

Dynamic water quality evaluation based on fuzzy matter–element model and functional data analysis, a case study in Poyang Lake

Bing Li^{1,2,3} · Guishan Yang¹ · Rongrong Wan¹ · Georg Hörmann³

Received: 14 December 2016 / Accepted: 23 May 2017 / Published online: 28 June 2017
© Springer-Verlag Berlin Heidelberg 2017

Abstract Comprehensively evaluating water quality with a single method alone is challenging because water quality evaluation involves complex, uncertain, and fuzzy processes. Moreover, water quality evaluation is limited by finite water quality monitoring that can only represent water quality conditions at certain time points. Thus, the present study proposed a dynamic fuzzy matter–element model (D–FME) to comprehensively and continuously evaluate water quality status. D–FME was first constructed by introducing functional data analysis (FDA) theory into a fuzzy matter–element model and then validated using monthly water quality data for the Poyang Lake outlet (Hukou) from 2011 to 2012. Results showed that the finite water quality indicators were represented as dynamic functional curves despite missing values and irregular sampling time. The water quality rank feature curve was integrated by the D–FME model and revealed comprehensive and continuous variations in water quality. The water quality in Hukou showed remarkable seasonal variations, with the best water quality in summer and worst water quality in winter. These trends were significantly correlated with water level fluctuations ($R = -0.71$, $p < 0.01$). Moreover, the extension weight curves of key indicators indicated that total

nitrogen and total phosphorus were the most important pollutants that influence the water quality of the Poyang Lake outlet. The proposed D–FME model can obtain scientific and intuitive results. Moreover, the D–FME model is not restricted to water quality evaluation and can be readily applied to other areas with similar problems.

Keywords Water quality evaluation · Fuzzy matter–element model · Functional data analysis · Poyang Lake

Introduction

Water is essential to all life forms and access to freshwater with desirable quality is a prerequisite to sustainable development (Srebotnjak et al. 2012). However, water is often polluted with chemical, physical, and biological contaminants mainly caused by anthropogenic activities (Smith et al. 1999; Vitousek et al. 1997). Many monitoring programs and protocols, such as the Harmonized Monitoring Scheme in Britain (Hurley et al. 1996), the National Water Quality Assessment in the USA (Kolpin et al. 1998), and the National Monitoring and Assessment Program in Denmark (Conley et al. 2002), have been implemented to provide a comprehensive image of water quality status. These monitoring programs, however, produce finite parameter observations that can only represent water quality condition at certain time points. Moreover, water quality evaluation involves complex, uncertain and fuzzy processes (Ip et al. 2007). Therefore, developing a comprehensive and dynamic water quality evaluation model is necessary for water resource management and protection.

Water quality in freshwater bodies is a complex issue that results from physical, chemical, and biological processes, and the interactions among these processes (Taner et al. 2011). Various methods, including multivariate statistical analysis

Responsible editor: Marcus Schulz

✉ Guishan Yang
gsyang@niglas.ac.cn

¹ Key Laboratory of Watershed Geographic Sciences, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China

² University of Chinese Academy of Sciences, Beijing 100049, China

³ Department of Hydrology and Water Resources Management, Institute of Natural Resource Conservation, Kiel University, 24118 Kiel, Germany

(Kazi et al. 2009; Ouyang et al. 2006; Vega et al. 1998), artificial neural networks (Gazzaz et al. 2012), water quality index (Abtahi et al. 2015; Lermontov et al. 2009), fuzzy comprehensive assessment (Zou et al. 2006), and matter element analysis (Liu and Zou 2012; Wong and Hu 2014), have been developed in recent decades to characterize and evaluate water quality. However, these mathematical evaluation methods have limitations. For example, multivariate statistical analysis requires large water quality samples to obtain reasonable results, and the artificial neural network lacks accurate analysis for each performance index (Deng et al. 2015); Although fuzzy theory is widely used because of its advantages in solving problems with fuzzy boundaries and controlling the effect of sampling errors on evaluation results (Liu et al. 2010; Ocampo-Duque et al. 2006), it cannot effectively distinguish adjacent characteristic indicators. The fuzzy matter–element (FME) model combines the advantages of both fuzzy and matter element theories (Deng et al. 2015). It can analyze data in the forms of intervals and solve contradictory problems (Wong and Hu 2014). In addition, using fuzzy membership in FME can efficiently deal with the fuzziness and uncertainty of the evaluation process. Deng et al. (2015) applied the FME model to assess the river health status of the Taihu Lake basin. They found that the FME model has more advantages in reflecting objective factors compared with other comprehensive assessment methods. The FME model efficiently considers the uncertainty and fuzziness in the optimized selection of environmental monitoring points (Wang et al. 2015). Zhang et al. (2011) also found that the FME model is practical and reliable in the evaluation of environmental impact for land use planning.

The FME model, however, is effective only when no value is missing from the dataset, which is difficult to satisfy in field monitoring. Furthermore, dynamic water quality evaluation is always limited by finite sampling campaigns that can only represent water quality at finite discrete time points (e.g., weekly, monthly, and seasonal samplings) (Yan et al. 2015). Functional data analysis (FDA) has emerged as an effective approach toward modeling time series data and has received attention in public health and biomedical applications (Ullah and Finch 2013). The principle of FDA is to express discrete observations from a time series in the form of a function that represents all monitored data as a single observation (Ramsay 2006; Ramsay and Silverman 2002). The temporal variations of water quality parameters are described by continuous smooth dynamics which allows accurate estimates of parameters to use in analysis. Given that the FDA approach is highly flexible, the timing intervals of data observations do not have to be equally spaced and missing values are acceptable (Müller et al. 2011). FDA has received considerable attention (Haggarty et al. 2012; Henderson 2006) since its introduction in water quality research in 1997 (Champely and Doledec 1997). FDA has also been used in water quality evaluation.

Yan et al. (2015) incorporated FDA with commonly used water quality index to solve the shortcomings in the dynamic assessment, which uses variable weights and water quality index. The same authors also introduced FDA into fuzzy set theory to develop a dynamic variable fuzzy set assessment model (Yan et al. 2016).

This study firstly developed the dynamic fuzzy matter–element (D-FME) model by introducing the FDA theory into the FME model, to comprehensively and continuously evaluate the water quality evaluation. A 2-year monthly water quality dataset in the outlet of Poyang Lake (Hukou) was used to examine the reasonability and strength of the D-FME model.

