ORIGINAL ARTICLE

Spatial–temporal characteristics of surface water quality in the Taihu Basin, China

Guangju Zhao · Junfeng Gao · Peng Tian ·
Kun Tian · Guangheng Ni

Abstract Taihu Basin is one of the most developed and industrialized regions in China. In the last two decades, rapid development of economy as well as an increase in population has resulted in an increase of pollutants produced and discharged into rivers and lakes. Much more attention has been paid on the serious water pollution problems due to high frequency of algal blooming. The dataset, obtained during the period 2001–2002 from the Water Resources Protection Bureau of the Taihu Basin, consisted of eight physicochemical variables surveyed monthly at 22 sampling sites in the Taihu Basin, China. Principal component analysis (PCA) and cluster analysis (CA) were used to identify the characteristics of the surface water quality in the studied area. The temporal and spatial variations of water quality were also evaluated by using the fuzzy synthetic evaluation (FSE) method. PCA extracted the first two principal components (PCs), explaining 86.18% of the total variance of the raw data. Especially, PC1 (73.72%) had strong positive correlation with DO, and was negatively associated with COD_{Mn}, COD, BOD, NH_4^+–N, TP and TN. PC2 (12.46%) was characterized by pH. CA showed that most sites were highly polluted by industrial and domestic wastewater which contributed significantly to PC1. The sites located in the west of Lake Taihu were influenced by farmland runoff which may contribute to nitrogen pollution of Lake Taihu, whereas the monitoring sites in the eastern of Lake Taihu demonstrated that urban residential subsistence and domestic wastewater are the major contaminants. FSE indicates that there is no obvious variance between 2001 and 2002 among most sites. Only several sites free from point-source pollution appear to exhibit good water quality through the studied period.

Keywords Surface water quality · Principal component analysis · Cluster analysis · Fuzzy synthetic evaluation · Taihu Basin

G. Zhao (✉)
Institute of Soil and Water Conservation, Chinese Academy and Sciences and Ministry of Water Resources, P.O. Box 26, Yangling 712100, Shannxi, People's Republic of China
e-mail: guangjuzhao@yahoo.com

G. Zhao
Department of Hydrology and Water Resources Management, Institute of Natural Resources Conservation, Kiel University, Olshausenstr. 75, 24118 Kiel, Germany

J. Gao
Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, People's Republic of China

P. Tian
College of Water Resources and Architectural Engineering, Northwest A & F University, Yangling 712100, People's Republic of China

K. Tian · G. Ni
State Key Laboratory of Hydro-Science and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing, People's Republic of China

Introduction

With the increase of population and rapid economic development, surface water has received large amounts of pollution from a variety of sources such as urban wastewater, industrial and agricultural activities, as well as the assimilation and transport of pollution effluents (Bowen and Depledge 2006; Milovanovic 2007). Human health is largely influenced by water pollution, as water pollution

affects sustainability of the aquatic ecosystem and social economic development in some regions. Nutrients from farmland fertilizer and urban wastewater discharged into rivers contribute main pollutants to the surface water in a catchment, thereby tending to induce serious ecological problems such as eutrophication and environmental pollution (Wang et al. 2007). It is therefore essential and urgent to control water pollution and improve the water quality, and regularly implement monitoring programs which give the tools to help understand the spatial and temporal variations of the surface water quality.

Taihu Basin is one of the most industrialized regions in China with high population density, urbanization, and economic development. The area covers only 0.4% of territory of China while contributing about 11% of Gross National Product (GNP) and more than 14% of China's gross domestic production (Qin et al. 2007). Since the 1980s, rapid development of local economy and increased population and urbanization has resulted in pollutants being produced and discharged into rivers and lakes. In recent years, serious water pollution problems have extracted much attention due to the fact that algal blooms occurred much more frequently, extending its coverage, while simultaneously persisting throughout the summer. This environmental issue seriously affects the lake, as a supply of drinking water (Qin et al. 2007). In the late May of 2007, Lake Taihu was overtaken by a major algae bloom, leaving approximately two million people without drinking water for at least 1 week.

In order to examine the pollution sources and improve the surface water quality, numerous studies have focused on eco-environmental issues in Lake Taihu and analyzed the loss and load of pollutants from different pollution sources (Zhang et al. 2003; Gao et al. 2004; Guo et al. 2004; Qin et al. 2007). However, research on the water quality at the whole catchment scale is very limited, thus improvements to research should be undertaken. A large number of studies focused on the pollution and eutrophication of Lake Taihu (Chen et al. 2003), whereas others focused on the status of pollution in principal rivers around the lake (Wang et al. 2007; Xie et al. 2007; Xu et al. 2009). These studies provided good insights into the spatial and temporal characteristics of the surface water quality in the Taihu Basin. However, the intensive agricultural activities and the rapid urbanization of the basin heavily influenced the natural flows, well developed drainage network which received large amounts of contaminants from industry, domestic wastewater, agriculture, aquiculture and livestock, and thus not only making it difficult to identify the pollutants sources but also assess characteristics of the surface water quality (Xu et al. 2009).

