methods were realized based on the AHP-FCE model.
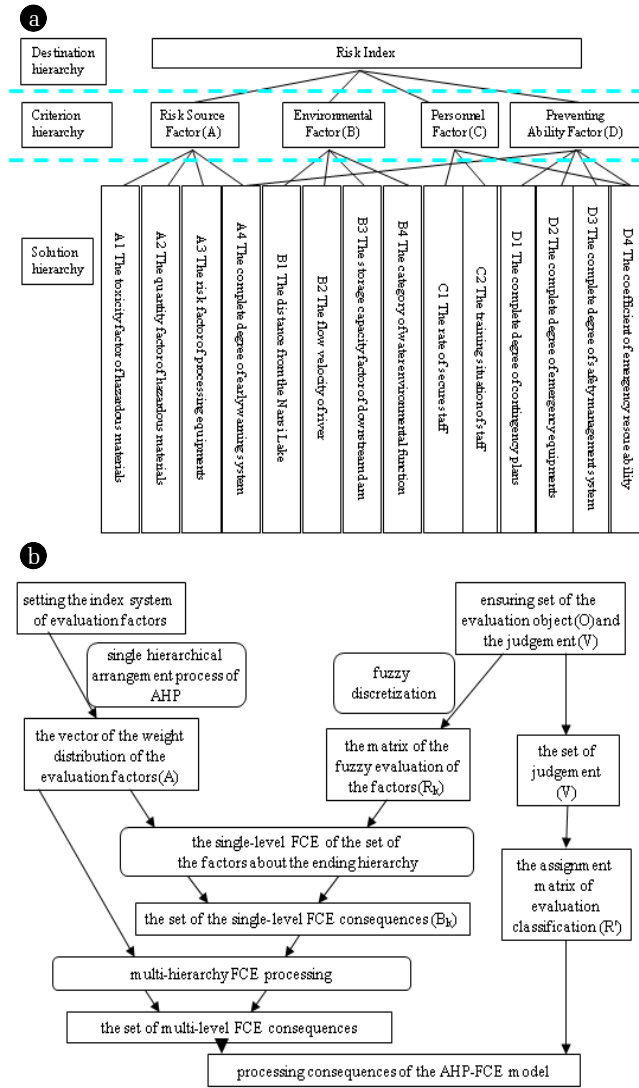
Fig. 2. The design of AHP-FCE model for Nansi Lake Basin. (a) The structure of environmental hazard identification and evaluation system. (b) The processing of AHP-FCE model.

According to the characteristics of the hazard installations at Nansi Lake Basin, factors that would affect evaluation results include: a risk source factor, an environmental factor, a human factor, and a preventing ability factor (Fig. 2). Fourteen specific solution hierarchy factors are also listed.

The risk index evaluation system and method system was constructed based on risk source and industrial characteristics, basin characteristics of the hydrological and hydraulic, and a consideration of the terrain of Nansi Lake Basin, site of risk source, developing process of risk, and the controlling process of risk.

2.3. Model Description

Simple fuzzy classification (SFC), fuzzy similarity method (FSM) and FCE are all subtitles of fuzzy synthetic evaluation (FSE). FCE has been used recently by a number of researchers in various environmental areas [25]. Since 1965, when the FCE model was first put forward by L.A. Zadeh with Fuzzy Set, fuzzy mathematics has become an indispensable part of mathematical science and has been widely used in social and natural science fields, including psychics, autocontrol, bioengineering and environmental science [26, 27]. Based on the complexity and uncertainty of water body environment, the fuzzification of water quality grade and water quality standard, William Silver was the first to assess seawater environment via fuzzy mathematics theory [28]. Following his lead, domestic and overseas scholars began evaluating water environment quality with the usage of the FCE model.

The PCA model practically transforms the original data matrix into a product of two matrices, one of which contains the information about the samples and the other about the variables. The principle components are the uncorrelated variables obtained by multiplying original correlated variables with eigenvector and are weighted linear combinations of the original variables. That is to say that the PCA model provides information on the most meaningful parameters, which describe whole data set affording data reduction with minimum loss of the original information. The PCA model has been widely used in water quality assessment of rivers, diggings, underground water reservoirs and lakes [29].

2.2.1. Fuzzy comprehensive evaluation model

The following procedure describes the FCE model.

(a) Select assessment parameters and establish assessment criteria: It is crucial to select representative, rational, and accurate water quality assessment parameters to form an assessment factor set U based on the actual local situation. This is expressed as

$$U=\{u_1,\ u_2,\ u_3,\ ...,\ u_n\} \tag{1}$$

where n is the number of selected assessment parameters. The assessment criteria set V is established from National Surface Water Environmental Quality Standards of China. This is expressed as

$$V=\{v_1,\ v_2,\ v_3,\ ...,\ v_m\} \tag{2}$$

where m is the number of assessment criteria categories.