Materials and methods

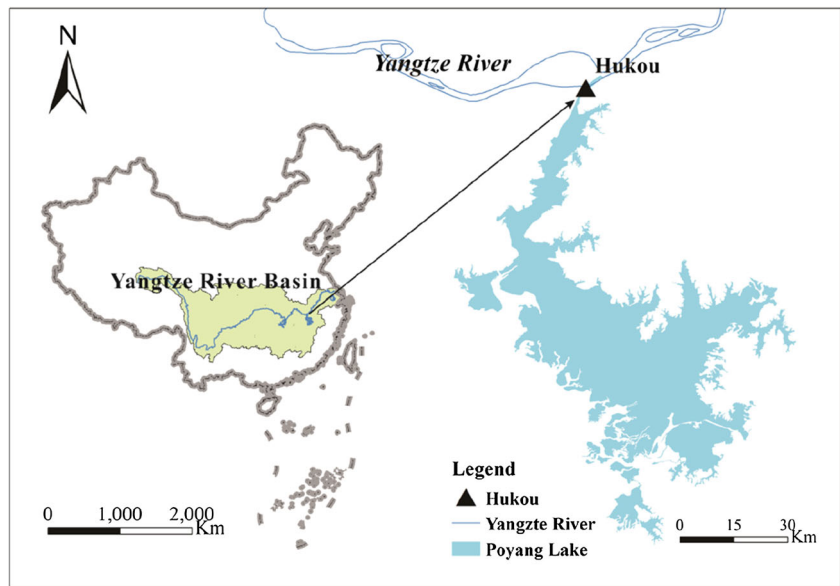
Study area

The Poyang Lake (115°47′–116°45′E, 28°22′–29°45′N) is located on the south bank of the Yangtze River. The lake provides a habitat for rare migratory birds in winter and has been designated as a globally important eco-region by the World Wide Fund for Nature (Tang et al. 2016). The lake nourishes a drainage area of 16.22×10^4 km² and plays a significant role in supplying freshwater and fish, restricting flooding, regulating local climate, and degrading pollutants in the area. It has a subtropical monsoon climate with an average annual temperature of 17.6 °C and a mean annual precipitation of 1450–1550 mm that falls mostly during summer. The lake is freely connected to the Yangtze River at Hukou (Fig. 1). The water quality of Hukou partially determines the water environment of the lower reaches of the Yangtze River.

Data

According to the pollution features of the Poyang lake and traditional water quality indices in China (Ban et al. 2014; Li et al. 2017), dissolved oxygen (DO), chemical oxygen demand (COD_{Mn}), ammonium nitrogen (NH₄-N), total nitrogen (TN), and total phosphorus (TP) were selected as key indicators for water quality evaluation. DO reflects the overall state of the water quality and low DO concentrations are harmful to aquatic organisms. COD_{Mn} is an indicator of organic pollution. NH₄-N was proved to be an important nutrient indicator of that controls the presence frequency of aquatic vegetation (Zhang et al. 2016). TN and TP are key indicators of lake eutrophication. Water quality was classified into five classes (Table 1) in accordance with the National Surface Water Environmental Quality Standards of China (Chinese Environmental Protection Agency 2002). It should be

Fig. 1 Location of the Poyang Lake and monitoring sites



noted that the water quality standards are expressed in intervals rather than values, and classification I indicates the best water quality. Monthly water quality data from 2011 to 2012 including key indicators and corresponding sampling data were used in this study (Hydrology Bureau of the Poyang Lake 2010). No sampling was conducted in January 2012. DO and TN concentrations were missing in September 2011 and November 2012, respectively. Thus, they were excluded from the FME model. The daily water level data at Hukou gauge station was obtained from the Hydrological Bureau of Jiangxi Province, China.

Fuzzy matter–element model

Generation of compound fuzzy matter–element

The complexity and uncertainty of water quality evaluation have been pointed out by previous studies (Beck 1987; Ip et al. 2007). The matter–element model, which is based on classical and fuzzy mathematics, is commonly used to solve complex and uncertain problems (Cai 1999). The fuzzy

matter–element is defined with triple ordered “objects, characteristics, and fuzzy values,” which were denoted as $R = (N, C, u)$ (Cai 1999). Accordingly, the compound fuzzy matter–element of the m th water quality object (R^m) is established by n feature vectors C_i (indicators), k water quality classifications G_j , and the corresponding fuzzy value u_{ij} and can be expressed as the following matrix:

$$R^m_{nk} = \begin{bmatrix} & G_1 & G_2 & \cdots & G_k \\ C_1 & u_{11} & u_{12} & \cdots & u_{1k} \\ C_2 & u_{21} & u_{22} & \cdots & u_{2k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ C_n & u_{n1} & u_{n2} & \cdots & u_{nk} \end{bmatrix} \tag{1}$$

where C_i is the i th indicator and $i = 1, 2, \dots, n$; G_j is the j th water quality classification and $j = 1, 2, \dots, k$; u_{ij} is the fuzzy membership degree of the i th indicator of j th classification and is calculated based on fuzzy membership functions and corresponding water quality classifications. Previous studies have shown that numerous observations can be approximately regarded as normal distributions with membership functions under a similar category (Li et al. 2011). Therefore, in the present study, we used normal membership function instead of triangular or lower semi-trapezoidal functions. The normal membership function is expressed as follows:

$$u_{ij} = \exp \left[- \left(\frac{x_i - a_{ij}}{b_{ij}} \right)^2 \right] \tag{2}$$

where x_i is the measured concentration value of the i th indicator, a_{ij} and b_{ij} are the characteristic parameters of the normal distribution function that meet the conditions of $a_{ij} > 0$ and $b_{ij} > 0$. In Formula (2), $u_{ij} = 1$ when $x_i = a_{ij}$, therefore a_{ij} is the average value of the j th classification criterion of the i th

Table 1 Water quality classifications based on the Environmental Quality Standards for Surface Water, China (GB3838-2002)

Parameter	Classification of water quality				
	I	II	III	IV	V
DO (mg/L) \geq	7.5	6	5	3	2
COD _{Mn} (mg/L) \leq	2	4	6	10	15
TN (mg/L) \leq	0.2	0.5	1	1.5	2
TP (mg/L) \leq	0.01	0.025	0.05	0.1	0.2
NH ₄ -N (mg/L) \leq	0.15	0.5	1	1.5	2

indicator, which is the middle value of each interval $[x_p, x_q]$. Here $j = 1, 2, \dots, k$. Therefore, the formula of a_{ij} is as follows:

$$a_{ij} = \frac{x_p + x_q}{2} \tag{3}$$

where x_p and x_q are the lower and upper boundary values of the j th classification criterion of the i th indicator, respectively.