To better understand temporal and spatial patterns of water pollution and variance of the aquatic ecosystems,

monitoring data are usually interpreted by applying multivariate statistical techniques (Yidana et al. 2008), such as: cluster analysis (CA), principal component analysis (PCA), discriminant analysis (DA) and factor analysis (FA). The multivariate statistical techniques are considered as useful tools for both simplifying the complicated data sets consisting of water quality variables, and extracting meaningful interpretation (Vega et al. 1998; Wunderlin et al. 2001; Bu et al. 2009). In this study, physicochemical parameters of water quality at 22 sites were surveyed and analyzed monthly from 2001 to 2002 in the Taihu Basin. Three different multivariate statistical techniques: PCA, CA, and fuzzy synthetic evaluation (FSE) were adopted to assess the characteristics of water quality and classification of water samples in the surface water bodies. The objectives of this paper are to: (1) identify the spatial and temporal characteristics of the physicochemical variables of water quality, (2) assess surface water quality using FSE, and (3) determine important factors and sources influencing the water quality.

Materials and methods

Study area

Taihu Basin is located near the downstream of the Yangtze River (Fig. 1), including parts of Jiangsu Province, Zhejiang Province, Anhui province and Shanghai city in administration with an area of approximately 36,895 km². It is dominated by subtropical summer monsoons, with an average annual rainfall about 1,177 mm. There are various kinds of topographical situations, not only high mountainous and hilly area towards west, but also low alluvial plains in northern and eastern parts, occupying 80% of the whole basin (Fig. 1). The drainage networks are well developed and heavily diverted by human beings for flood control and agricultural irrigation (Fig. 1). There are more than 200 rivers distributed in the whole watershed, and 172 rivers or channels connected to Lake Taihu (Xu and Qin 2005). The total length of the rivers in the Taihu Basin is ca. 12,000 km, i.e., about 3.24 km km⁻² (Qin et al. 2007). The greatest inflow rivers that bring most of the pollutants into the lake are Chendonggang, Xitiaoxi, and Yincungang, located in the west and southwest. The main outflow rivers are Taipu, Xinyunhe, and Xijiang in the southeast.

Sampling sites and analysis

Due to the large area of the Taihu Basin, it is difficult to collect the recent measurement data in the whole region. Water samples from 67 different sites in the Taihu Basin were collected at monthly intervals from 2000 to 2002