The weights set A is established with different degree of importance. This is expressed as

$$A=(a_1,\ a_2,\ a_3,\ ...,\ a_n) \tag{3}$$

$$\sum_{i=1}^{n} a_i = 1, a_i \geq 0, \quad i = 1, 2, ..., n \tag{4}$$

(b) Single-factor fuzzy evaluation: Single-factor fuzzy evaluation is defined as ensuring the membership between the evaluation object and the assessment criteria set V_j based on the ith factor u_i. The single-factor fuzzy evaluation set R_i is expressed as

$$R_i=(r_{i1},\ r_{i2},\ ...,\ r_{im}) \tag{5}$$

Thus, the single-factor fuzzy evaluation sets of all factors can be obtained as follows:

$$R_{1i} = (r_{11}, \quad r_{12}, \quad ..., \quad r_{1m})$$
$$R_{2i} = (r_{21}, \quad r_{22}, \quad ..., \quad r_{2m})$$
$$R_{ni} = (r_{n1}, \quad r_{n2}, \quad ..., \quad r_{nm}) \tag{6}$$

Based on the above single-factor fuzzy evaluation set, the single-factor fuzzy evaluation matrix R is expressed as

$$R = \begin{bmatrix} r_{11}, & r_{11}, & \cdots & r_{1m} \\ r_{21}, & r_{22}, & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{m1}, & r_{n2}, & \cdots & r_{nm} \end{bmatrix} \tag{7}$$

where r_{ij} (i = 1, 2, …, n; j = 1, 2, …, m) is the membership degree of the ith assessment parameter at the jth level.

(c) Fuzzy Comprehensive Evaluation: Based on single-factor fuzzy evaluation lack of the interaction between evaluation factors and evaluation object, the omnibus evaluation set B is expressed as

$$B = A \bullet R = (a_1, a_2, \cdots, a_n) \begin{bmatrix} r_{11}, & r_{11}, & \cdots & r_{1m} \\ r_{21}, & r_{22}, & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1}, & r_{n2}, & \cdots & r_{nm} \end{bmatrix} = (b_1, b_2, \cdots, b_n) \tag{8}$$

where ith row is the influence extent between ith factor and all evaluation factors, and jth column is the influence extent between jth factor and all evaluation factors. The results can reasonably represent the comprehensive influence of all factors.

2.2.2. Principal component analysis model

The following procedure describes the PCA model.

(a) Establish original data matrix X: The original data matrix X is expressed as

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \tag{9}$$

where m is the sample number, and n is factor number. To eliminate the influence of dimension and order of magnitudes, the original data matrix X was normalized via the Z-Score Method. The normalization is expressed as

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \tag{10}$$

where $\bar{x}_j$ is the mean number of jth samples, and S_j is the standard deviation of jth samples. The formulas of $\bar{x}_j$ and S_j are expressed as

$$\bar{x}_j = \frac{1}{m} \sum_{i=1}^{m} x_{ij} \tag{11}$$

$$S_j = \sqrt{\frac{1}{m-1} \sum_{i=1}^{m} (x_{ij} - x_j)^2} (\tag{12}$$

(b) Compute correlation coefficient and correlation coefficient matrix: R_{ij} is the correlation coefficient between X_i and X_j, and it is expressed as

$$r_{ij} = \frac{\sum_{k-1}^{n} (x_{ki} - \bar{x}_i) \ (x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k-1}^{n} (x_{ki} - x_i)^2 \sum_{k-1}^{n} (x_{kj} - x_j)^2}} \tag{13}$$

Then, the correlation coefficient matrix R is as following:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{m1} & r_{m2} & \cdots & r_{nm} \end{bmatrix} \tag{14}$$

(c) Solve eigenvalue and eigenvector: The eigenequation $|\lambda I - R| = 0$ is solved by using the Jacobi Method, and eigenvalue λ is sorted according to the size as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$.

(d) Compute contribution ratio (CR) and accumulating contribution ratio (ACR), and determine principal component: The CR and ACR are expressed as

$$CR = \frac{\lambda_i}{\sum_{k-1}^{p} \lambda_k} (i = 1, 2, .. p) \tag{15}$$

$$ACR = \frac{\sum_{k-1}^{i} \lambda_k}{\sum_{k-1}^{p} \lambda_k} (i = 1, 2, .. p) \tag{16}$$

In general, when the ACR of eigenvalue is greater than 70%, the corresponding pollutants are defined as principal components and shown as the 1st, 2nd, 3rd, ..., mth (m≥4) principal components.