In Table 1, the boundary values of the classifications are transition values from one classification to another. Therefore, the membership functions of two adjacent grades should be equal at boundary values, as shown by the following equation:

$$u_{ij} = \exp \left[- \left(\frac{x_q - x_q + x_p}{b_{ij}} \right)^2 \right] = \frac{1}{2} \tag{4}$$

$$b_{ij} = \frac{|x_q - x_p|}{2\sqrt{\ln 2}} \tag{5}$$

a_{ij} and b_{ij} are calculated for each indicator and water quality grade in accordance with the water quality classification boundary values of all indicators. Thus, the compound FME matrix was obtained.

Determination of extension weights

In accordance with the matter–element extension theory, the extension weights of each indicator for each object were calculated with the simple correlative function (Wong and Hu 2014). The simple correlative function assigns weights based on the correlation between the values of each indicator and the corresponding water quality classification. Therefore, worse indicators have higher weights, which is consistent with the idea of the “bucket effect” (Li et al. 2017). For the m th evaluation object, let

$$r_{ij} = \begin{cases} \frac{2(x_i - x_p)}{x_q - x_p}, & \text{if } x_i \leq \frac{x_p + x_q}{2} \\ \frac{2(x_q - x_i)}{x_q - x_p}, & \text{if } x_i > \frac{x_p + x_q}{2} \end{cases} \tag{6}$$

where r_{ij} is the simple correlative function of the i th indicator of the j th water quality classification; x_i is the measured value of the i th indicator, and let

$$r_{ij_{\max}} = \max_{j=1,2,\dots,k} (r_{ij}) \tag{7}$$

Then, the representative water quality indicators are identified either as benefit (i.e., the larger, the better) or cost (i.e., the smaller, the better). For benefit indicators, such as DO, we define

$$r_i = \begin{cases} (n - j_{\max}) \times (1 + r_{ij_{\max}}), & \text{if } r_{ij_{\max}} \geq -0.5 \\ (n - j_{\max}) \times 0.5, & \text{if } r_{ij_{\max}} < -0.5 \end{cases} \tag{8}$$

Meanwhile, for cost indicators

$$r_i = \begin{cases} j_{\max} \times (1 + r_{ij_{\max}}), & \text{if } r_{ij_{\max}} \geq -0.5 \\ j_{\max} \times 0.5, & \text{if } r_{ij_{\max}} < -0.5 \end{cases} \tag{9}$$

The limiting condition of Formulas (8)–(9) occurs when certain indicators, especially for pollutants like TN and TP, exceed the water quality classification boundary, which may happen in natural water bodies. Finally, the weight of indicator w_i is given by

$$w_i = \frac{r_i}{\sum_{i=1}^n r_i} \tag{10}$$

Calculation of fuzzy neartude

Fuzzy neartude is the measure of proximity between evaluated objects and water quality standards (Deng et al. 2015). It avoids negative values in the matter–element evaluation process by introducing a similarity measure (Wong and Hu 2014). A greater value of fuzzy neartude indicates that the evaluated objects are closer to a certain grade. The fuzzy neartude matrix of the m th evaluation object was obtained with Hamming neartude (ρH_j) (Teng et al. 2012), as follows:

$$\rho H_j = 1 - \sum_{i=1}^k w_i |\tilde{u}_{ij} - \tilde{u}_{i0}| \tag{11}$$

where \tilde{u}_{ij} is the scaled fuzzy matter–element matrix (u_{ij}), and

$$\tilde{u}_{i0} = \begin{vmatrix} r_{10} \\ r_{20} \\ \vdots \\ r_{n0} \end{vmatrix} = \begin{vmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{vmatrix} \tag{12}$$

Generation of non-integral feature value

The water quality feature value of the m th evaluation object was calculated as follows:

$$J_m = \sum_{j=1}^k j \times \left(\frac{\rho H_j}{\sum_{j=1}^k \rho H_j} \right) \tag{13}$$

where J_m is the non-integral water quality rank feature value of the m th evaluation object. A low value for J_m indicates better water quality and vice versa. Here, in reference to similar studies that used the FME model (Deng et al. 2015), we defined water quality as excellent when $1 < J_m < 1.5$ and as bad when $4.5 < J_m < 5$. The water quality grades are good, medium, and poor when $i - 0.5 < J_m < i + 0.5$, $i = 2, 3, 4$, respectively.

Dynamic fuzzy matter–element model

The FDA theory was introduced into the proposed FME model to dynamically evaluate water quality based on finite water quality sampling campaigns. The basic principle of FDA is to express discrete observations in the form of continuous function curves, thus enabling accurate parameter estimation and effective data noise reduction (Ramsay and Dalzell 1991; Ramsay and Silverman 2002). FDA methods are not necessarily based on the assumption that the concentration values observed at different time points are independent (Ullah and Finch 2013). In FDA, the sampling timing intervals of data observations do not have to be equally spaced, which is often the case in reality. The key step of FDA is smoothing, which converts raw discrete data points into smoothly varying functions. Various smoothing techniques, such as B-splines, Fourier smoothing, cubic smoothing, and wavelet smoothing, have been proposed (Henderson 2006). Cubic smoothing is unsuitable for regions of complex change, where it may get over-smooth easily. Fourier smoothing is more suitable for periodic data. Wavelet smoothing can efficiently deal with data with frequent and severe fluctuations, whereas B-spline smoothing is the best choice for representing non-periodic data (Henderson 2006; Ullah and Finch 2013), which is the case in the present study. B-spline smoothing is the most popular technique because of its simplicity and flexibility in addressing nonparametric and semi-parametric modeling situations (Ullah and Finch 2013; Yan et al. 2015). Therefore, the present study utilized a fourth-order B-spline smoothing technique. The smoothing procedure is expressed as follows:

$$x_i(t) = \sum_{p=1}^K d_{ip} \varphi_p(t) \quad (14)$$

where $x_i(t)$ is the concentration curves of the i th indicator; $\varphi_p(t)$ is the p th basis function and $p = 1, 2, \dots, K$, K is the total number of B-spline basis functions; and d_{ip} is the coefficient of the i th indicator of the p th basis function.

