

An improved index for water quality evaluation in an estuary region: a case study in the Eastern Pearl River Delta, China

Shenlin Li^{a,b,c,d}, Xiaohong Chen^{a,b,c,*}, V. P. Singh^d, Yanhu He^{a,b,c}
and Xiaoyan Bai^e

^aCenter for Water Resources and Environment, Sun Yat-sen University, Guangzhou 510275, China

^bGuangdong Engineering Technology Research Center of Water Security Regulation and Control for Southern China, Sun Yat-sen University, Guangzhou 510275, China

^cKey Laboratory of Water Cycle and Water Security in Southern China of Guangdong High Education Institute, Sun Yat-sen University, Guangzhou 510275, China

*Corresponding author. E-mail: eescxh@mail.sysu.edu.cn

^dDepartment of Biological and Agricultural Engineering, Texas A&M University, College Station, TX 77843, USA

^eSchool of Environmental Science and Engineering, Guangdong University of Technology, Guangzhou 510006, China

Abstract

Water quality evaluation is fundamental for water resources management. In this study, a water quality index (WQI) was constructed to evaluate water quality in an estuary region. First, principal component analysis and the Bartlett method were used to select more important water quality parameters from multivariables. Second, quality curves and weights of selected parameters were assigned, and then WQI scores were calculated. The WQI method was applied to the Eastern Pearl River Delta in China as a case study. Results showed that water quality in the upstream area and the coastal region was better than in the central delta, with an average WQI of 72, 55 and 14, respectively. Results further revealed that water quality in the coastal region was more variable (the standard variation of WQIs is near 20) due to more rapid changes in hydrologic features, while water quality in the inland area was more stable (the standard variation is around 10). Comparison between the WQI and fuzzy evaluation methods indicated the reliability of the WQI method. This WQI method can evaluate water quality in the estuarine delta area well, and statistical techniques used in this paper can be applied in different geographical areas considering their specific characteristics.

Keywords: Eastern Pearl River Delta; Estuary region; Principal component analysis; Water quality index

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY-NC-ND 4.0), which permits copying and redistribution for non-commercial purposes with no derivatives, provided the original work is properly cited (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

doi: 10.2166/wp.2019.151

© 2019 The Authors

Introduction

Water quality evaluation is basic to water resources management and policy making (Wang *et al.*, 2007; Abaurrea *et al.*, 2011; Kong *et al.*, 2014; Akhtar & Iqbal, 2017; Yang *et al.*, 2018). It is even more important nowadays because water resources managers are paying greater attention to the benefits of water environmental and ecological systems (Savenije & Van der Zaag, 2008; Gichuki *et al.*, 2009; Buytaert *et al.*, 2014; Friesen *et al.*, 2017). However, evaluation of water quality is complicated because water quality is impacted by natural processes as well as anthropogenic inputs (Singh *et al.*, 2004; Kazi *et al.*, 2009; Lobato *et al.*, 2015).

Water quality index (WQI) is a widely used method for water quality evaluation. WQI usually considers general water quality parameters, such as dissolved oxygen, pH, temperature, and total dissolved solids (Ott, 1978; Canter, 1985). This was considered as a promising method, and has been constantly improved and developed. Dojlido *et al.* (1994) took both basic and additional parameters into consideration to get more accurate results of evaluation in special cases. Boyacioglu (2007) proposed a universal WQI to evaluate the quality of surface water used for drinking water supply. Lobato *et al.* (2015) constructed a novel WQI, considering the hydrological cycle, for the evaluation of water quality in reservoir areas. Bassi & Kumar (2017) applied WQI as a tool for wetland restoration. The calculation steps, such as parameter selection, assignment of weights, and aggregation of sub-indices to produce an overall index, are still being developed (Boyacioglu, 2007).

Besides the traditional WQI approach, some new approaches have also been used in recent studies, such as statistical and stochastic approach, fuzzy set approach, and artificial intelligence (AI) approach.

Principal component analysis (PCA) is a well-known multivariate analysis technique employed to transfer samplings to another low dimension system that is more convenient for data analysis (Bro & Smilde, 2014). With this method, multivariate problems can be reduced to a small number of indices. This method has been applied to several water quality assessment studies (Boyacioglu, 2006; Selle *et al.*, 2013; Vonberg *et al.*, 2014). Partial least squares analysis (PLSA) is also a powerful statistical tool that requires fewer assumptions (Sawatsky *et al.*, 2015). PLSA is appropriate when there are more predictor variables than observations and multicollinearity problems are required to be taken into account among the predictor variables (Abdi & Williams, 2013; Liu *et al.*, 2013). It has the potential to be applied in water quality evaluation.

The fuzzy set method is another widely used method for water quality evaluation (Chang *et al.*, 2001; Lu & Lo, 2002; Ghosh & Mujumdar, 2006). It provides a logical reasoning to clarify water quality levels under consideration in that all water quality parameters may not be included in a single class. This method can take into consideration more qualitative evaluation aspects, because sophisticated statements can be translated from natural language into a mathematical formalism (McNeill *et al.*, 1996).

Artificial intelligence methods, such as artificial neural networks (ANNs) and genetic algorithm, have also been applied in some studies (Aguilera *et al.*, 2001; Gentry *et al.*, 2003; Schulze *et al.*, 2005; Kuo *et al.*, 2006).

However, even with numerous mature evaluation methods, some challenges still exist in water quality evaluation research. One of them is how to select the right water quality parameters in special regions or under particular requirements. Water quality evaluation is a very wide research range, and different parameters could be involved for certain research targets, such as eutrophication of lakes or reservoirs

(Ignatiades *et al.*, 1992; Parinet *et al.*, 2004), drinking water evaluation (Avvannavar & Shrihari, 2008; Haydar *et al.*, 2016), urban storm water quality evaluation (Vaze & Chiew, 2003) or heavy metals in mining or industrial areas (Prasad & Bose, 2001; Edet & Offiong, 2002). These discrete studies chose different parameters to evaluate water quality. However, there is no common method to select the right parameters in unusual regions.

As a result, this paper proposed an improved WQI method to address this challenge. The proposed method aims to select the key water quality parameters from all redundant information based on both experiential and statistic methods. The key parameters are required to be sufficient to evaluate local water quality accurately, and also avoid invalid information. In this study, we want to introduce an acceptable method for water quality parameters' selection and water quality evaluation, and apply it to various regions or situations.