Fig. 1 Location of the study
area and monitoring sites
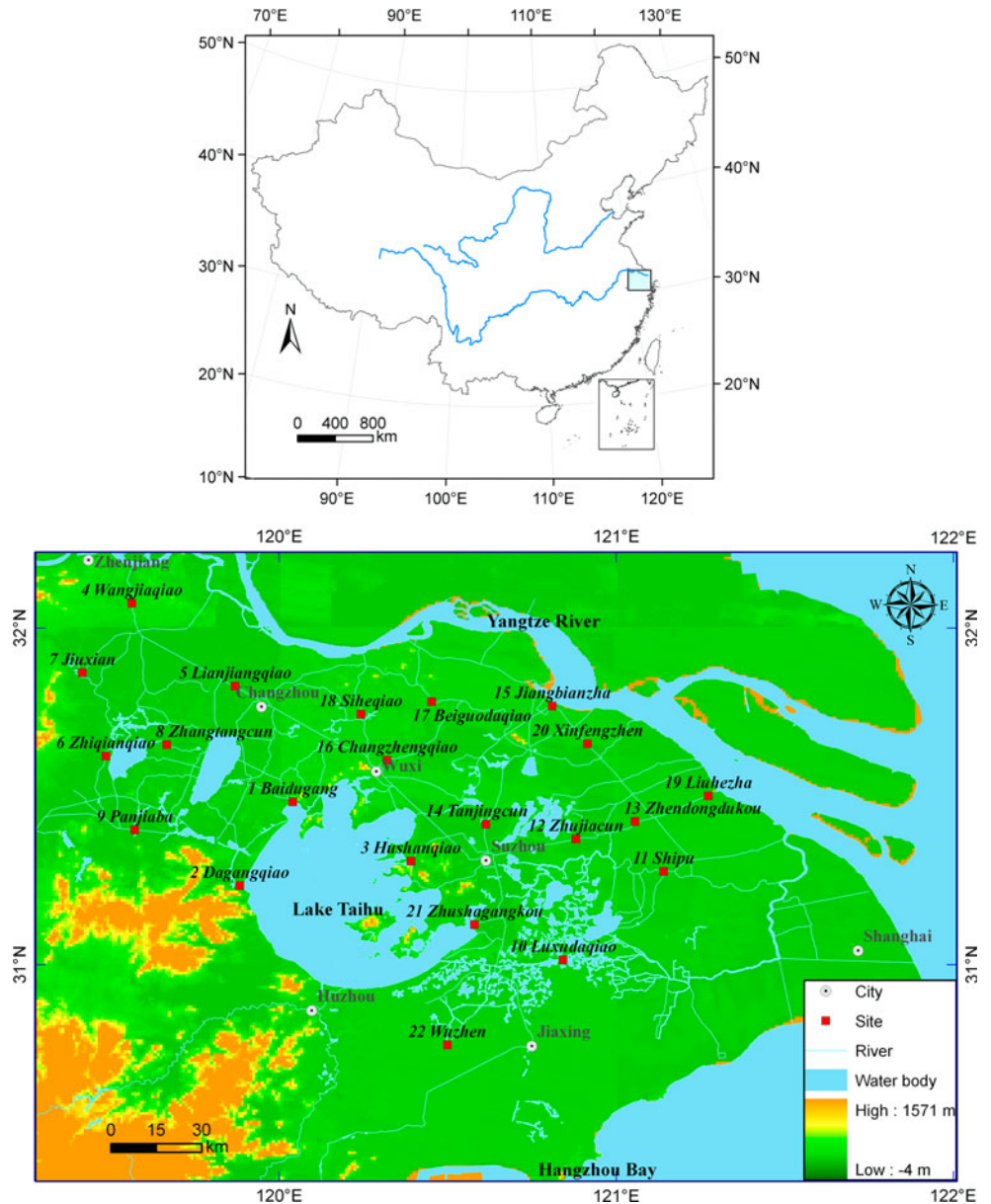
from the Water Resources Protection Bureau of the Taihu
Basin (Fig. 1). According to the data completeness and
homogeneity of the sites in the same reaches, we here only
selected 22 sites between 2001 and 2002 for water quality
analysis. Four sites were selected at the mouth of the rivers
connected with Lake Taihu, and 18 stations were selected
from the main streams and tributaries. The datasets con-
sisted of 24 water quality variables: temperature, electrical
conductivity, pH, total suspended solid (TSS), Secchi disc
depth (transparency), total nitrogen (TN), ammonia nitro-
gen (NH_4^+–N), chlorophyll-*a*, total phosphorus (TP),
biologic oxygen demand (BOD), chemical oxygen demand
in manganese (COD_{Mn}), chemical oxygen demand in chro-
mium (COD), dissolved oxygen (DO), petroleum, volatile

phenol, mercury, copper, lead, zinc, cadmium, chromium,
iron, hardness, and nitrate (nitrogen). Due to adverse weather
conditions, several samples were blank during some periods,
such as April 2001. In addition, the monitoring of trace metal
contaminants is discontinuous. Based on sampling continu-
ity of the sites, we selected 8 physicochemical variables for
our study, which consists of physical properties, organic
constituents, and nutrients.

Data analysis methods

Multivariate statistical approaches are widely and com-
prehensively used to identify the variances or similarities
of the environmental areas in term of surface water quality

parameters (Maria and Graca 2006). PCA and CA were employed to sort the water quality variables and sampling sites, respectively. PCA was performed using Statistica 7.0, and CA was applied using the SPSS 16.0 statistic software packages. FSE was adopted to identify the pollution level of surface water using Matlab R2008a.

Principal component analysis

PCA is used to reduce the dimensionality of the data set by explaining the correlation among several random uncorrelated environmental variables in terms of a small number of underlying factors or principal components without extreme loss of information (Vega et al. 1998). The methodology will not be extensively reviewed here due to its wide application in environmental science and a larger number of documents providing a thorough description of its formulation and properties (e.g. Emery and Thomson 1997; Jolliffe 2002; Wall et al. 2003).

To objectively isolate the most important modes of variance of the water quality parameters, we applied PCA on the normalized data set of selected eight indicators, including pH, DO, COD, COD_{Mn}, BOD, NH_4^+–N, TP, and TN. The factor loadings are usually presented as correlation coefficients between the associated principal components (PCs), and can be considered as a measure of the relative importance in the extracted PCs. Any factor with an eigenvalue greater than unity (eigenvalue >1) was considered significant according to the criteria of Cattell and Jaspers (1967).