The implementation of the D–FME model is illustrated in Fig. 2. A roughness penalty approach was defined to trade off

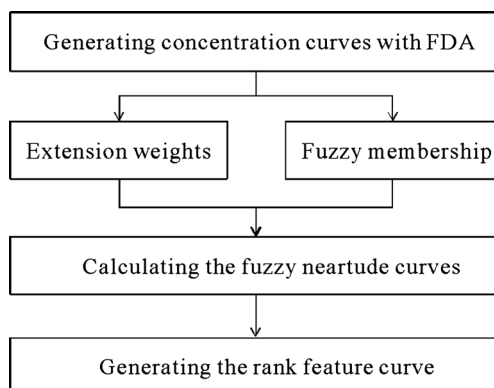


Fig. 2 Water quality evaluation process with the D–FME model

curve roughness against data fitting, and the smoothing parameter (λ) was used to control the smoothness of the curves (Ramsay et al. 2009). Furthermore, the optimal value for the smoothing parameter was identified with the generalized cross-validation measure (GCV) that was developed by Wahba and Craven (1978). The Package *fda* that was developed by Ramsay and Silverman (2002; 2006) for R version 3.3.1 (Team R 2012) specifically to support FDA is available at <<https://cran.r-project.org/web/packages/fda/index.html>>.

Results and discussion

Concentration curves of indicators

The 2-year monthly sampling observations for each indicator were converted to a continuous concentration curve via FDA (Fig. 3). As shown in Fig. 3a, the DO concentration curve showed similar variations in 2011 and 2012. A decreasing trend was observed during the first half of the year, reached the minimum during summer (July–September), and then gradually increased during the second half of the year. These trends were in accordance with temperature variations, which are negatively correlated with DO dynamics (Vega et al. 1998). Compared with other lakes, the average COD_{Mn} concentration in Hukou was relatively low at 2.6 mg/L (Beyhan and Kaçikoç 2014; Taner et al. 2011). In addition, the COD_{Mn} concentration curve, which is shown in Fig. 3b, showed insignificant temporal variations with little oscillation from 2011 to 2012. These patterns indicated a relatively low level of organic pollution in Poyang Lake that was contributed by anthropogenic sources, such as industrial and domestic sewage. Similar with DO, the concentration curves of $\text{NH}_4\text{-N}$, TN, and TP all showed low peaks during summer in 2011–2012, which is in accordance with the highest discharge and water levels in summer. During summer, the pollutants in the lake are more degraded and diluted because lake tributaries produce substantial freshwater discharge, which significantly increases the degradation capacity of the water (Li et al. 2016). Moreover, the degradation capacities are low in winter and high in summer because of the variations in water quantity and temperature (Yan et al. 2015). In 2011, $\text{NH}_4\text{-N}$, TN, and TP all decreased before July and then sharply increased after July; these trends can be attributed to the drastic transition between flood and dry in 2011. The water level decreased substantially after July 2011 (Fig. 6). The TN and $\text{NH}_4\text{-N}$ concentration curves peaked in December and January 2012 (winter), whereas the TP concentration peaked in October 2011 and remained low from January to October in 2012. This phenomenon may be attributed to the decrease in TP pollutant flux in 2012.

It is to be noted that the sampling time of every month was not equally spaced, which is common in practical sampling