Case study area and data

The Eastern Pearl River Delta is located near Dongguan City, China, at a latitude of $22^{\circ}45'–22^{\circ}10'N$ and longitude of $113^{\circ}30'–113^{\circ}55'E$. It is a triangular area embraced by two tributaries of the Eastern Pearl River, and the Shiziyang estuary, as shown in Figure 1. Numerous rivers flow through this

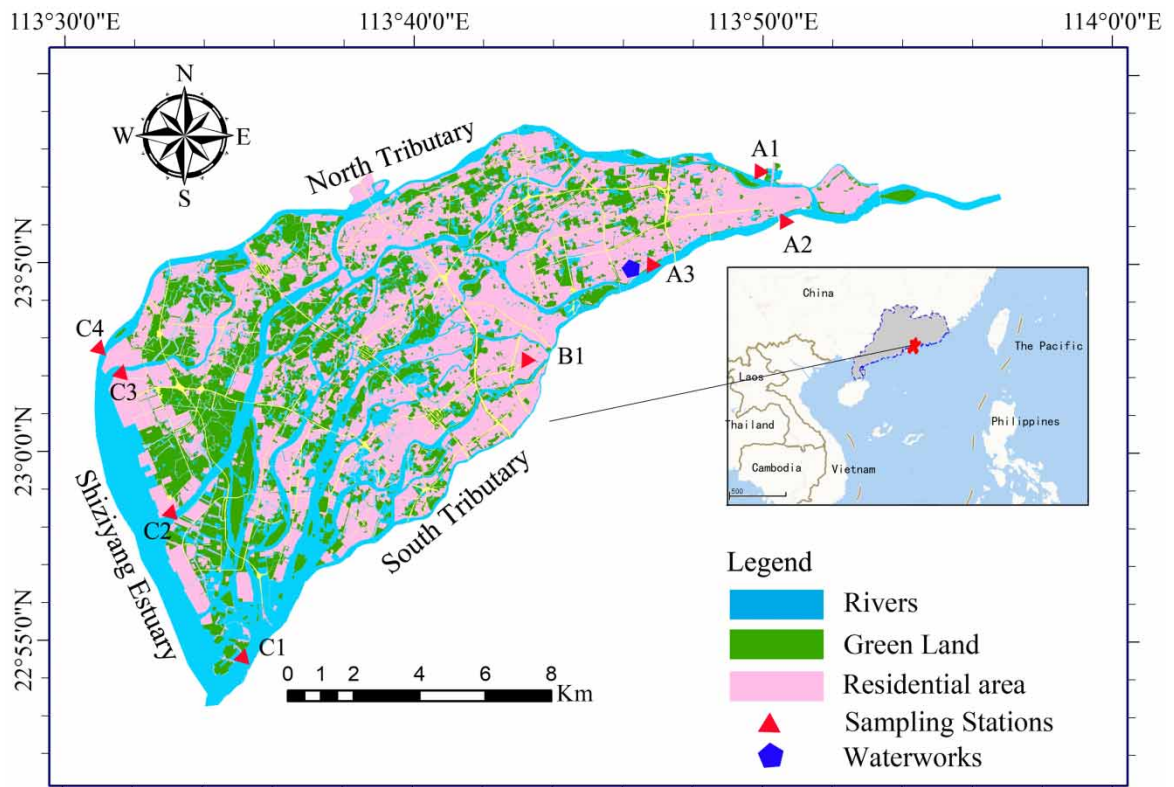


Fig. 1. Map of Eastern Pearl River Delta and features of its land use.

region, and cover an area of 63.7 km², accounting for 24% of the whole delta. The average annual precipitation is 1,766 mm, 80% of which is concentrated in the rainy season (from April to September) (Wang et al., 2015). The northeast part of the delta is a major industrial and residential area; in the west part some undeveloped woodlands and grasslands remain. The international port Humen Port is located on the west coastline.

We chose this as a case study area because of its complicated hydrological environment and important location. The area is affected by tidal influence. Besides pollution sources from upstream and the riverine area, sources from Shiziyang estuary can also influence water quality. Moreover, it is a geometric centre of the Pearl River Delta urban agglomeration, and conjoins Guangzhou and Shenzhen City. As an important area, its sustainable development regarding the environmental and ecological system is a noted programme for the local government.

Eight water sampling stations, named A1, A2, A3, B1, C1, C2, C3 and C4, respectively, exist in the study area, as shown in Figure 1. A1, A2 and A3 are upstream stations; B1 is located in the urban centre area; and C1, C2, C3 and C4 are coastal stations, where tidal influence is dominant and can be violent. A total of 196 water samples were collected from 2009 to 2012, and these are summarized in Table 1. The samplings are carried out every month or every two months at the corresponding station. All detection results are supported by the official detection centre of Guangdong Province.

The following physico-chemical parameters were detected: temperature (T), pH, faecal coliforms (FC), electrical conductivity (EC), dissolved oxygen (DO), chemical oxygen demand (COD), 5 days' biological oxygen demand (BOD₅), ammonium (NH₄-N), total phosphorus (TP), fluorides (F), cyanide, phenols, chromium VI, copper (Cu), zinc (Zn), lead (Pb), cadmium (Cd), mercury (Hg), arsenic, selenium, anionic, chloride (Cl), sulfates (SO₄) and nitrate (NO₃). The upstream discharge (Q) is also recorded at the same time as part of the hydrological measurements. The threshold values of different water quality classifications in the Chinese national standards are shown in Table 2.

Methodology

The water quality evaluation method entails two steps. First, the important water quality parameters are chosen from the 25 parameters in order to reflect the specific characteristics of the study area. Factorial analysis, Bartlett method and experience-based judgement are used in parameter selection.

Table 1. Locations of eight water sampling stations and water sampling frequency per study year.

Sampling station	Water sampling frequency			
	2009	2010	2011	2012
A1	12	12	12	3
A2	12	5	6	11
A3	7	6	6	12
B1	4	4	4	5
C1	6	5	5	3
C2	6	5	5	3
C3	6	5	5	2
C4	6	5	5	3

Table 2. Threshold values of different water quality classifications in the Chinese national standards.

Parameters	Units	Classification				
		I	II	III	IV	V
pH		6–9				
FC	Units/L	200	2,000	10,000	20,000	40,000
DO	mg/L	7.5	6	5	3	2
COD	mg/L	2	4	6	10	15
BOD ₅	mg/L	3	3	4	6	10
NH ₄ -N	mg/L	0.15	0.5	1	1.5	2
TP	mg/L	0.02	0.1	0.2	0.3	0.4
F	mg/L	1	1	1	1.5	1.5
Cyanide	mg/L	0.005	0.05	0.02	0.2	0.2
Phenols	mg/L	0.002	0.002	0.005	0.01	0.1
Chromium VI	mg/L	0.01	0.05	0.05	0.05	0.1
Cu	mg/L	0.01	1	1	1	1
Zn	mg/L	0.05	1	1	2	2
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
Hg	mg/L	0.00005	0.00005	0.0001	0.001	0.001
Arsenic	mg/L	0.05	0.05	0.05	0.1	0.1
Selenium	mg/L	0.01	0.01	0.01	0.02	0.02
Anionic	mg/L	0.2	0.2	0.2	0.3	0.3
Cl	mg/L	≤ 250				
SO ₄	mg/L	≤ 250				
NO ₃	mg/L	≤ 10				
T	°C	No standards				
EC	μs/cm	No standards				
Q	m ³ /s	No standards				

Second, using the quality curves and weights of selected parameters, an integrated WQI score is constructed. These steps are now discussed.