(e) Compute loading and score of principal component: The loading of principal component is expressed as

$$I_{ij} = \sqrt{\lambda_i} e_{ij} (i, \ j = 1, 2, \cdots p) \tag{17}$$

Then, the Principal Component Score Matrix Z shows as following:

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ z_{m1} & z_{m2} & \cdots & z_{nm} \end{bmatrix} \tag{18}$$

(f) Determine expression of principal component and comprehensive evaluation function: The expression of principal component is expressed as

$$\begin{cases} F_1 = l_{11}Z_1 + l_{21}Z_2 + \cdots + l_{n1}Z_n \\ F_2 = l_{12}Z_1 + l_{22}Z_2 + \cdots + l_{n2}Z_n \\ \cdots\cdots \\ F_m = l_{1m}Z_1 + l_{2m}Z_2 + \cdots + l_{nm}Z_n \end{cases} \tag{19}$$

And the expression of comprehensive evaluation function is expressed as

$$F = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \cdots + \lambda_p}F_1 + \frac{\lambda_2}{\lambda_1 + \lambda_2 + \cdots + \lambda_p}F_2 + \cdots\cdots + \frac{\lambda_p}{\lambda_1 + \lambda_2 + \cdots + \lambda_p}F_p \tag{20}$$

According to the circumstance of the Nansi Lake Basin, The FCE-PCA model had been improved and applied to assess the water quality.

3. Results and Discussion

3.1. Model Simulations

The membership functions represent the degree to which the specified concentration belongs to the fuzzy set. The smaller monitoring index is, the better the water quality is, and when the data is close to the critical value, the water quality will be terrible suddenly. Meanwhile, as the response relationship between water quality and monitoring index accords with the Nor-Half Sinusoidal Distribution Function, the membership functions were established via the Nor-Half Sinusoidal Distribution Method as follows.

$$R_1 = \begin{cases} 0 & , \quad (X > X_2) \\ \sin\frac{\pi}{2}\left(\frac{X_2 - X}{X_2 - X_1}\right) & , \quad (X_1 \leq X \leq X_2) \\ 1 & , \quad (X < X_1) \end{cases} \tag{21}$$

$$R_2 = \begin{cases} 0 & , \quad (X > X_3; X < X_1) \\ \sin\frac{\pi}{2}\left(\frac{X - X_1}{X_2 - X_1}\right) & , \quad (X_1 \leq X \leq X_2) \\ \sin\frac{\pi}{2}\left(\frac{X_3 - X}{X_3 - X_2}\right) & , \quad (X_2 \leq X \leq X_3) \end{cases} \tag{22}$$

$$R_3 = \begin{cases} 0 & , \quad (X > X_4; X < X_2) \\ \sin\frac{\pi}{2}\left(\frac{X - X_2}{X_3 - X_2}\right) & , \quad (X_2 \leq X \leq X_3) \\ \sin\frac{\pi}{2}\left(\frac{X_4 - X}{X_4 - X_3}\right) & , \quad (X_3 \leq X \leq X_4) \end{cases} \tag{23}$$

$$R_3 = \begin{cases} 0 & , \quad (X > X_4; X < X_2) \\ \sin\frac{\pi}{2}\left(\frac{X - X_2}{X_3 - X_2}\right) & , \quad (X_2 \leq X \leq X_3) \\ \sin\frac{\pi}{2}\left(\frac{X_4 - X}{X_4 - X_3}\right) & , \quad (X_3 \leq X \leq X_4) \end{cases} \tag{24}$$

$$R_5 = \begin{cases} 0 & , \quad (X < X_4) \\ \sin\frac{\pi}{2}\left(\frac{X - X_4}{X_5 - X_4}\right) & , \quad (X_4 \leq X \leq X_5) \\ 1 & , \quad (X > X_5) \end{cases} \tag{25}$$

where X is the actual monitoring data for the ith assessment parameter, and X_j is the criteria value of the ith assessment parameter at the jth level (i = 1, 2, …, n; j = 1, 2, …, m).

Weight was calculated via the Exceeding Standard Multiple Method as follows:

$$W_i = \frac{X}{S} \tag{26}$$

$$S = \frac{1}{n}\sum_{j-1}^{n} X_j \tag{27}$$

where X is the real concentration of ith pollutant parameter, S is the standard concentration of ith pollutant parameter, X_j is the standard values of jth water quality, and n is the water quality class. The normalization of W_i is expressed as

$$W' = \frac{W_i}{\sum_{i-1}^{n} W_i} \tag{28}$$

After normalization, the weight set A was obtained. Water quality data of five relau monitoring sections in spring, summer, autumn, winter and annual mean were analyzed and calculated by the according membership function. The membership degree was calculated by because it is relevant to the weight of water quality indexes. Through the setting of assessment factor set, the building of the assessment criteria set, the membership function, and the weight set, the computational procedure can be realized.

To eliminate the influence of dimension and magnitudes about pollutant indexes, the original data was operated under normalization via SPSS 19.0. The correlation coefficient matrix can be obtained according to the normalization data. The absolute value of the correlation coefficient between any two indexes is usually greater than 0.1 and it is shown that the PCA model can be used to analyze the inter-indexes correlation. The eigenroot, eigenvector, CR and ACR were computed through the use of this correlation coefficient matrix.