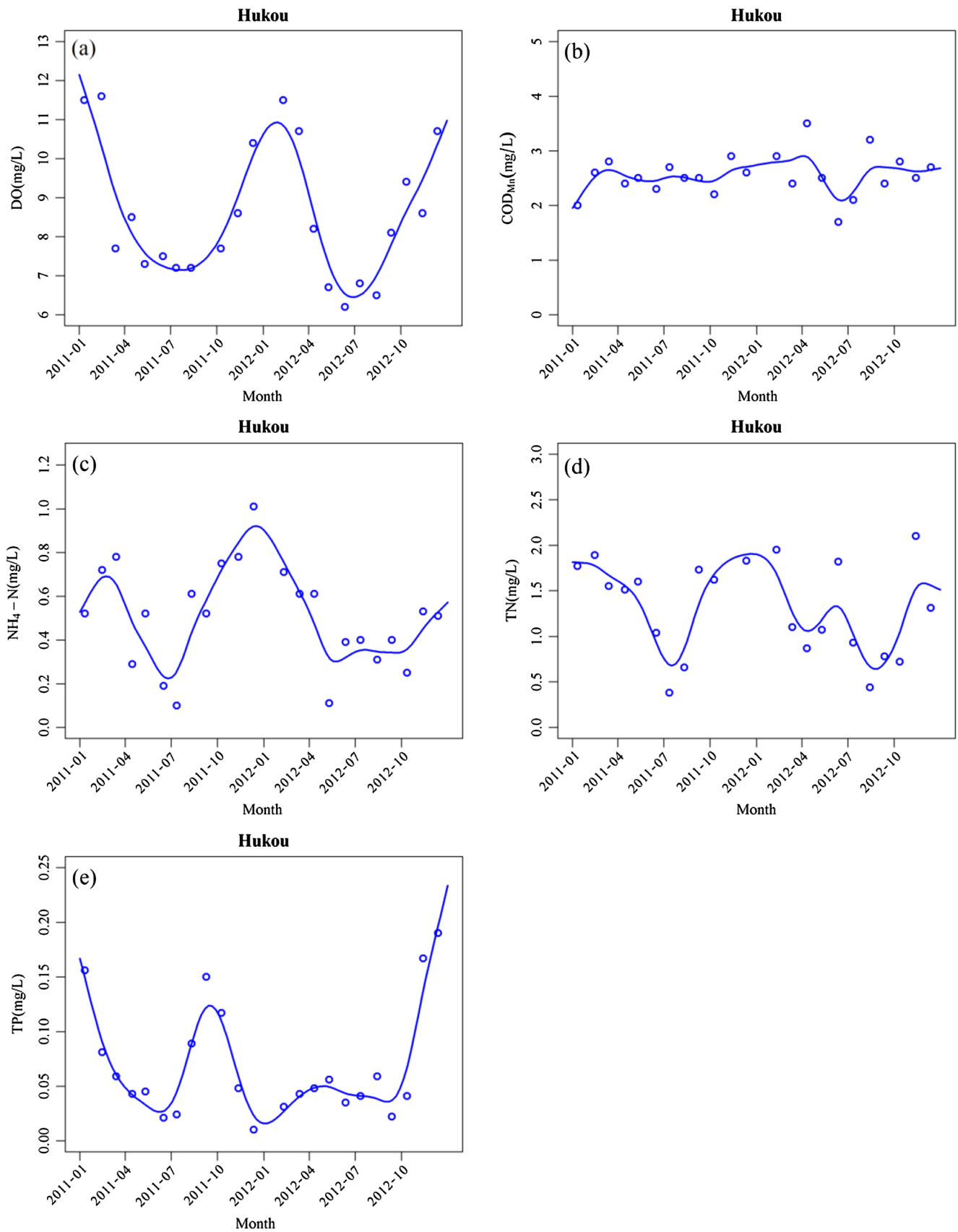


Fig. 3 Observed data and concentration values of **a** DO, **b** COD_{Mn}, **c** NH₄-N, **d** TN, and **e** TP

campaigns. For example, the first three sampling campaigns were implemented on 11/1/2011, 14/2/2011, and 13/3/2011 (Fig. 3). Moreover, missing values exist for each indicator in February 2012, and DO and TN were also missing for September 2011 and June 2012, respectively. Nonetheless, these problems can be efficiently addressed by FDA via the B-spline smoothing technique. Figure 3 showed that although the concentration curves did not precisely pass through each observation value, the observed variation trends of all indicators were correctly captured. This observation can be attributed to monitoring errors, which are assumed to be inevitable in FDA. The optimal curve was determined as the one with the minimum mean square error (Yan et al. 2016). However, uncertainty still existed when high intensity rainfall events caused monitoring outliers and uncontrolled sewage was discharged into lake, which contributed to deviations between the concentration curves and the observation values (Seiler et al. 2015; Zhou et al. 2012), for example, the deviation of TN concentration curve in June and November in 2011.

Dynamic extension weight curves

The dynamic extension weight curves, which are based on simple correlative function, were calculated with Formulas (6)–(10). The results are presented in Fig. 4. In 2011 and 2012, TN and TP had average weights of 0.32 and 0.26. These values were remarkably higher than those of other indicators and indicated their high influence on the overall water quality of Poyang Lake. From a temporal perspective, the weight of TN was high from January to April when TN concentration was relatively high and reached its minimum in July for both 2011 and 2012 when the TN concentration was lowest within the year. Similar variation can be observed for $\text{NH}_4\text{-N}$ weight curve. By contrast, the weight of DO was low from

January to March (winter) when the oxygen demand of aquatic organisms was low. Then, DO gradually increased and reached a maximum average weight of 0.19 during the spring and summer because of various redox reaction activities, including breeding of fishes and growth of planktons and aquatic vegetation. The weight of DO began to decrease gradually again in winter. Remarkable low peak in weight curve of TP was observed in January 2012. The weight curve of COD_{Mn} exhibited little fluctuations through the year and had a mean weight of 0.13. These dynamics of the weight curves were consistent with those of concentration curves, which were influenced by pollutant effluents, dilution effect, and degradation processes.

Unlike in previous studies, the combination of FDA and matter–element extension theory produced the dynamic curves of the weight variations in indicators throughout the year (Akkoyunlu and Akiner 2012; Gharibi et al. 2012; Abtahi et al. 2015). The simple correlative function method determines weights based on concentration curves and corresponding water quality grades (Wong and Hu 2014). Moreover, the more polluting the indicator is, the higher its weight would be during evaluation. This method is similar to the principle of the “bucket effect,” which emphasizes the effect of the extreme indicator on the overall evaluation results. It is also consistent with an existing weighting method that is based on the relative importance and degree of pollution (Semiromi et al. 2011; Yan et al. 2015).

Results of dynamic water quality evaluation

The different parameter groups (a and b) of normal distribution for each water classification were determined with Formulas (3,4 and 5) and in accordance with the water quality classifications of the five indicators presented in Table 1. Fuzzy membership degree was obtained via Formula (2) and the concentration curves. Then, the finite and continuous fuzzy neartude curves to the five water quality classifications (i.e., I, II, III, IV, and V) were generated. For simplicity, only fuzzy neartude to classifications I, II, IV, and V are shown in Fig. 5. Fuzzy neartude measures the fuzzy closeness degree of the water sample to water quality classifications. This method outperforms the traditional correlative degree method in matter–element theory by avoiding negative values in the correlative degree (Wong and Hu 2014). A high fuzzy neartude value indicates a close proximity between the water sample and water quality classifications (Teng et al. 2012). Figure 5a showed that the water quality conditions in Poyang Lake were related differently to classifications I, II and classes IV, V. These relationships can be divided into three periods. Firstly, during January and December, the fuzzy neartude to classification V was considerably higher than that of other classifications, indicating relatively poor water quality during this period. Secondly, during summer (June–August), the fuzzy

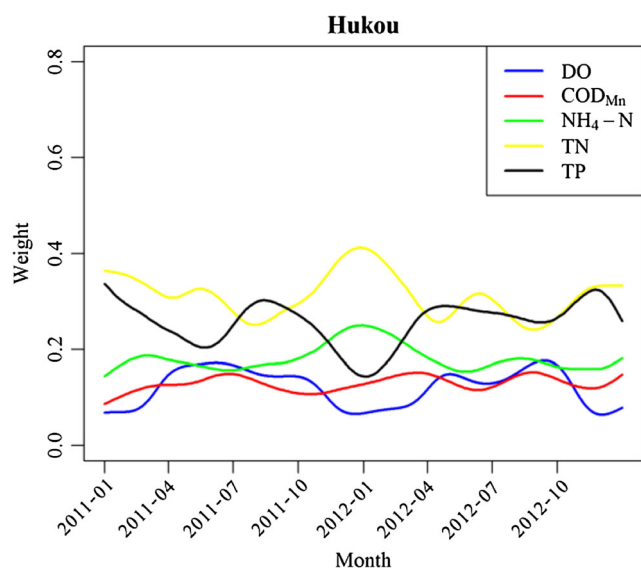


Fig. 4 Dynamic weight curves of the five indicators

neartude to classifications I and II was significantly higher than that of classifications IV and V, which indicated that the water quality condition during this period was the most desirable. The fuzzy neartude to classification V was almost zero, which signified zero possibility that water quality belonged to classification V during summer. The third period (February to May and September to November) was mixed without clear information on which classification the water quality should belong to (Fig. 5a). Therefore, judging water quality condition based on the maximum fuzzy neartude is not scientific. Instead, water quality conditions should be objectively and intuitively characterized with the non-integral water quality rank feature value.