Parameter selection

The experience-based judgement is exercised first. Some parameters may be inert, having tiny values and little fluctuation. These parameters are excluded, assuming that they would be inert under the local conditions as well. Table 3 shows the average values, the standard deviation (StD) and the coefficient of variation (CV) of all 25 water quality parameters. It can be seen that the underlined parameters are cyanide, phenols, chromium VI, Cu, Zn, Pb, Cd, Hg, arsenic and anionic. These parameters have very tiny average values and StD. Moreover, parameters with a larger CV, such as cyanide and Cu with a CV of 7.45 and 7.32, respectively, also fluctuate within a good quality range. Because their sum of mean and StD (0.012 and 0.09, respectively) are still lower than the threshold values of classification II (0.05 and 1.0, respectively, Table 2), they were considered inert, and were therefore removed from the data set.

Then, a PCA of the remaining parameters was done. PCA is a well-known multivariate analysis technique and was first proposed by Pearson (1901). The method allows the identification of patterns in a

Table 3. Average values and variances of detected water quality parameters.

Parameters	T	pH	FC	EC	DO	COD	BOD ₅	NH ₄ -N	TP
Mean	22.7	6.94	8,986	336	3.95	4.2	4.1	1.918	0.264
StD	5.90	0.19	2,8250.00	282.63	2.13	2.53	2.47	2.78	1.65
CV	0.26	0.03	3.14	0.84	0.54	0.61	0.60	1.45	6.25
Parameters	F	Cl	SO₄	NO₃	Cyanide	Phenols	Chromium VI	Cu	Zn
Mean	0.38	154.33	29.28	0.98	0.00191	0.00050	0.002	0.011	0.015
StD	0.20	1,142.45	93.78	0.89	0.01	0.001	0.003	0.08	0.01
CV	0.51	7.40	3.20	0.91	7.45	1.61	1.54	7.32	0.68
Parameters	Pb	Cd	Hg	Arsenic	Selenium	Anionic	Q		
Average value	0.003	0.0007	0.000096	0.002	0.000	0.02	564		
StD	0.004	0.0004	0.00002	0.001	0.0004	0.05	285.00		
CV	1.41	0.66	0.21	0.61	1.53	2.34	0.51		

data series and expressing them in such a way that the similarities and differences can be observed, reducing the dimensionality without losing too much information (Sauthier *et al.*, 2018). With the eigenvalues (A_j , j is the sequence of components) and their eigenvectors (e_{ij} , i is the sequence of parameters) of principal components, the loading matrix (a_{ij}) and the commonalities (S) of parameters are calculated by the following equations:

$$a_{ij} = A_j * \text{sqrt}(e_{ij}) \quad (1)$$

$$S_i = \sum_{j=1}^n a_{ij}^2 \quad (2)$$

Results of PCA are shown in Table 4, which shows that five principal components were selected, with a cumulative variance of 71%.

According to the Bartlett method, if the commonalities are larger than 0.5, the corresponding parameters are chosen for the WQI construction (Hair *et al.*, 1998; Toledo & Nicoletta, 2002). In this step, EC and NO₃ with a commonality of 0.46 and 0.45, respectively, were removed. The remaining parameters, including temperature (T), pH, faecal coliforms (FC), dissolved oxygen (DO), chemical oxygen demand (COD), 5 days' biological oxygen demand (BOD₅), ammonium (NH₄-N), total phosphorus (TP), fluorides (F), chloride (Cl), sulfates (SO₄) and upstream discharge (Q) were then selected. A new PCA of these parameters was done and one can see from Table 5 that all parameters passed the Bartlett test, indicating that they were the main parameters for explaining the water quality of the study area.

Construction of WQI

A general WQI formula was expressed as:

$$\text{WQI} = \prod_{i=1}^m q_i^{w_i} \quad (3)$$

Table 4. Selected principal components and commonalities of original parameters.

Parameters	Commonalities	Components				
		1	2	3	4	5
T	0.67	−0.03	0.34	−0.74	−0.04	−0.01
pH	0.67	0.20	−0.02	0.13	0.76	−0.17
FC	0.68	0.01	0.38	0.02	0.57	−0.46
EC	0.46	−0.61	0.00	0.17	0.23	0.01
DO	0.74	−0.75	−0.23	0.33	−0.10	0.09
COD	0.86	−0.92	−0.03	−0.12	0.01	−0.04
BOD ₅	0.80	−0.85	−0.15	0.12	0.08	0.16
NH ₄ -N	0.63	−0.74	−0.02	−0.24	0.12	−0.13
TP	0.90	−0.04	0.04	−0.04	0.50	0.80
F	0.75	−0.79	0.16	−0.28	−0.10	−0.13
Cl	0.87	−0.14	0.82	0.42	−0.08	−0.01
SO ₄	0.92	−0.24	0.85	0.36	−0.13	0.03
NO ₃	0.45	0.24	0.56	0.01	−0.10	0.24
Q	0.62	0.07	−0.44	0.63	−0.08	−0.13
Eigenvalue (A)		3.85	2.27	1.57	1.30	1.03
Variance		27%	16%	11%	9%	7%
Cumulative variance		27%	44%	55%	64%	71%

Table 5. Selected principal components and commonalities of selected parameters.

Variables	Commonalities	Components				
		1	2	3	4	5
T	0.72	0.07	0.38	−0.74	−0.14	0.08
pH	0.64	−0.23	0.07	0.07	0.75	−0.12
FC	0.72	−0.01	0.40	−0.02	0.54	−0.51
DO	0.73	−0.71	0.28	−0.36	0.05	−0.10
COD	0.86	−0.92	0.09	0.09	−0.05	0.05
BOD ₅	0.82	0.84	−0.19	0.14	0.14	0.17
NH ₄ -N	0.73	0.77	−0.10	−0.20	0.21	−0.21
TP	0.87	0.03	0.00	−0.08	0.54	0.76
F	0.76	0.81	0.09	−0.24	−0.08	−0.17
Cl	0.95	0.19	0.84	0.44	−0.07	0.09
SO ₄	0.95	0.29	0.83	0.38	−0.13	0.10
Q	0.65	−0.11	−0.45	0.64	0.01	−0.17
Eigenvalue (A)		3.51	2.05	1.56	1.26	1.01
Variance		29.2%	17.1%	13.0%	10.5%	8.4%
Cumulative variance		29.2%	46.3%	59.3%	69.8%	78.2%

where WQI is the water quality index in the range of 0–100; q_i is the quality of the i -th parameter; m is the number of parameters selected to compose the index; and w_i is the weight corresponding to the i -th parameter.