3.2. Calculation Results via the FCE- PCA Model

The rivers of the Nansi Lake Basin are seasonal rivers and the water quality of the Nansi Lake Basin is discrepant in different seasons. The principal pollutants should be paid close attention to and the corresponding measure should be taken in the Nansi Lake Basin. The FCE- PCA model was designed according to conditions of the Nansi Lake Basin and the result would be useful for the water environment management. The water quality of five relau monitoring sections was tested in four seasons.

Through analyzing water quality data of five relau monitoring sections by the FCE model, the seasonal and annual water quality results were obtained as follows in Table 2.

Table 2. Water Quality Results of FCE Results (a) in Spring, (b) in Summer, (c) in Autumn, (D) in Winter and (E) FCE Results of Annual Mean

Monitoring section	FCE results in spring					Water quality
	I	II	III	IV	V	
Point 1	0.317,	0.256,	0.199,	0.205,	0.235	I
Point 2	0.258,	0.221,	0.385,	0.247,	0.000	III
Point 3	0.354,	0.303,	0.353,	0.228,	0.000	I
Point 4	0.221,	0.200,	0.143,	0.450,	0.211	IV
Point 5	0.296,	0.202,	0.404,	0.284,	0.000	III

Monitoring section	FCE results in summer					Water quality
	I	II	III	IV	V	
Point 1	0.341,	0.245,	0.192,	0.160,	0.226	I
Point 2	0.145,	0.318,	0.291,	0.219,	0.147	II
Point 3	0.324,	0.317,	0.150,	0.241,	0.100	I
Point 4	0.193,	0.249,	0.102,	0.281,	0.277	IV
Point 5	0.301,	0.226,	0.281,	0.338,	0.000	IV

Monitoring section	FCE results in autumn					Water quality
	I	II	III	IV	V	
Point 1	0.271,	0.189,	0.286,	0.395,	0.000	IV
Point 2	0.179,	0.315,	0.408,	0.213,	0.008	III
Point 3	0.345,	0.273,	0.232,	0.243,	0.000	I
Point 4	0.270,	0.146,	0.536,	0.267,	0.000	III
Point 5	0.189,	0.049,	0.302,	0.230,	0.314	V

Monitoring section	FCE results in winter					Water quality
	I	II	III	IV	V	
Point 1	0.200,	0.215,	0.199,	0.459,	0.142	IV
Point 2	0.210,	0.258,	0.348,	0.350,	0.015	IV
Point 3	0.302,	0.101,	0.155,	0.525,	0.082	IV
Point 4	0.199,	0.126,	0.269,	0.159,	0.363	V
Point 5	0.367,	0.173,	0.392,	0.346,	0.000	III

Monitoring section	FCE results of annual mean					Water quality
	I	II	III	IV	V	
Point 1	0.362,	0.155,	0.324,	0.248,	0.135	I
Point 2	0.205,	0.300,	0.354,	0.239,	0.038	III
Point 3	0.378,	0.238,	0.304,	0.256,	0.000	I
Point 4	0.296,	0.226,	0.164,	0.320,	0.225	IV
Point 5	0.367,	0.173,	0.392,	0.346,	0.000	III

The results showed that the annual water quality was relatively good: two monitoring sections (Nanyang sublake-point 1 and Zhaoyang sublake-point 3) belonged to class I (best level), two monitoring sections (Dushan sublake-point 2 and Weishan sublake-point 5) belonged to class III (normal level), and the rest one monitoring sections (point 4 close to Weishan sublake) belonged to class IV (bad level). Small amounts of water in winter and more pollution sources near point 4 have the main effects on the annual water quality. The results also indicated that 80% of the monitoring sections were either superior to or flatting to class III (normal level) in spring, 60% of the monitoring sections were superior to class III in summer, 60% of the monitoring sections were either superior to or flatting to class III in autumn, 20% of the monitoring sections belonged to class III in winter, and 80% of the monitoring sections were either superior to or

flatting to class III in annual mean.

The eigenvalue, principal component, contribution ratio and accumulating contribution ratio were computed by the PCA model and the results expressed in Fig. 3.

The eigenroot of the 1st principal component in spring was 5.640, and the contribution ratio was 40.289%. The eigenroot of the 1st principal component in summer was 4.226, and the contribution ratio was 30.184%. The eigenroot of the 1st principal component in autumn was 3.698, and the contribution ratio was 26.416%. The eigenroot of the 1st principal component in winter was 5.927, and the contribution ratio was 42.336%. The eigenroot of the 1st principal component in the annual mean was 5.477, and the contribution ratio was 39.118%.

In the operating procedure of extracting principal component loading, the results were expressed in Table 3.