Figure 5b presents the fuzzy neartude to water quality classifications as a bar plot. Although the results of fuzzy neartude on sampling date were almost identical with Fig. 5a at specific time points, the fuzzy neartude curves were more intuitive and comprehensive. Moreover, because of the missing values in the dataset, the fuzzy neartude in June 2011, January 2012, and November 2012 cannot be calculated with the FME model. The efficient use of smoothing techniques in FDA has also been reported by Yan et al. (2016).

The non-integral water quality rank feature curve was generated based on Formula (13). The monthly discrete values from 2011 to 2012 are exhibited in Fig. 6. The comprehensive water quality rank feature curve in Poyang Lake outlet (Hukou) showed similar seasonal variations in 2011 and 2012; specifically a decreasing trend from January to July and an increasing trend in the second half of the year. These trends indicated that water quality condition improved from winter to summer and deteriorated from autumn to winter. We extracted the finite rank feature values of each day from the curve. The average water quality rank feature values were 3.27 and 3.05 in 2011 and 2012, which can be classified as medium in accordance with the rules set in the “Materials and methods” section. The

highest rank feature value was observed in December and January (winter), whereas the lowest was observed in July. Similar to the concentrations curves of indicators, a sharp increase was observed after July 2011. The water quality evaluation results for the Poyang Lake that were obtained by the D–FME model were consistent with those of obtained by fuzzy theory (Li et al. 2017). The rank feature value clearly showed a reverse variation trend with water level variation in Poyang Lake. Correlation analysis showed that water quality rank feature value and water level were significantly correlated ($R = -0.71, p < 0.01$), which indicated that water level variations greatly influenced pollutant transportation and degradation. This result was consistent with previous studies. For example, Li et al. (2016) examined the seasonal pollutant concentration variations and their relationship with hydrological conditions and found that water quality is likely to be good during periods of high water levels. Liu et al. (2016) also concluded that water level fluctuations are the principal drivers of physicochemical variables in Poyang Lake. Similar studies have also shown the great influence of water level variations on water quality and aquatic ecosystem health (Coops et al. 2003; O’Farrell et al. 2011).

FME theory focuses on solving fuzzy incompatibility, as well as comprehensively and objectively reflecting the influence of all indicators. Li and Li (2014) have proven that the FME model with Hamming approach degree proficiently handles fuzzy and incompatible problems in evaluating the ecosystem health of Beijing and Shanghai. Zhu et al. (2016) also reported that the FME model was comparable to the fuzzy optimum method and Technique for Order of Preference by Similarity to Ideal Solution method when evaluating different reservoir flood control operation alternatives. The evaluation results from finite observations were consistent with the rank feature curve. Although it captured the overall variation of synthetic water quality, less information was obtained. Moreover,

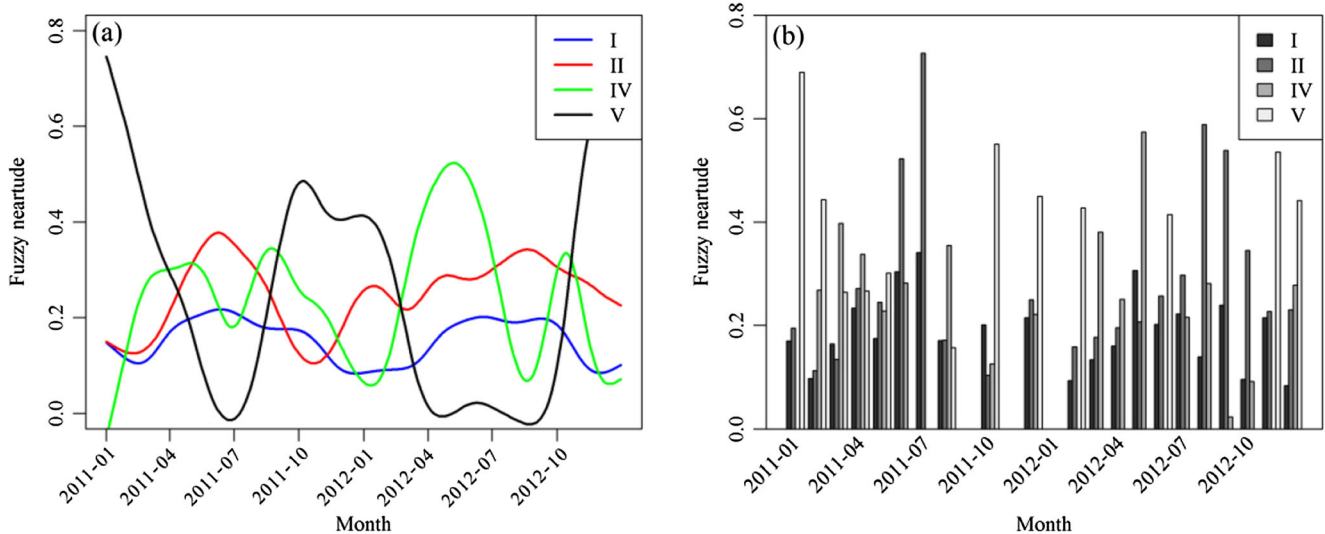


Fig. 5 Fuzzy neartude curves (a) and finite values (b) to classifications I, II, IV, and V