The categories of WQI (Lobato *et al.*, 2015) results are shown in Table 6. A higher value of WQI indicates better water quality. The index value from 0 to less than 20 represents poor quality, 20

Table 6. Categories proposed for WQI.

WQI classification	0–19	20–36	37–51	52–79	80–100
	Poor	Bad	Acceptable	Good	Excellent

to less than 37 bad quality, 37 to 51 acceptable quality, 52 to 79 good quality, and above excellent quality.

As Equation (3) shows, two key steps to yield WQI are calculating q_i and w_i , respectively. To calculate q_i , the quality curve of every selected parameter was plotted, as shown in Figure 2. For every parameter, the classifications of ‘excellent, good, acceptable, bad and poor’, respectively, correspond

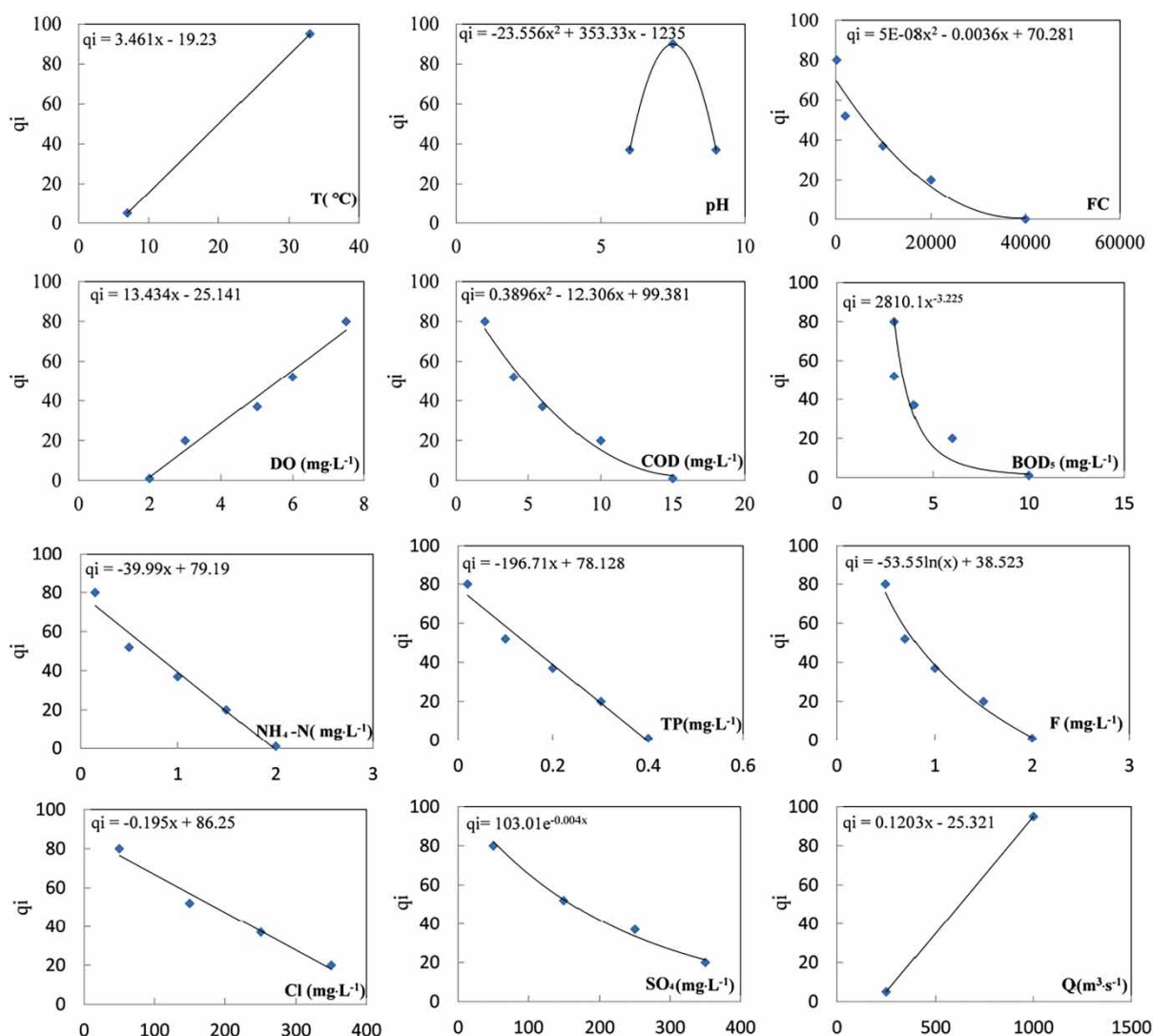


Fig. 2. Water quality curves of selected parameters.

to ‘I, II, III, IV, V’ shown in Table 2. The threshold values of different water quality classifications were plotted as a scatterplot first. A curve was then fitted as a water quality curve on the basis of such scatters. Moreover, the parameters of temperature and discharge are without standard threshold values in the Chinese national standards, so their quality curves were generated following Lobato et al. (2015). Quality curves established the functional relationship between original data and q_i , by which we can yield the q_i scores.

The quantity w_i is the weight corresponding to the i -th parameter, a number between 0 and 1, assigned according to its importance to the overall water quality status. The sum of weights w_i equals 1 (Toledo & Nicolella, 2002). The yield of w_i requires the results from PCA, as Equation (4) shows:

$$w_i = \frac{\sum_{j=1}^n A_j * a_{ij}}{\sum_{i=1}^m \sum_{j=1}^n A_j * a_{ij}}; \quad \sum_{i=1}^m w_i = 1 \quad (4)$$

where w_i is the weight of the i -th WQI parameter; m is the number of selected parameters; and n is the number of principal components. A_j is eigenvalue of j -th principal component (as Table 5 shows), a_{ij} is the loading matrix as Equation (1) has shown.