Table 3. Results of Principal Component Loading

Indexes	Spring				Summer				Autumn				Winter				Annual mean			
PC	1st	2nd	3rd	4th	1st	2nd	3rd	4th	1st	2nd	3rd	4th	1st	2nd	3rd	4th	1st	2nd	3rd	4th
TN	0.571	0.623	-0.397	-0.023	0.322	0.677	0.397	-0.400	0.732	0.310	-0.325	0.210	0.002	0.463	0.691	-0.451	-0.239	0.488	0.770	-0.152
Oils	0.760	-0.083	0.195	0.257	0.612	-0.007	0.135	-0.063	-0.230	0.393	0.028	-0.063	0.727	0.107	0.315	0.033	-0.0169	0.593	-0.141	0.309
TP	0.605	0.521	-0.510	0.107	0.570	0.284	-0.367	-0.533	0.408	0.441	-0.310	0.529	-0.051	0.975	-0.097	0.049	-0.0347	0.762	0.328	-0.280
BOD	0.474	0.478	0.406	-0.594	0.481	0.390	-0.320	0.417	0.232	0.847	0.256	-0.071	-0.268	0.117	0.485	0.611	-0.406	0.516	0.073	0.535
NH₃-N	0.767	0.469	-0.307	-0.203	0.687	-0.014	-0.404	-0.469	-0.012	0.411	-0.084	0.756	-0.076	0.580	0.627	-0.307	-0.354	0.743	0.349	0.086
V-phenol	-0.399	0.696	0.332	0.285	0.587	0.452	0.297	0.402	0.568	0.529	0.058	-0.264	-0.119	0.900	-0.359	0.162	-0.334	0.716	0-.432	-0.358
COD₍Cr₎	0.736	0.199	0.357	-0.502	0.732	-0.380	-0.315	0.230	-0.606	0.625	0.099	-0.055	-0.264	-0.010	0.493	0.571	-0.453	0.567	0-.331	0.448
Cr(VI)	-0.514	0.767	0.223	0.107	0.011	0.828	0.093	-0.007	0.675	0.090	-0.216	-0.313	0.991	0.033	-0.034	0.050	0.926	0.309	0.083	-0.035
Hg	0.680	0.209	0.515	0.419	0.676	0.214	0.124	0.527	0.051	0.819	0.135	-0.340	0.989	0.051	-0.044	0.050	0.931	0.352	-0.046	-0.008
Pb	-0.740	0.598	-0.015	-0.062	0.601	-0.811	0.371	-0.011	-0.864	0.396	-0.257	-0.012	0.989	0.006	-0.050	0.040	0.925	0.313	0.039	-0.059
Cd	-0.625	0.709	0.069	0.019	0.581	-0.753	-0.013	0.137	-0.902	0.357	-0.073	0.014	0.990	0.048	-0.043	0.050	0.929	0.354	-0.049	-0.004
As	-0.271	0.292	0.151	-0.010	0.236	-0.018	0.780	-0.070	0.045	0.067	0.934	0.161	0.989	0.046	-0.040	0.046	0.933	0.343	-0.016	0.011
CN⁻	0.655	0.589	-0.173	0.315	0.731	0.321	-0.133	-0.112	0.348	0.370	-0.207	-0.150	-0.039	0.896	-0.395	0.131	-0.253	0.717	-0.476	-0.377
Cu	0.830	-0.282	0.301	0.241	-0.308	0.393	-0.324	0.465	0.240	0.003	0.906	0.145	0.579	0.008	0.386	0.107	0.534	0.149	0.005	0.454

Note. Not available N.A; Standard hydrogen electrode (SHE); Hydraulic retention time (HRT); Saturated calomel electrode (SCE); Applied voltage (A.P)

The related indexes of the 1st principal component were As (0.933), Hg (0.931), Cd (0.929), Cr(VI) (0.926), Pb (0.925), and Cu (0.534) in the annual mean, and the 2nd principal component were TP(0.762), NH₃-N(0.743), CN⁻(0.717), V-phenol(0.716), Oils(0.593), COD₍Cr₎(0.567), BOD(0.516), and the 3rd principal component were TN(0.770), and the 4th principal component were BOD(0.535). The related indexes of 1st principal component were Cu (0.830), NH₃-N (0.767), Oils (0.760), Pb (-0.740), COD₍Cr₎ (0.736), Hg (0.680), and CN⁻ (0.655) in spring. The related indexes of the 1st principal component were COD₍Cr₎ (0.732), CN⁻(0.731), NH₃-N (0.687), Hg (0.676), Oils (0.612), and V-phenol (0.587) in summer. The related indexes of the 1st principal component were Cd (-0.902), Pb (-0.864), TN (0.732), Cr(VI) (0.675), and V-phenol (0.568) in autumn. The related indexes of the 1st principal component were Cr (VI) (0.991), Cd (0.990), Hg (0.989), Pb (0.989), As (0.989), Oils (0.727), and Cu (0.579) in winter.

3.3. Water Quality Assessment and Principal Pollutant Evaluation

The FCE model was used to assess the integrated water quality of the Nansi Lake Basin. In accordance with the water quality changing with seasons, the water quality in the spring and summer was relatively good, but in the winter the water quality was relatively bad. The water quality order of the Nansi Lake Basin is as follows: Zhaoyang Lake > Nanyang Lake > Dushan Lake > Weishan Lake. These results were calculated by FCE model and were similar to Wu Zhouhu [30]. After analyzing the water quality via the FCE model, the next task was to extract the principal pollutants via the PCA model.