Cluster analysis

Cluster analysis is a multivariate technique that can classify categories or clusters to reveal their intrinsic characteristics based on their similarity (Vega et al. 1998). Hierarchical clustering is the most common approach that uses Ward's (1963) linkage as a measure of similarity. Unlike PCA that normally uses only two or three PCs for displaying key variances of the parameters, CA uses all the variance or information contained in the original data set to demonstrate the similarity and proximity. The results of CA are usually presented by a dendrogram which provides a visual summary of the clustering results. The visual summary explains the internal homogeneity and external heterogeneity of the objects (Singh et al. 2004; Shrestha and Kazama 2007). Many studies have shown that CA reliably classifies surface water quality and can guide future sampling strategies (Singh et al. 2004; Xu et al. 2009). In the present study, CA was performed on the normalized data set by means of the Wads' method, using Euclidean distances as a measure of similarity.

Fuzzy synthetic evaluation

Fuzzy set theory is used for decision-making or pattern recognition when the context of the problem is unclear and boundaries are undefined or imprecise. The FSE method evaluates each individual variable's value according to predefined quality criteria in the fuzzy environment by designing a suitable membership function and using the fuzzy operators (Cude 2001). In this approach, water classes are defined as fuzzy sets in terms of degrees of membership with flexible boundaries rather than binary/crisp sets (Dahiya et al. 2007). Recent studies indicate that FSE has become a useful tool which is extensively applied throughout the world in decision-making and evaluation processes in imprecise environments. Icaga (2007) proposed an index model to evaluate the surface water quality using fuzzy logic method in terms of physical and inorganic chemical parameters. Dahiya et al. (2007) used FSE to assess the physicochemical variables of groundwater for drinking purposes. Since the FSE method has been deeply discussed in Yen and Langari (1999), Ross (2004), Dahiya et al. (2007) and Lu et al. (2009), only a brief description of the procedure is presented below.

1. Select assessment variables and establish assessment criteria:

   The first important step is to select the representative and rational water quality assessment variables. For each site, an assessment indicator matrix U can be expressed as:

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

   where n is the number of the selected variables. The water quality is classified into five levels in terms of National Surface Water Environmental Quality Standards (Chinese Environmental Protection Agency 2002b, GB3838-2002). Afterwards, the assessment criteria matrix V can be expressed as follows:

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

   where m is the number of assessment criteria categories, which equal 5 in this study.

2. Establish membership functions:

   The membership functions represent the degree to which the water quality contaminants belong to the fuzzy set. For an element u_i of U, the value r_{ij} is called the membership degree of u_i in the fuzzy set V. The value 0 means that u_i is not a member of the fuzzy set; the value 1 means that u_i is fully a member of the fuzzy set. The values between 0 and 1 characterize fuzzy members, which belong to the fuzzy set only partially. Various kinds of membership functions can be used to quantify the membership degree, such as: the Gaussian distribution function, the sigmoid curve, and

quadratic and cubic polynomial curves. In this study, the triangular membership function is applied to evaluate the water quality variables as follows. The parameters in the following equations are shown in Table 1, obtained from the National Surface Water Environmental Quality Standards (Chinese Environmental Protection Agency 2002b, GB3838-2002).

$$r_{i1}(u_i) = \begin{cases} 1 & u_i \leq a \\ (u_i - b)/(a - b) & a < u_i < b \\ 0 & u_i \geq b \end{cases} \quad (3)$$

$$r_{i2}(u_i) = \begin{cases} 0 & u_i \leq a, u_i \geq c \\ (u_i - a)/(b - a) & a < u_i < b \\ (u_i - c)/(b - c) & b < u_i < c \end{cases} \quad (4)$$

$$r_{i3}(u_i) = \begin{cases} 0 & u_i \leq b, u_i \geq d \\ (u_i - b)/(c - b) & b < u_i < c \\ (u_i - d)/(c - d) & c < u_i < d \end{cases} \quad (5)$$

$$r_{i4}(u_i) = \begin{cases} 0 & c \leq v_{ij-1}, u_i \geq e \\ (u_i - c)/(d - c) & c < u_i < d \\ (u_i - e)/(d - e) & d < u_i < e \end{cases} \quad (6)$$

$$r_{i5}(u_i) = \begin{cases} 0 & u_i \leq d \\ (u_i - d)/(e - d) & d < u_i < e \\ 1 & u_i \geq e \end{cases} \quad (7)$$

3. Calculate the membership function matrix:

After substituting the monitoring data of each assessment variables, the fuzzy matrix R can be expressed as:

$$R = (r_{ij})_{n \times m} = \begin{pmatrix} r_{11} & r_{12} & \ldots & r_{1m} \\ r_{21} & r_{22} & \ldots & r_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ r_{n1} & r_{n2} & \ldots & r_{nm} \end{pmatrix} \quad (8)$$

where $r_{ij}(i = 1, 2, \ldots, n; \quad j = 1, 2, \ldots, m)$ is the membership degree of the ith assessment variable at the jth level.