Fuzzy method

In this study, the fuzzy method was used to yield results for comparison. The fuzzy method is well employed in water quality evaluation (Chang et al., 2001; Lu & Lo, 2002; Ghosh & Mujumdar, 2006; Lai et al., 2015). To employ the method, a fuzzy evaluation matrix is obtained by membership function first (Sun et al., 2014). As mentioned above, there are five different water quality levels. For each original data x_i , its membership degree r to a certain level can be calculated by Equations (5)–(9):

$$r_{i1}(x_i) = \begin{cases} 1, & (x_i < V_1) \\ \frac{x_i - V_1}{V_2 - V_1}, & (V_1 < x_i < V_2) \\ 0, & (x_i \geq V_2) \end{cases} \quad (5)$$

$$r_{i2}(x_i) = \begin{cases} 0, & (x_i \leq V_1 \text{ or } x_i \geq V_3) \\ \frac{x_i - V_1}{V_2 - V_1}, & (V_1 < x_i \leq V_2) \\ \frac{V_3 - x_i}{V_3 - V_2}, & (V_2 < x_i < V_3) \end{cases} \quad (6)$$

$$r_{i3}(x_i) = \begin{cases} 0, & (x_i \leq V_2 \text{ or } x_i \geq V_4) \\ \frac{x_i - V_2}{V_3 - V_2}, & (V_2 < x_i \leq V_3) \\ \frac{V_4 - x_i}{V_4 - V_3}, & (V_3 < x_i < V_4) \end{cases} \quad (7)$$

$$r_{i4}(x_i) = \begin{cases} 0, & (x_i \leq V_3 \text{ or } x_i \geq V_5) \\ \frac{x_i - V_3}{V_4 - V_3}, & (V_2 < x_i \leq V_3) \\ \frac{V_4 - x_i}{V_4 - V_3}, & (V_3 < x_i < V_4) \end{cases} \quad (8)$$

$$r_{i1}(x_i) = \begin{cases} 0, & (x_i < V_4) \\ \frac{x_i - V_4}{V_5 - V_4}, & (V_4 < x_i < V_5) \\ 1, & (x_i \geq V_5) \end{cases} \quad (9)$$

where V_j ($j = 1, 2, 3, 4, 5$) is the threshold of water quality level.

Applying x_i to Equations (5)–(9), an evaluation matrix R shown as Equation (10) can be yielded. Along with weight vector, the comprehensive membership degree B is calculated by Equation (11):

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{15} \\ r_{21} & r_{22} & \dots & r_{25} \\ \dots & \dots & \dots & \dots \\ r_{i1} & r_{i2} & \dots & r_{i5} \end{bmatrix} \quad (i = 1, 2, \dots, m) \quad (10)$$

$$B = W * R = [w_1, w_2, \dots, w_i] * \begin{bmatrix} r_{11} & r_{12} & \dots & r_{15} \\ r_{21} & r_{22} & \dots & r_{25} \\ \dots & \dots & \dots & \dots \\ r_{i1} & r_{i2} & \dots & r_{i5} \end{bmatrix} = [b_1, b_2, b_3, b_4, b_5] \quad (11)$$

where W is the weight vector. In vector B , the final quality lever can be determined by maximum membership degree law (Xue & Yang, 2014), which identifies the final quality level as where the maximum membership degree is located.

Results

In the case study area, initial data of 25 quality parameters were collected in every sampling. We removed 11 ‘inert’ parameters by experience method (Table 3), and removed another two parameters according to the Bartlett test (Table 4). Applying the PCA method for the 12 reserved parameters, five major components, whose cumulative variance is 78.2%, were chosen (Table 5). A weight vector and sub-scores for parameters were then calculated, employed by the PCA method and quality curves respectively. After that, the WQI scores were yielded.

The WQI scores are shown in Table 7. The average scores of stations A1, A2 and A3 were 74, 71 and 73, respectively, which were all rated as ‘good’. Station B1 was rated as ‘poor’, with an average WQI of 14. The average WQI scores of C1, C2, C3 and C4 stations were 48, 56, 52 and 58, respectively, which were categorized as ‘acceptable’ or ‘good’.

The results of evaluation were consistent with empirical experience. A1, A2 and A3 stations are located upstream, where water resources discharged from the mountainous area are of good quality.

Table 7. Mean of the results obtained for the WQI sampling stations in 2009, 2010, 2011 and 2012.

Station	A1	A2	A3	B1	C1	C2	C3	C4
Average score	74	71	73	14	48	56	52	58
Standard deviation	10.20	10.71	10.24	7.05	24.41	20.09	17.32	15.58
Level	Good	Good	Good	Poor	Acceptable	Good	Good	Good

On the other hand, as shown in Figure 1, the residential area has an efficient sewerage system and water protection is strict as the water works of Dongguan City are located near the area of A3. However, the B1 station is situated in the central urban area which is densely populated and suffers from serious anthropogenic pollution, with the result that water quality is very poor. Stations C1, C2, C3 and C4 are situated in a more natural area, where scattered agriculture is the main source of pollution and water quality is generally good. In particular, water quality of C1 is worse than that of C2, C3 and C4, as runoff here is from a more urban area and receives more pollution.

The standard deviation (StD) of WQI provided a further indication of water quality. The StD values of A1, A2, A3 and A4 were 10.2, 10.71, 10.24, 7.05, respectively, and those of stations C1, C2, C3 and C4 were, respectively, 24.41, 20.09, 17.32 and 15.58, which was higher than that of the former four stations. A higher StD value meant the sampling evaluation results were more unstable, which is illustrated in Figure 3, in which the WQI scores of C1, C2, C3 and C4 are located in almost every quality level, while scores of other stations are more concentrated in one or two quality levels.

The dispersive WQI scores of stations C1, C2, C3 and C4 can be explained by noting that these four stations are coastal, where hydrological conditions are highly influenced by tides. The water quality of the area is not only influenced by local water runoff but also by the round-trip runoff from Shiziyang Channel and discharge from Guangzhou City and the South China Sea. In contrast, stations A1, A2, A3 and B1 are inland, where hydrological conditions are much more stable and have a more consistent WQI.

To further demonstrate the WQI method, a fuzzy evaluation method was applied for comparison. Numerical results, translated into quality level, are shown in Figure 4.

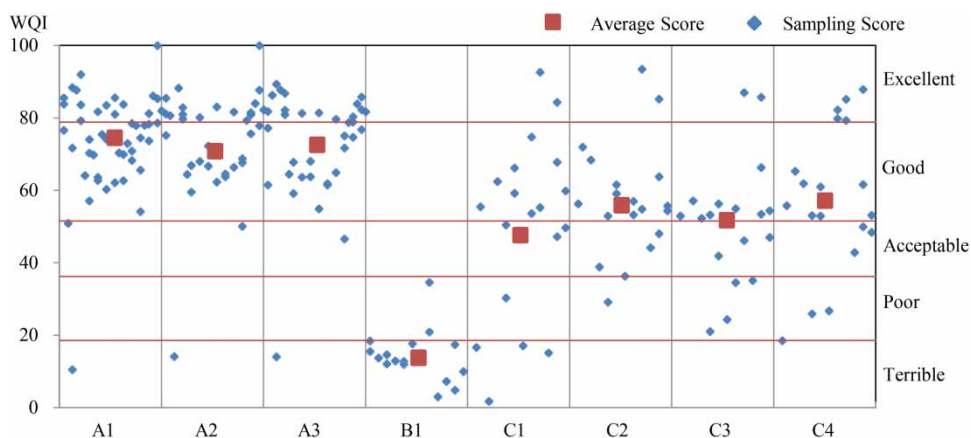


Fig. 3. Scatter plot of WQI scores.