There were four principal components extracted via the PCA model, where the accumulating contribution ratio reached 84.634%, 74.533%, 73.032%, 86.085%, and 85.365% in spring, summer, autumn, winter, and annual mean, respectively. The four principal components extracted via the PCA model can replace the fourteen pollutant indexes for assessing water quality. These results show that the combined pollution of heavy metals, toxic pollutants and organic pollutants should be given much more attention in regards towards the Nansi Lake basin; these pollutants are Cd, Pb, Cr(VI), Hg, As, Cu, CN⁻, V-phenol, Oils, COD₍Cr₎ and BOD. Furthermore, the water eutrophication should be given attention, especially towards TN, NH₃-N and TP.

Enhancing the capacity to understand and control water environmental pollution is a critical issue not only in China and the Nansi Lake Basin, but also in other areas around the world. Brömssen et al. [31] assessed the sustainability of low-arsenic aquifers as a safe drinking water source in southeastern Bangladesh, and illustrated the effects of high-arsenic groundwater on human health based on model analysis. In facts, The FCE-PCA model simulation is not only capable of supporting the regulation and evaluation of water quality effectively and scientifically, but will be able to further assist in finding principal components for targeted control of water pollution. PCA model has been used to reduce the number of environmental variables to abstain from multicollinearity [32]. It is reported that principal components for targeted pollution control of the Bouhamdane River, Guelma (north-east Algerian) are evaluated using PCA to realize the water basin management [33].
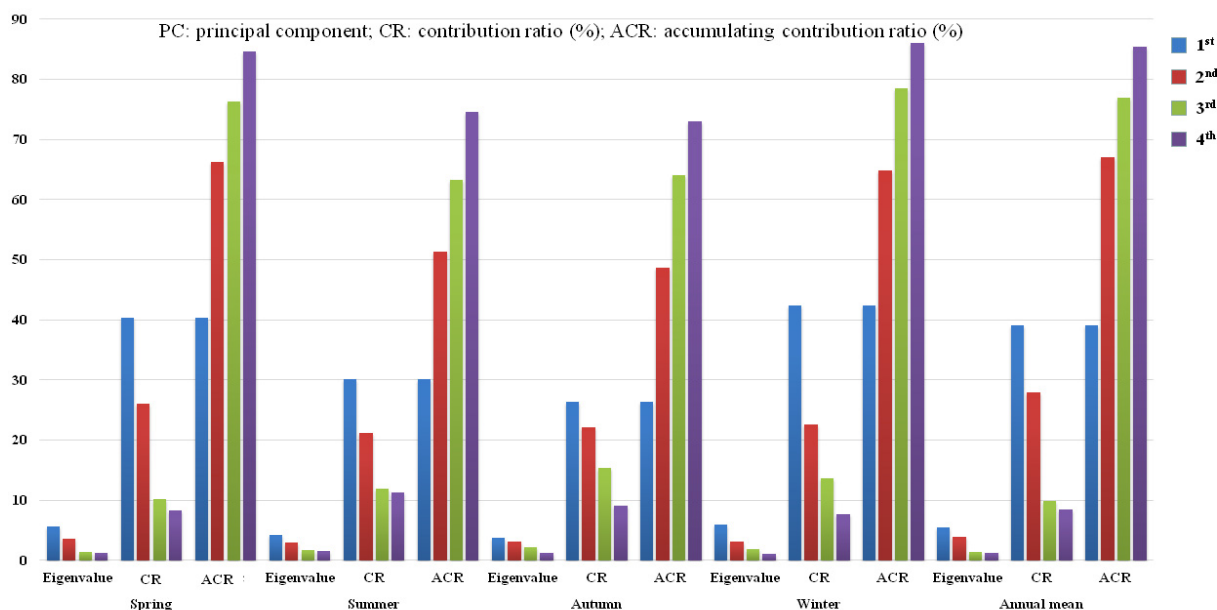
Fig. 3. Results of eigenvalue and principal component.

Besides the understanding of aquatic pollution using the FCE-PCA model, this model method is also used in the fields related to land use [34], and material research [35], which suggested the great potential of the FCE-PCA model in dealing with different problems. In future research, the accuracy and applicability of the combined FCE-PCA should be enhanced, allowing the index and method to function to its full potential by optimizing and by detailing more.

3.4. Environment Implications

To achieve the South-to-North Water Diversion Project and ensure the water quality safety of the Nansi Lake Basin, it is essential to evaluate the water quality and decide the key pollutants. The paroxysmal and gradual warning index of water environment based on the combined FCE-PCA model was structured to ensure the water quality safety of the Nansi Lake Basin and .the South-to-North Water Diversion Project. Results indicated that the heavy metal and organic pollutants may cause some degree environmental risk, and the prevention and control system of risk should pay attention to pollutants (Cu, Cr(VI), Pb, As, V-phenol, Oils, COD) and its sources (paper, chemical, coal mining, operation, textile, etc.). It is the hope that the combined FCE-PCA model will provide valuable information to assess the water quality for the Nansi Lake Basin and the East Route of South-to-North Water Diversion Project.