Table 1 The limits of the membership function based on surface water environmental quality standards

Variables	The parameters of the membership functions				
	a	b	c	d	e
pH ≥ 7.5	7.5	8	8.5	8.75	9
pH ≤ 7.5	6	6.25	6.5	7	7.5
DO (mg/l)[a]	7.5	6	5	3	2
COD$_{Mn}$ (mg/l)	2	4	6	10	15
COD (mg/l)	12	15	20	30	40
BOD (mg/l)	2	3	4	6	10
NH$_4^+$–N (mg/l)	0.25	0.5	1.0	1.5	2.0
TP (mg/l)	0.02	0.1	0.2	0.3	0.4
TN (mg/l)	0.2	0.5	1.0	1.5	2.0

[a] The number membership functions in reverse order

4. Determine the weights matrix:

Water quality cannot be determined according to a single water quality indicator. Each water quality variable has its own contribution to water quality (Shen et al. 2005; Lu et al. 2009). Therefore, it is necessary to determine the weights of each variable in FSE. In this study, the entropy method is used to determine the weights of assessment variables (Zou et al. 2006; Lu et al. 2009). The weight of entropy of the ith assessment variable is defined as:

$$w_i = \frac{1 - H_i}{n - \sum_{i=1}^{n} H_i} \quad (9)$$

where $0 \leq w_i \leq 1$ $\sum_{i=1}^{n} w_i = 1$. H_i is the entropy of the ith indicator, calculated as:

$$H_i = -k \sum_{j=1}^{p} f_{ij} \ln f_{ij}, \quad i = 1, 2, \ldots, n \quad (10)$$

where $f_{ij} = r_{ij}/\sum_{j=1}^{p} r_{ij}$, $k = 1/\ln p$, p is the number of monitoring sites (22 in this article), and f_{ij} is the normalized value of the ith variable at the jth monitoring site. If $f_{ij} = 0$, then $\ln f_{ij}$ is supposed to be zero.

5. Calculate evaluation results:

The surface water quality assessment results can be obtained from:

$$D = W \cdot R = (d_1, d_2, \ldots, d_m) \quad (11)$$

in which the fuzzy matrix $R = (r_{ij})_{n \times m}$, weight matrix $W = (w_i)_{1 \times n}$, and the results matrix D is calculated as:

$$d_j = \sum_{1}^{n} w_i r_{ij}, \quad j = 1, 2, \ldots, m, \quad (12)$$

The final evaluation result is:

$$\text{Class} = \text{index}(\max(d_j)) \quad j = 1, 2, \ldots, m. \quad (13)$$

The index of the maximum value of d_j (i.e. j) denotes the water quality class of the water samples.

Results and discussion

Status of surface water quality in the Taihu Basin

Due to a large quantity of observed data, only the mean values and standard deviations of the measured water quality variables at 22 sites are shown in Table 2. The mean values of the pH vary from 7.04 to 7.87 which falls within the range of National Surface Water Environmental Quality Standards of China (Chinese Environmental Protection Agency 2002b, GB3838-2002).

The lowest DO value was found at Changzhengqiao (1.08 mg/L), and the relatively lower values at Xinfengzhen,

Table 2 Mean and standard deviations (SD) of physicochemical variables of surface water quality in the Taihu Basin