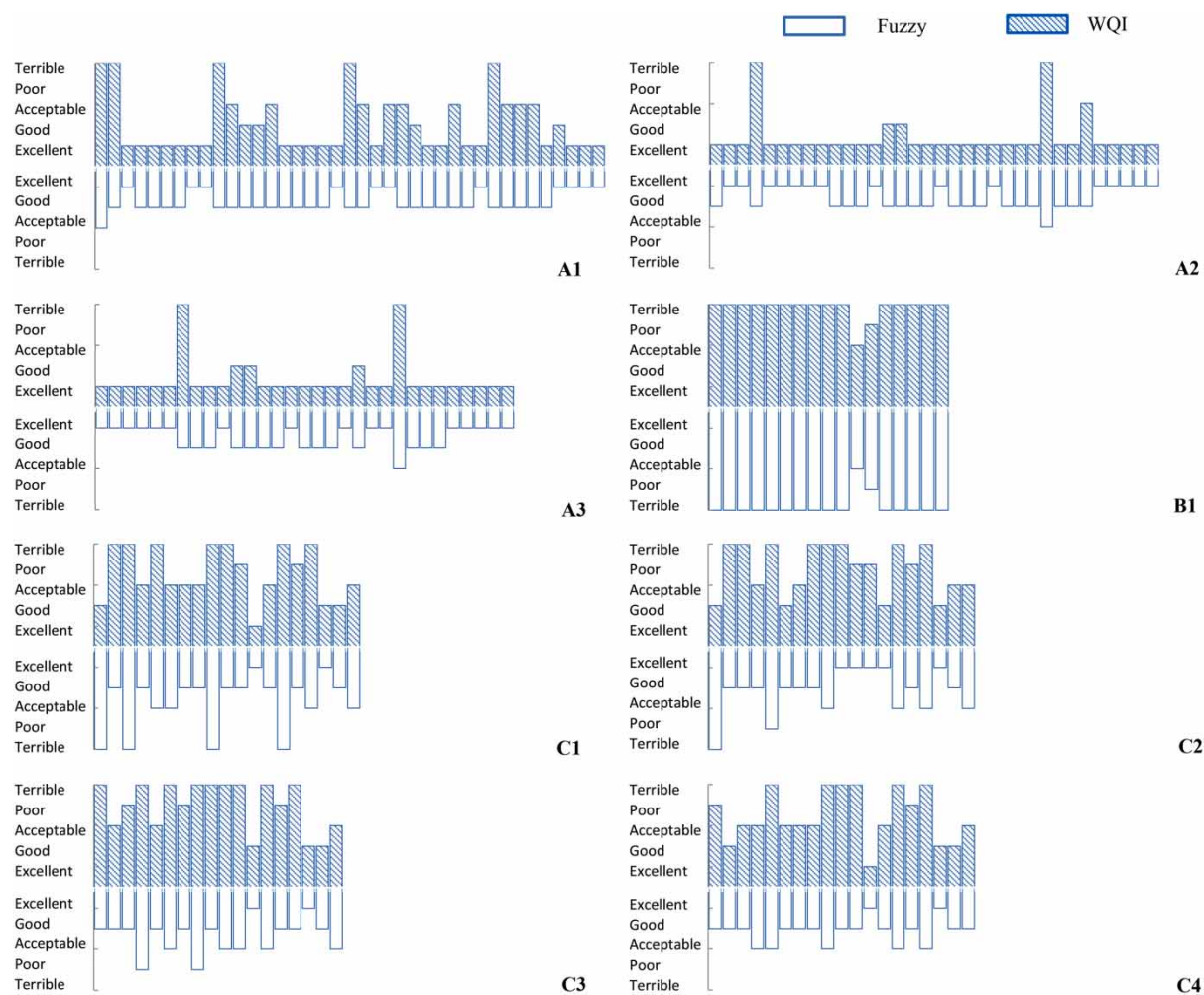


Fig. 4. Comparison of water quality levels of every sample from WQI and fuzzy methods.

It can be seen that the results of the two different methods were more consistent in inland stations, and less similar for stations located offshore. For example, for station B1, the two methods had exactly the same evaluation results for all 17 samples. Stations A1, A2 and A3 had the same results in, respectively, 31%, 50% and 61% of the samples, and for above 85%, 94% and 94% of the samples the results were within a difference of less than one level (Table 8). However, for the C1, C2, C3 and C4 stations, the differences between the two methods were larger. Taking C2 as an example, only 11% of samples' water quality levels are the same, and 58% of samples' differences are more than one level. Further, evaluation results of the fuzzy method were worse than the results of WQI. For station C2, for example, there was only one sample evaluated as 'terrible' by WQI, while there were eight samples evaluated in this way by the fuzzy method.

The consistent results obtained for inland stations were mainly because their water quality status was stable. Taking B1 as an example, water quality here was recognized as terrible, caused by perennial

Table 8. Difference in water quality level between WQI and fuzzy methods.

	A1	A2	A3	B1	C1	C2	C3	C4
Difference = 0	31%	50%	61%	100%	37%	11%	11%	21%
Difference ≤ 1	85%	94%	94%	100%	63%	42%	44%	58%

pollution. Many quality parameters badly exceeded the state standard. Therefore, it was easier to get a consistent water quality evaluation result. In contrast, water quality at offshore stations was more uncertain and hence less consistent results at these stations were expected.

However, the WQI method generated more considerate evaluation results, because it not only indicated the fluctuating status in these areas, but also illustrated the relative state among stations. For example, one can see from Figure 4 that with the fuzzy method, the evaluation of water quality of C1 was in general better than of C2 (with nine ‘poor’ or ‘terrible’ samplings of C1 and 11 ‘poor’ or ‘terrible’ samplings of C2), while by WQI, the quality at C1 was worse (with four ‘poor’ or ‘terrible’ samplings of C1 and two ‘poor’ or ‘terrible’ samplings of C2). The latter was a more acceptable result. C1 had a worse water quality as its discharge had gone through the whole urban area and carried more pollution, while C2 was expected to have better water quality because its upstream area was almost a natural area and encountered less pollution sources.