4. Conclusions

The FCE-PCA model had been applied to assess the water quality of the Nansi Lake Basin in this paper. Based on the analysis of fourteen monitor indexes of the five relau monitoring sections, the water quality classification of the Nansi Lake Basin was determined via the FCE model. The annual water quality was relatively good, with 40% of the monitoring sections belonging to class I (best level), 40% of the sections belonging to class III (normal level), and only 20% of the sections belonging to class IV (bad level). Considering the difference of the temporal and spatial distribution in average, water quality of Level I was 28.9%, 28.1%, 25.1%, 25.6%, respectively in spring, summer, autumn, and winter, which suggested that water quality in spring and summer was better than in autumn and winter. Lastly, the order of water quality was Zhaoyang Lake > Nanyang Lake > Dushan Lake > Weishan Lake.

Based on the analysis of fourteen monitoring indexes of forty-one monitoring sections, the principal pollutants of the Nansi Lake Basin was extracted via the PCA model. There were four extracted principal components that can replace the fourteen pollutant indexes for assessing water quality for the Nansi Lake Basin. Through analyzing the annual mean data, the 1st principal components were heavy metals (e.g. As, Hg, Cd, Cr (VI), Pb etc.), the 2nd principal components were organic and toxic pollutants (e.g., V-phenol, Oils, COD_{Cr}, CN^- etc.), the 3rd and 4th principal components were TN and BOD.

In conclusion, the water quality of Nansi Lake Basin was relatively good especially in spring and summer, and the key pollution indicators are COD_{Cr}, V-phenol, TN, TP, CN^-, Hg, As, Cd, Cr, and Pb. The water quality assessment should be chosen from more comprehensive indexes and a more reasonable weight calculation method like the methods found in the FCE-PCA model. It is our hope that the combined FCE-PCA model will explain the discrepancies, enhance the efficiency and goodness-of-fit, and predict power and robustness.

Acknowledgment

Author Contributions

S.X. (Professor) wrote and revised the manuscript, C.Y. (M.S. student) analyzed the data. Y.C (M.S. student) conducted the experiments. S.W. (Ph.D.), W.D. (Ph.D.), L.H. (Ph.D.) and L.H (Professor) made the model analysis. C.L. (Professor) revised the manuscript. Z.R. (Professor) and W.W (Professor) organized project and revised the manuscript.