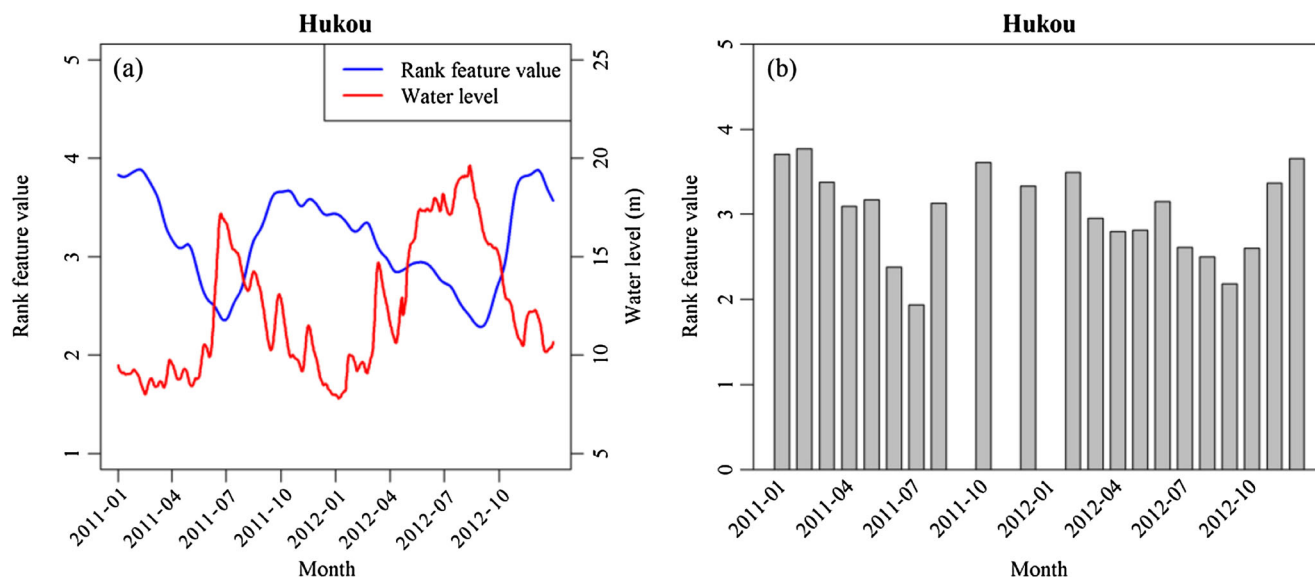


Fig. 6 Water quality rank feature curve (a) and monthly finite values (b) of Poyang Lake

sampling months with missing values were excluded in the FME model. However, the D-FME model efficiently solved the problem of missing data by calculating the quantitative concentration curves that were generated by smoothing. The FME model is a case of D-FME model at specific time points. Therefore, the combination of FME theory and FDA is of great use for the comprehensive and continuous evaluation of water quality. The D-FME model can provide intuitive and efficient results and can be utilized for similar problems in other fields.

Conclusion

The present study proposed the D-FME model, which was constructed by incorporating the FME model with FDA theory, for the effective and continuous evaluation of variations in water quality. The weights of water quality indicators for each sample were calculated with the simple relative function. Finite water quality data were converted into continuous concentration curves by the B-spline smoothing technique. The non-integral water quality rank feature value was obtained via fuzzy nearness method. The proposed model was successfully applied to the outlet of the Poyang Lake (Hukou).

FME theory is suitable for multifactor evaluation and can comprehensively and objectively reflect the impacts of all indicators. However, it cannot address the problem of missing values. Moreover, FME can only represent water quality conditions at specific time points. The D-FME model efficiently solved these problems and integrated finite water quality indicators into a single water quality rank feature curve, thereby making water quality evaluation comprehensive and intuitive.

The average non-integral water quality rank feature value of Poyang Lake was 3.27 and 3.05 in 2011 and 2012, respectively, which denoted medium water quality. The DO, $\text{NH}_4\text{-N}$,

TN, and TP concentration curves all showed similar seasonal variation within the year. The water quality evaluation results based on the D-FME model also showed the same trend, with the best water quality in summer and worst water quality in winter. Moreover, the water quality in Poyang Lake was significantly correlated with water level variations. The proposed D-FME is not restricted to water quality evaluation and can be readily applied to other areas.

Acknowledgments This research was financially supported by the National Basic Research Program of China (973 Program) (Grant 2012CB417006) and the National Scientific Foundation of China (Grant 41571107 and 41601041). The authors would also like to thank the editor and reviewers for their extensive review that significantly improved the manuscript.

References

- Abtahi M, Golchinpour N, Yaghmaeian K et al (2015) A modified drinking water quality index (DWQI) for assessing drinking source water quality in rural communities of Khuzestan Province, Iran. *Ecol Indic* 53:283–291
- Akkoyunlu A, Akiner ME (2012) Pollution evaluation in streams using water quality indices: a case study from Turkey's Sapanca Lake Basin. *Ecol Indic* 18:501–511
- Ban X, Wu Q, Pan B et al (2014) Application of composite water quality identification index on the water quality evaluation in spatial and temporal variations: a case study in Honghu Lake, China. *Environ Monit Assess* 186(7):4237–4247
- Beck MB (1987) Water quality modeling: a review of the analysis of uncertainty. *Water Resour Res* 23(8):1393–1442
- Beyhan M, Kaçikoç M (2014) Evaluation of water quality from the perspective of eutrophication in Lake Eğirdir, Turkey. *Water Air Soil Poll* 225(7):1–13
- Cai W (1999) Extension theory and its application. *Chinese Sci Bull* 44(17):1538–1548