Discussion

In this study, an improved WQI method was proposed. WQI is widely used in water quality evaluation. However, compared with the traditional WQI method, the proposed method has significant improvement, especially in the step of selecting parameters. As Canter (1985) expressed, the traditional WQI takes into consideration dissolved oxygen, faecal coliforms, pH, biochemical oxygen demand, nitrates, phosphates, temperature, turbidity, and total solids. However, the proposed WQI did not apply this mechanically. We proposed to select parameters according to the local natural and social situation. As a result, 12 quality parameters from 25 parameters were selected in the case study area. COD, NH₄-N, F, Cl, SO₄ and the upstream discharge were added in the integrated WQI. NO₃ was excluded (turbidity and total solids were not recorded in the original data set). This means that COD, NH₄-N, F, Cl, SO₄ and upstream discharges were more variable factors than NO₃ to influence water quality in the study area. Choosing important water quality parameters can eliminate the influence of unimportant parameters and hence yields more accurate evaluation results.

Moreover, a comparison between the proposed WQI method and fuzzy method were also present in this study. The fuzzy method is also a popular method employed in water quality evaluation (Chang et al., 2001; Lu & Lo, 2002; Ghosh & Mujumdar, 2006). We have illustrated that both methods evaluated the water quality well and the proposed WQI method even got better results. Moreover, beyond the results, the improved WQI has two more advantages. The first is the yield of sub-score. WQI generates the sub-score (q_i in Equation (3)) by quality curves, while the fuzzy method uses linear relationship (as Equations (5)–(9) show). A nonlinear relationship is more suitable for some quality parameters, such as pH and FC. The second is the determination of parameters’ weights. In the proposed WQI method, once the parameters are selected from PCA, their weights are also assigned at the same

time and the relative importance of each parameter is thus assessed (as Equation (4) shows). It is thus more objective when doing water quality evaluation and comparing the results of different methods.

A potential problem existing in the method is from the ‘inert’ parameters. They are removed from the considered list at the first step because of their tiny and safe value. However, the value could produce explosive growth if an emergency happened. As a result, the proposed method is more workable in normal situations than emergencies. Another question concerns the removed parameters in PCA. From a regular point of view, EC and NO₃ are critical parameters in water quality evaluation. However, in this study, the two parameters did not pass the statistical test and were removed. Without the parameters, we still got good evaluation results; however, the inner mechanism of the statistical test results has not been researched, which is worthy of further study.

Conclusion

Considering the different regional features of water quality evaluation, a WQI method is proposed in which principal component analysis is used to select more important water quality parameters, and each parameter is assigned a weight reflecting its relative important. Evaluation scores of individual parameters are calculated by fitting a water quality curve for every selected parameter. Corresponding with the individual score and the weight, an integrated WQI score is computed, which is applied to the Eastern Pearl River Delta as a case study area. A fuzzy method is also applied for comparison. The proposed method selected 12 key parameters from all parameters, with which the WQI scores were calculated. The results show the upstream region, the central delta and the coastal region have an average WQI of 72, 55 and 14, respectively. Evaluation results reflect the specific quality features of different sampling stations. Comparison with the fuzzy method further improves the validation of the proposed WQI. The method is well-behaved as it has the ability to select specific parameters according to their hydrological and social characteristics, and yields accurate evaluation results, and thus has enormous potential to be applied to different regions.

Acknowledgements

The research is financially supported by National Key R&D Program of China (2017YFC0405900), National Natural Science Foundation of China (Grant No. 91547202, 51210013, 51479216, 51509127, 51509040), the Chinese Academy of Engineering Consulting Project (2015-ZD-07-04-03), the Project for Creative Research from Guangdong Water Resources Department (Grant No. 2016-07, 2016-01), Research program of Guangzhou Water Authority (2017), the State Scholarship Fund of China (201706380071).

References

- Abaurrea, J., Asín, J., Cebrián, A. C. & García-Vera, M. A. (2011). Trend analysis of water quality series based on regression models with correlated errors. *Journal of Hydrology* 400(3–4), 341–352.
- Abdi, H. & Williams, L. J. (2013). Partial least squares methods: partial least squares correlation and partial least square regression. *Methods in Molecular Biology* 930, 549–579.

- Aguilera, P. A., Frenich, A. G., Torres, J. A., Castro, H., Vidal, J. M. & Canton, M. (2001). Application of the Kohonen neural network in coastal water management: methodological development for the assessment and prediction of water quality. *Water Research* 35(17), 4053–4062.
- Akhtar, S. M. & Iqbal, J. (2017). Assessment of emerging hydrological, water quality issues and policy discussion on water sharing of transboundary Kabul River. *Water Policy* 19(4), 650–672.
- Avvannavar, S. M. & Shrihari, S. (2008). Evaluation of water quality index for drinking purposes for river Netravathi, Mangalore, South India. *Environmental Monitoring and Assessment* 143(1–3), 279–290.
- Bassi, N. & Kumar, M. D. (2017). Water quality index as a tool for wetland restoration. *Water Policy* 19(3), 390–403.
- Boyacioglu, H. (2006). Surface water quality assessment using factor analysis. *Water SA* 32(3), 389–393.
- Boyacioglu, H. (2007). Development of a water quality index based on a European classification scheme. *Water SA* 33(1), 101–106.
- Bro, R. & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods* 6(9), 2812–2831.
- Buytaert, W., Zulkafli, Z., Grainger, S., Acosta, L., Alemie, T. C., Bastiaensen, J., Bievre, B. D., Bhusal, J., Clark, J., Dewulf, A., Foggin, M., Hannah, D. M., Hergarten, C., Isaeva, A., Karpouzoglou, T., Pandeya, B., Paudel, D., Sharma, K., Steenhuis, T., Tilahun, S., Hecken, G. V. & Zhumanova, M. (2014). Citizen science in hydrology and water resources: opportunities for knowledge generation, ecosystem service management, and sustainable development. *Frontiers in Earth Science* 2, 26.
- Canter, L. W. (1985). *Environmental Impact of Water Resources Projects*. Lewis Publishers, Chelsea, Michigan, USA.
- Chang, N. B., Chen, H. W. & Ning, S. K. (2001). Identification of river water quality using the fuzzy synthetic evaluation approach. *Journal of Environmental Management* 63(3), 293–305.
- Dojlido, J., Raniszewski, J. & Woyciechowska, J. (1994). Water quality index – application for rivers in Vistula river basin in Poland. *Water Science and Technology* 30(10), 57–64.
- Edet, A. E. & Offiong, O. E. (2002). Evaluation of water quality pollution indices for heavy metal contamination monitoring. A study case from Akpabuyo-Odukpani area, Lower Cross River Basin (southeastern Nigeria). *GeoJournal* 57(4), 295–304.
- Friesen, J., Sinobas, L. R., Foglia, L. & Ludwig, R. (2017). Environmental and socio-economic methodologies and solutions towards integrated water resources management. *Science of the Total Environment* 581–582, 906–908.
- Gentry, R. W., Larsen, D. & Ivey, S. (2003). Efficacy of genetic algorithm to investigate small scale aquitard leakage. *Journal of Hydraulic Engineering* 129(7), 527–535.
- Ghosh, S. & Mujumdar, P. P. (2006). Risk minimization in water quality control problems of a river system. *Advances in Water Resources* 29(3), 458–470.
- Gichuki, F. N., Kodituwakku, D. C., Nguyen-Khoa, S. & Hoanh, C. T. (2009). Cross-scale trade-offs and synergies in aquaculture, water quality and environment: research issues and policy implications. *Water Policy* 11(S1), 1–12.
- Hair, J. F., Anderson, R. E., Tatham, R. L. & Black, W. C. (1998). *Multivariate Data Analysis*. Prentice-Hall International, Upper Saddle River, New Jersey, USA.
- Haydar, S., Arshad, M. & Aziz, J. A. (2016). Evaluation of drinking water quality in urban areas of Pakistan: a case study of Southern Lahore. *Pakistan Journal of Engineering and Applied Sciences* 5, 16–23.
- Ignatiades, L., Karydis, M. & Vounatsou, P. (1992). A possible method for evaluating oligotrophy and eutrophication based on nutrient concentration scales. *Marine Pollution Bulletin* 24(5), 238–243.
- Kazi, T. G., Arain, M. B., Jamali, M. K., Jalbani, N., Afridi, H. I., Sarfraz, R. A., Bajj, J. A. & Shah, A. Q. (2009). Assessment of water quality of polluted lake using multivariate statistical techniques: a case study. *Ecotoxicology and Environmental Safety* 72(2), 301–309.
- Kong, Y., Pang, Z., Wu, C. & Jing, M. (2014). A comparative study of single factor and multivariate statistical methods for surface water quality assessment. *Water Policy* 16(1), 157–167.
- Kuo, J. T., Wang, Y. Y. & Lung, W. S. (2006). A hybrid neural-genetic algorithm for reservoir water quality management. *Water Research* 40(7), 1367–1376.
- Lai, C., Chen, X., Chen, X., Wang, Z., Wu, X. & Zhao, S. (2015). A fuzzy comprehensive evaluation model for flood risk based on the combination weight of game theory. *Natural Hazards* 77(2), 1243–1259.
- Liu, K., Elliott, J. A., Lobb, D. A., Flaten, D. N. & Yarotski, J. (2013). Critical factors affecting field-scale losses of nitrogen and phosphorus in spring snowmelt runoff in the Canadian prairies. *Journal of Environmental Quality* 42, 484–496. doi:10.2134/jeq2012.0385.
- Lobato, T. C., Hauser-Davis, R. A., Oliveira, T. F., Silveira, A. M., Silva, H. A. N., Tavares, M. R. M. & Saraiva, A. C. F. (2015). Construction of a novel water quality index and quality indicator for reservoir water quality evaluation: a case study in the Amazon region. *Journal of Hydrology* 522, 674–683.