References

1. Zhang G, Chen L, Liu Y, et al. A New OMMS of Accidental Pollution Events Developed for Regional Water Basin in Ningbo, China. *Water Sci. Technol.* 2011;64(9):1828-1934.
2. Ma J, Qiao L, Ji L, et al. The online monitoring and assessment of thallium stress using oxygen consumption rate and carbon dioxide excretion rate of zebrafish (Danio rerio). *Chemosphere* 2019;216:103-109.
3. Hartnett M, Nash S. An integrated measurement and modeling methodology for estuarine water quality management. *Water Sci. Eng.* 2015;8(1):9-19.
4. Wang L, Ren Z, Kim H, Xia C, Fu R, Chon T.-S. Characterizing response behavior of medaka (Oryzias latipes) under chemical stress based on self-organizing map and filtering by integration. *Ecol. Inform.* 2015;29:107-118.
5. Altenburger R, Ait-Aissa S, Antczak P. Future water quality monitoring-Adapting tools to deal with mixtures of pollutants in water resource manage-ment. *Sci. Total Environ.* 2015;512: 540-551.
6. Ban YU, Woo HM, Han KM, Baek JI. Building an Integrated Governance Model and Finding Management Measures for Nonpoint Source Pollution in Watershed Management of Korea. *Environ. Eng. Res.* 2013;18(3):199-208.
7. Giansanti D, Maccioni G, Cesinaro S, Benvenutib F, Macellari V. Assessment of fall-risk by means of a neural network based on parameters assessed by a wearable device during posturography. *Med. Eng. Phys.* 2008;30(3):367-372.
8. Hartnett M, Nash S. An integrated measurement and modeling methodology for estuarine water quality management. *Water Sci. Eng.* 2015;8(1):9-19.
9. Ahn J, Na Y, Park SW. Assessment of water quality in an artificial urban canal: A case study of Songdo City in South Korea. *Environ. Eng. Res.* 2019;24(4):582-590.
10. Lee SH, Jung SG, Park SM, Lee BD. Evaluation of the tributaries by influence index on the mid-lower portion of the Nakdong River basin. *Environ. Eng. Res.* 2018;23(2):150-158.
11. Wang X, Pan H. Fuzzy comprehensive evaluation of water quality management information system. *Adv. Mater. Res.* 2014;1044:486-489.
12. Guleda OE, Ibrahim D, Halil H. Assessment of urban quality in Istanbul using fuzzy synthetic evaluation. *Atmos. Environ.* 2004;38:3809-3815.
13. Ji X, Dahlgren RA, Zhang M. Comparison of seven water quality assessment methods for the characterization and management of highly impaired river systems. *Environ. Mon. Assess.* 2016;188(1):15-30.
14. Hou GY, Tang DS. Fuzzy Comprehensive Evaluation of Water Resources Carrying Capacity based on Vague Method. *Appl. Mechan. Mater.* 2014;501:2040-2044.
15. Zhang QW, Zhong M. Using multi-level fuzzy comprehensive evaluation to assess reservoir induced seismic risk. *J. Comput.* 2011;6(8):1670-1676.
16. Sahoo MM, Patra KC, Khatua KK. Inference of Water Quality Index Using ANFIA and PCA. *Aquat. Procedia* 2015;4: 1099-1106.
17. Arslan O. Spatially weighted principal component analysis (PCA) method for water quality analysis. *Water Resour.* 2013;40(3):315-324.
18. Winter TC, Mallory SE, Allen TR. The use of principal component analysis for interpreting ground water hydrographs. *Groundwater* 2000;38(2):234-246.
19. Mustapha A, Aris AZ, Juahir H. River water quality assessment using environmentric techniques: case study of Jakara River Basin. *Environ. Sci. Pollut. Res.* 2013;20(8):5630-5644.
20. Perkins RG, Underwood GJC. Gradients of chlorophyll a and water chemistry along an eutrophic reservoir with determination of the limiting nutrient by in situ nutrient addition. *Water Res.* 2000;34(3):713-724.
21. Wang W, Dong C, Dong W, Yang C, Ju T, Huang L, Ren Z. The design and implementation of risk assessment model for hazard installations based on AHP–FCE method: A case study of Nansi Lake Basin. *Ecol. Inform.* 2016;36:162–171
22. Lv JS, Zhang ZL, Li S. Assessing spatial distribution, sources, and potential ecological risk of heavy metals in surface sediments of the Nansi Lake, Eastern China. *J. Radioanal. Nucl. Ch.* 2014;299(3):1671-1681.
23. Guo YS. On historical change of lakes in Shandong Province. *T. Oceanol. Lim.* 1990;3:15-22.
24. Wang WL, Liu XH, Wang YF, Guo XC, Lu SY. Analysis of point source pollution and water environmental quality variation trends in the Nansi Lake basin from 2002 to 2012. *Environ. Sci. Pollut. R.* 2016;23:4886-4897.
25. Osmi SFC, Malek MA, Yusoff M, Azman NH, Faizal WM. Development of river water quality management using fuzzy techniques: a review. *Int. J. River Basin Manag.* 2015;14(2): 243-254.
26. Zadeh LA. Fuzzy sets. *Inf. Control.* 1965;8(3):338-353.
27. Lu XW, Li LY, Lei K. Water quality assessment of Wei River, China using fuzzy synthetic evaluation. *Environ. Earth Sci.* 2010;60(8):1693-1699.
28. Silvert W. Ecological impact classification with fuzzy sets. *Ecol. Model.* 1997;96(1):1-10.
29. Liu S, Xu L, Li D. Prediction of Aquaculture Water Quality Based on Combining Principal Component Analysis and Least Square Support Vector Regression. *Sensor Lett.* 2013;11(6): 1305-1309.
30. Wu ZH, Zhang J, Zhu J. A monitoring project planning technique

of the water quality spatial distribution in Nansi Lake. *Procedia Environ. Sci.* 2011;10:2320-2328.

31. Brömssen MV, Markussen L, Bhattacharya P, Ahmed KM, Hossain M, Jacks G. Hydrogeological investigation for assessment of the sustainability of low-arsenic aquifers as a safe drinking water source in regions with high-arsenic groundwater in matlab, southeastern bangladesh. *J. Hydro.* 2014;518:373-392.

32. Zhang SY, Zhou QH, Xu D, He F, Cheng SP, Liang W. Vertical-flow constructed wetlands applied in a recirculating aquaculture system for channel catfish culture: effects on water quality and zooplankton. *P. J. Environ. Stud.* 2010;19(5):1063.

33. Zeghaba N, Laraba A. The Principal Components Analysis (PCA) to evaluate the quality of water of the Bouhamdane river, Guelma (north-east Algerian). *J. Fund. Appl. Sci.* 2018;10(2):383-395,.

34. Li Y, Qiao Y, Wu K. Comprehensive evaluation of land use benefit based on fuzzy comprehensive evaluation model. *J. Comput. Theor. Nanos.* 2016;13(5):2952-2958.

35. Tian P, Zhan GF, Nai L. Comprehensive evaluation of asphalt-mixture performance based on principal component analysis. *Adv. Mater. Res.* 2015;1095:280-283.