Variables		PH	DO	COD$_{Mn}$	COD	BOD	NH$_4^+$–N	TP	TN
1. Baidugang	Mean	7.31	4.06	7.72	24.79	9.00	3.70	0.29	4.97
	SD	0.21	2.88	2.38	9.44	4.22	2.55	0.18	2.62
2. Dagangqiao	Mean	7.04	5.95	3.58	12.77	3.75	0.62	0.09	2.66
	SD	0.46	3.51	1.87	6.46	2.32	0.84	0.08	0.96
3. Hushanqiao	Mean	7.64	8.28	3.87	19.21	2.77	0.53	0.08	1.66
	SD	0.37	2.46	0.91	7.57	1.01	0.84	0.03	1.49
4. Wangjiaqiao	Mean	7.40	6.30	4.93	15.25	1.85	2.37	0.19	4.24
	SD	0.20	1.56	0.92	6.42	0.61	2.06	0.12	2.20
5. Lianjiangqiao	Mean	7.52	7.01	4.50	15.43	2.47	0.82	0.23	3.27
	SD	0.23	1.82	1.58	9.25	1.56	0.55	0.09	1.35
6. Zhiqianqiao	Mean	7.56	4.57	5.70	21.28	3.71	1.81	0.23	5.92
	SD	0.35	1.94	0.93	5.60	1.26	0.86	0.07	2.18
7. Jiuxian	Mean	7.63	5.45	5.36	20.81	3.48	1.58	0.21	5.20
	SD	0.40	3.11	1.22	3.97	1.14	1.33	0.09	2.04
8. Zhangtangcun	Mean	7.55	5.35	4.89	18.25	3.55	0.92	0.18	3.76
	SD	0.40	2.59	1.16	4.44	1.34	1.09	0.13	1.85
9. Panjiaba	Mean	7.32	4.84	5.86	19.78	6.90	1.94	0.22	4.06
	SD	0.18	1.82	1.03	6.26	3.49	0.91	0.10	1.51
10. Luxudaqiao	Mean	7.53	6.90	4.38	16.81	1.73	0.68	0.11	1.48
	SD	0.23	1.81	0.73	4.32	0.94	0.67	0.06	0.87
11. Shipu	Mean	7.42	5.11	6.46	24.57	4.37	2.20	0.23	3.32
	SD	0.35	1.94	0.73	9.03	1.60	0.98	0.09	1.17
12. Zhujiacun	Mean	7.35	5.02	6.62	22.07	3.21	2.32	0.20	4.32
	SD	0.17	2.77	1.11	4.45	0.78	1.23	0.07	1.49
13. Zhendongdukou	Mean	7.46	3.95	5.97	24.73	5.28	2.25	0.25	3.49
	SD	0.34	1.86	0.92	4.37	2.14	0.79	0.11	1.19
14. Tanjingcun	Mean	7.58	5.82	5.01	16.09	2.89	1.17	0.11	2.10
	SD	0.20	0.84	1.11	2.59	0.99	0.83	0.04	1.19
15. Jiangbianzha	Mean	7.87	7.45	4.19	14.39	2.56	0.60	0.10	1.35
	SD	0.28	1.76	1.74	3.37	1.08	0.69	0.03	1.01
16. Changzhengqiao	Mean	7.42	1.08	12.00	57.04	11.50	5.83	0.64	11.01
	SD	0.27	0.78	3.88	26.41	7.54	2.79	0.40	4.35
17. Beiguodaqiao	Mean	7.44	3.21	8.46	27.00	5.58	0.97	0.15	5.37
	SD	0.28	0.33	1.33	4.22	0.48	0.19	0.04	2.61
18. Siheqiao	Mean	7.62	4.41	6.90	24.13	4.97	0.93	0.16	5.54
	SD	0.25	1.19	1.84	4.51	0.73	0.29	0.04	2.92
19. Liuhezha	Mean	7.40	4.89	5.94	22.71	3.13	2.17	0.16	3.34
	SD	0.37	1.61	1.36	7.09	1.64	1.52	0.07	1.64
20. Xinfengzhen	Mean	7.49	3.80	6.40	24.71	4.97	2.32	0.23	3.78
	SD	0.28	1.61	0.94	6.04	2.42	0.95	0.14	1.04
21. Zhushagangkou	Mean	7.43	7.17	5.53	23.42	3.04	0.64	0.12	2.03
	SD	0.39	2.06	0.77	10.84	1.44	0.29	0.04	0.95
22. Wuzhen	Mean	7.49	5.74	4.31	21.27	2.02	0.70	0.22	1.89
	SD	0.25	1.89	0.69	11.92	0.86	0.47	0.11	1.61

Beiguodaqiao, Zhendongdukou and Baidugang were found ranging from 3.17 to 4.06 mg/L. This suggests that the discharge of industrial and domestic wastewater induced serious organic pollution at these regions, since the lower DO was mainly caused by the decomposition of organic compounds (Boyle and Fraleigh 2003). The highest

DO concentration value was found at Hushanqiao (8.28 mg/L).

There is no distinct difference of COD_{Mn}, COD, and BOD among the sites. The COD_{Mn}, COD and BOD concentrations ranged from 3.58 to 12.00 mg/L, 12.77 to 57.04 mg/L, 1.73 to 12.50 mg/L, without large fluctuation at most sites except Changzhengqiao. The higher contents of COD_{Mn}, COD, and BOD appeared in the water samples collected from Changzhengqiao and Baidugang.