- Lu, R. S. & Lo, S. L. (2002). Diagnosing reservoir water quality using self-organizing maps and fuzzy theory. *Water Research* 36(9), 2265–2274.
- McNeill, M., Thro, E. & McAllister, M. N. (1996). Fuzzy logic: a practical approach. *SIAM Review* 38(1), 173–173.
- Ott, W. (1978). *Water Quality Indices: A Survey of Indices Used in the United States*, Vol. 1. Office of Monitoring and Technical Support, Office of Research and Development, Environmental Protection Agency, Washington, DC, USA.
- Parinet, B., Lhote, A. & Legube, B. (2004). Principal component analysis: an appropriate tool for water quality evaluation and management – application to a tropical lake system. *Ecological Modelling* 178(3–4), 295–311.
- Pearson, K. (1901). Principal components analysis. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 6(2), 559.
- Prasad, B. & Bose, J. (2001). Evaluation of the heavy metal pollution index for surface and spring water near a limestone mining area of the lower Himalayas. *Environmental Geology* 41(1–2), 183–188.
- Sauthier, M. C. D. S., Silva, E. G. P. D., Santos, B. R. D. S., Caldas, J. D. C., Minho, L. A. C., Santos, A. M. P. D. & Santos, W. N. L. D. (2018). Screening of mangifera indica, l. functional content using PCA and neural networks (ANN). *Food Chemistry* 273, 115–123.
- Savenije, H. H. & Van der Zaag, P. (2008). Integrated water resources management: concepts and issues. *Physics and Chemistry of the Earth, Parts A/B/C* 33(5), 290–297.
- Sawatsky, M. L., Clyde, M. & Meek, F. (2015). Partial least squares regression in the social sciences. *The Quantitative Methods for Psychology* 11(2), 52–62.
- Schulze, F. H., Wolf, H., Jansen, H. W. & Van der Veer, P. (2005). Applications of artificial neural networks in integrated water management: fiction or future? *Water Science and Technology* 52(9), 21–31.
- Selle, B., Schwientek, M. & Lischeid, G. (2013). Understanding processes governing water quality in catchments using principal component scores. *Journal of Hydrology* 486, 31–38.
- Singh, K. P., Malik, A., Mohan, D. & Sinha, S. (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) – a case study. *Water Research* 38(18), 3980–3992.
- Sun, Z., Zhang, J., Zhang, Q., Hu, Y., Yan, D. & Wang, C. (2014). Integrated risk zoning of drought and waterlogging disasters based on fuzzy comprehensive evaluation in Anhui province, China. *Natural Hazards* 71(3), 1639–1657.
- Toledo, L. G. D. & Nicolella, G. (2002). Water quality index for agricultural and urban watershed use. *Scientia Agricola* 59(1), 181–186.
- Vaze, J. & Chiew, F. H. (2003). Comparative evaluation of urban storm water quality models. *Water Resources Research* 39(10).
- Vonberg, D., Vanderborght, J., Cremer, N., Pütz, T., Herbst, M. & Vereecken, H. (2014). 20 years of long-term atrazine monitoring in a shallow aquifer in western Germany. *Water Research* 50, 294–306.
- Wang, D., Singh, V. P. & Zhu, Y. (2007). Hybrid fuzzy and optimal modeling for water quality evaluation. *Water Resources Research* 43(5).
- Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S. & Bai, X. (2015). Flood hazard risk assessment model based on random forest. *Journal of Hydrology* 527, 1130–1141.
- Xue, X. H. & Yang, X. G. (2014). Seismic liquefaction potential assessed by fuzzy comprehensive evaluation method. *Natural Hazards* 71, 2101–2112.
- Yang, B., Lai, C., Chen, X., Wu, X. & He, Y. (2018). Surface water quality evaluation based on a game theory-based cloud model. *Water* 10(4), 510.

Received 11 July 2018; accepted in revised form 13 January 2019. Available online 18 February 2019