- Champely S, Doledec S (1997) How to separate long-term trends from periodic variation in water quality monitoring. *Water Res* 31(11):2849–2857
- Chinese Environmental Protection Agency (2002) National surface water environmental quality standards of China (GB3838-2002). China Standards Press, Beijing (in Chinese)
- Conley DJ, Markager S, Andersen J (2002) Coastal eutrophication and the Danish national aquatic monitoring and assessment program. *Estuaries* 25(4):848–861
- Coops H, Beklioglu M, Crisman TL (2003) The role of water-level fluctuations in shallow lake ecosystems—workshop conclusions. *Hydrobiologia* 506(1–3):23–27
- Deng X, Xu Y, Han L et al (2015) Assessment of river health based on an improved entropy-based fuzzy matter-element model in the Taihu plain, China. *Ecol Indic* 57:85–95
- Gazzaz NM, Yusoff MK, Aris AZ et al (2012) Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Mar Pollut Bull* 64(11):2409–2420
- Gharibi H, Mahvi AH, Nabizadeh R et al (2012) A novel approach in water quality assessment based on fuzzy logic. *J Environ Manag* 112:87–95
- Haggarty R, Miller C, Scott E et al (2012) Functional clustering of water quality data in Scotland. *Environmetrics* 23(8):685–695
- Henderson B (2006) Exploring between site differences in water quality trends: a functional data analysis approach. *Environmetrics* 17(1):65–80
- Hurley M, Currie J, Gough J et al (1996) A framework for the analysis of harmonised monitoring scheme data for England and Wales. *Environmetrics* 7(4):379–390
- Ip W, Hu B, Wong H et al (2007) Applications of rough set theory to river environment quality evaluation in China. *Water Resour* 34(4):459–470
- Kazi T, Arain M, Jamali M (2009) Assessment of water quality of polluted lake using multivariate statistical techniques: a case study. *Ecotox Environ Safet* 72(2):301–309
- Kolpin DW, Barbash JE, Gilliom RJ (1998) Occurrence of pesticides in shallow groundwater of the United States: initial results from the National Water-Quality Assessment Program. *Environ Sci Technol* 32(5):558–566
- Lermontov A, Yokoyama L, Lermontov M et al (2009) River quality analysis using fuzzy water quality index: Ribeira do Iguape river watershed, Brazil. *Ecol Indic* 9(6):1188–1197
- Li Y, Li D (2014) Assessment and forecast of Beijing and Shanghai's urban ecosystem health. *Sci Total Environ* 487:154–163
- Li H, Gu C, Liang T et al (2011) A new perspective of ecosystem health. *J Forestry Res* 22(1):127–132
- Li B, Yang G, Wan R et al (2016) Spatiotemporal variability in the water quality of Poyang Lake and its associated responses to hydrological conditions. *Water* 8(7):296
- Li B, Yang G, Wan R et al (2017) Using fuzzy theory and variable weights for water quality evaluation in Poyang Lake, China. *Chinese Geogr Sci* 27(1):39–51
- Liu D, Zou Z (2012) Water quality evaluation based on improved fuzzy matter-element method. *J Environ Sci* 24(7):1210–1216
- Liu L, Zhou J, An X et al (2010) Using fuzzy theory and information entropy for water quality assessment in three gorges region, China. *Expert Syst Appl* 37(3):2517–2521
- Liu X, Teubner K, Chen Y (2016) Water quality characteristics of Poyang Lake, China, in response to changes in the water level. *Hydrol Res* 47(S1):238–248
- Müller HG, Sen R, Stadtmüller U (2011) Functional data analysis for volatility. *J Econometrics* 165(2):233–245
- O'Farrell I, Izaguirre I, Chaparro G et al (2011) Water level as the main driver of the alternation between a free-floating plant and a phytoplankton dominated state: a long-term study in a floodplain lake. *Aquat Sci* 73(2):275–287
- Ocampo-Duque W, Ferre-Huguet N, Domingo JL et al (2006) Assessing water quality in rivers with fuzzy inference systems: a case study. *Environ Int* 32(6):733–742
- Ouyang Y, Nkedi-Kizza P, Wu Q et al (2006) Assessment of seasonal variations in surface water quality. *Water Res* 40(20):3800–3810
- Ramsay JO (2006) Functional data analysis. Wiley Online Library
- Ramsay JO, Dalzell C (1991) Some tools for functional data analysis. *Journal of the Royal Statistical Society Series B (Methodological)*: 539–572
- Ramsay JO, Silverman BW (2002) Applied functional data analysis: methods and case studies, vol 77. Citeseer
- Ramsay J O, Hooker G, Graves S (2009) Functional data analysis with R and MATLAB. Springer Science & Business Media
- Seiler LM, Fernandes EHL, Martins F et al (2015) Evaluation of hydrologic influence on water quality variation in a coastal lagoon through numerical modeling. *Ecol Model* 314:44–61
- Semiromi FB, Hassani A, Torabian A et al (2011) Evolution of a new surface water quality index for Karoon catchment in Iran. *Water Sci Technol* 64(12):2483–2491
- Smith VH, Tilman GD, Nekola JC (1999) Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. *Environ Pollut* 100(1):179–196
- Srebotnjak T, Carr G, Sherbinin A et al (2012) A global water quality index and hot-deck imputation of missing data. *Ecol Indic* 17:108–119
- Taner MÜ, Üstün B, Erdinçler A (2011) A simple tool for the assessment of water quality in polluted lagoon systems: a case study for Küçükçekmece Lagoon, Turkey. *Ecol Indic* 11(2):749–756
- Tang X, Li H, Xu X et al (2016) Changing land use and its impact on the habitat suitability for wintering Anseriformes in China's Poyang Lake region. *Sci Total Environ* 557:296–306
- Team R (2012) Development core. R: A language and environment for statistical computing
- Teng J, Zhang T, Lu W (2012) Structural stress identification using fuzzy pattern recognition and information fusion technique. *J of Civ Eng Architect* 6(4):479
- Ullah S, Finch CF (2013) Applications of functional data analysis: a systematic review. *BMC Med Ees Methodol* 13(1):1
- Vega M, Pardo R, Barrado E et al (1998) Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Res* 32(12):3581–3592
- Vitousek PM, Mooney HA, Lubchenco J et al (1997) Human domination of Earth's ecosystems. *Science* 277(5325):494–499
- Wahba G, Craven P (1978) Smoothing noisy data with spline functions. Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer Math* 31:377–404
- Wang H, Liu Z, Sun L et al (2016) Optimal design of river monitoring network in Taizihe River by matter element analysis. *PLoS One* 10(5):e0127535
- Wong H, Hu BQ (2014) Application of improved extension evaluation method to water quality evaluation. *J Hydrol* 509:539–548
- Yan F, Liu L, Li Y et al (2015) A dynamic water quality index model based on functional data analysis. *Ecol Indic* 57:249–258
- Yan F, Liu L, Zhang Y et al (2016) The research of dynamic variable fuzzy set assessment model in water quality evaluation. *Water Resour Manag* 30(1):63–78
- Zhang J, Wang K, Chen X et al (2011) Combining a fuzzy matter-element model with a geographic information system in eco-environmental sensitivity and distribution of land use planning. *Int J Environ Res Public Health* 8(4):1206–1221
- Zhang Y, Liu X, Qin B (2016) Aquatic vegetation in response to increased eutrophication and degraded light climate in Eastern Lake Taihu: implications for lake ecological restoration. *Sci Rep* 6:23867

- Zhou W, Yin K, Harrison PJ et al (2012) The influence of late summer typhoons and high river discharge on water quality in Hong Kong waters. *Estuar Coast Shelf S* 111:35–47
- Zhu F, Zhong P, Xu B et al (2016) A multi-criteria decision-making model dealing with correlation among criteria for reservoir flood control operation. *J Hydroinf* 18(3):531–543
- Zou Z, Yi Y, Sun J (2006) Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment. *J Environ Sci* 18(5):1020–1023