In contrast, the higher concentrations of TN, NH_4^+–N and TP were detected at most sites upstream of Lake Taihu, such as Baidugang, Jiuxian, Panjiaba and Changzhengqiao. This indicates that the nutrients coming from both agricultural activities and domestic wastewater extensively deteriorate the surface water quality. Though diffuse pollution contributed more than 60% nutrient (N, P) loading (Wang et al. 2004), high concentrations of nitrogen and phosphorus (e.g. NH_4^+–N) were found at many sites strongly affected by urbanization, which probably was the main reason for the hypertrophic state of the northern part of Lake Taihu (Chen et al. 2003). Previous studies also indicated that pollutants produced by industry were mainly from the east or southeast of the lake, such as Suzhou, Wuxi and Jiaxing, pollutants from agriculture (cropping, rice growing, etc.) accounted for 37% of COD, 49.5% of TN and 48% of TP produced in the basin, these non-point pollutant sources came mostly from the west (Huang et al. 2004; Qin et al. 2007).

Among all the monitoring sites, Changzhengqiao located in Wuxi city was found to have the worst water quality with most highest indicators (e.g. TP, TN, COD, BOD), which suggests that it is still urgent to control the point pollutions in the Taihu Basin. On the other hand, the water quality at Dagangqiao and Hushanqiao located near the hilly region were relatively better than the others.

Principal component analysis on variables of surface water quality

PCA was applied to identify characteristics of water quality variables at all studied sites, based on the normalized average values obtained from monitoring data surveyed in a monthly interval during the study period.

A planar plot of eight variables against their values is presented in Fig. 2. As shown in Fig. 2, the significant factor PC1, with an eigenvalue greater than unity extracted by PCA accounts for 73.72% of the total variance. The second factor (PC2) was almost equal to unity (0.99) which accounts for 12.46% of the total variance. It can therefore be clearly seen that the first factor captures all nutrient related variables, and has strong negative loadings on COD, COD_{Mn}, BOD, TP, NH_4^+–N and TN, and positive loading on DO. The second factor (PC2) accounts for
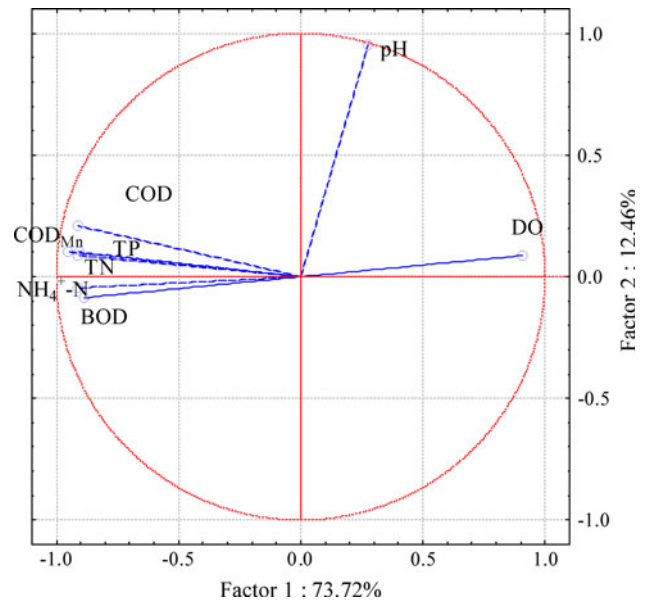

Fig. 2 Principal component analysis (PCA) ordination of the 22 sites by 8 environmental variables in the Taihu Basin

12.46% of the total variance and has strong positive loading on pH and moderate positive loading on DO. Using PCA, the eight original variables were reduced to two key uncorrelated factors. Each factor is significantly correlated to specific variables representing a different dimension of the water quality. The first factor represents the nutrient dimension, which is related to all nutrient variables. The high correlation of nutrient variables suggests that there is high consistency in their variations, which is in agreement with previous research conducted on Lake Taihu (Wang et al. 2007). Furthermore, this factor accounts for the majority of total variance, which can be concluded that overloading of nutrients is the major environmental problem to aquatic systems in the Taihu Basin (Luo and Pang 2005). The second factor is only strongly associated with pH, which may be influenced by other environmental variables (Boyle and Fraleigh 2003; Wang et al. 2007).

Clusters analysis on variables of water quality

A dendrogram of samples obtained by the Ward method using CA is shown in Fig. 3. Twenty-two sampling sites were divided into three groups. Group 1 consisted of sites 1, 3, 6, 7, 11–13 and 17–22. Group 2 consisted of sites 2, 4, 5, 8–10, 14 and 15, and group 3 only included site 16 (i.e. Changzhengqiao) which was detected as the heaviest polluted site. The CA results revealed that the similarities of the monitoring sites in each group are represented by the characteristics of the water quality variables.

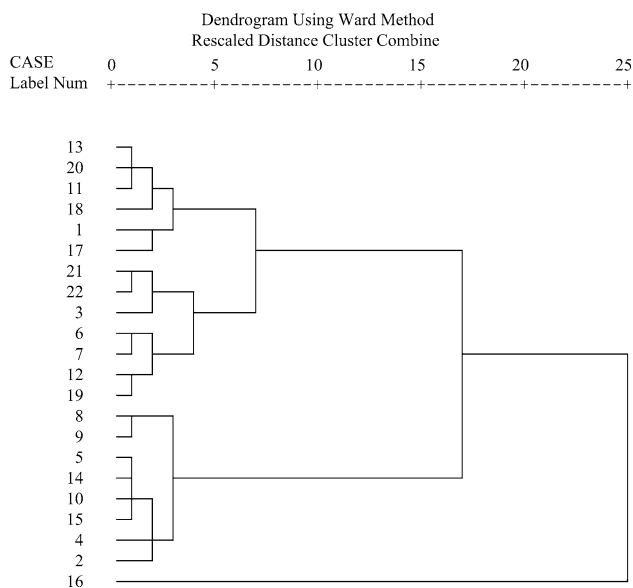As shown in Fig. 3, the first main group is formed from two subgroups that are linked at a rescaled distance of 7.

Fig. 3 Dendrogram showing sampling sites clusters in the Taihu Basin, China

Fig. 4 Scatter plot of the first two factor scores for the 22 sampling sites

The first subgroup includes sites 1, 11, 13, 17, 18, and 20 with higher concentrations of COD (>24 mg/L), NH_4^+–N, TP, and TN, which shows that surface water has become highly eutrophic and polluted. In the PCA method of classification these samples scored negative on factor 1 and close to 0 on factor 2 (see Fig. 4). The sites 1, 11, 13, and 20 show much higher concentration in NH_4^+–N (>2.2 mg/L), TP (>0.22 mg/L) and TN (>3.3 mg/L). This may have been caused by the large amounts of contaminants discharged into the surface water, produced in urban towns. Xie et al. (2007) investigated twelve most important rivers in Changshu city, Taihu region, and concluded that the highly NH_4^+–N pollutants mainly originated from point sources, such as domestic sewage, rural human and animal excreta, and industrial wastewater. The second subgroup consists of sites 3, 6, 7, 12, 19, 21, and 22 with COD in the range from 19 to 23 mg/L and relatively lower NH_4^+–N and TP, especially at sites 3, 21 and 22, the NH_4^+–N and TN contents are much lower, indicating that these sites are not intensively influenced by point source pollution.

In contrast, the sites in group 2, with infield and farmland around, were relatively moderately polluted. The contents of COD vary from 12.77 to 17.38 mg/L, and the NH_4^+–N values are much lower than those in group 1. In the PCA analysis, these samples were mostly grouped as positive values on factor 1 (Fig. 4). Most sites in this group indicate that the agriculture and households were dominant contributors of the nutrients flowing into the lake. This can be explained by observing the fact that a large portion of discharges from villages and livestock farms (especially those outside the core cities) are not appropriately treated, and effective technical measures to control agricultural
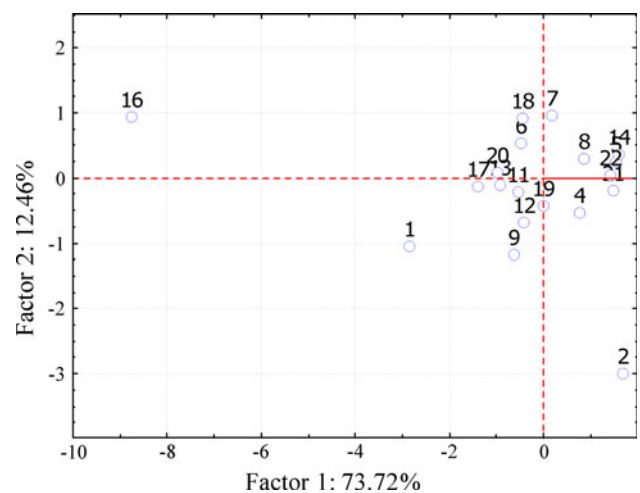
sources of pollution are not readily available in China at present (Wang et al. 2006). Intensive agricultural activities resulted in higher concentration of TN coming from fertilizers. This was the major reason of water eutrophication.

Group 3 corresponded to the highest polluted site which was only located at site 16 (i.e. Changzhengqiao). From Fig. 4, it also can be clearly seen that site 16 presented the worst quality due to the lowest negative value on factor 1. The site Changzhengqiao located in the north of Wuxi, received pollutants from point and non-point sources. Large amounts of urban wastewater, industry, and domestic wastewaters were discharged into the river directly, which brought site 16 to the super-polluted status.

Fuzzy synthetic evaluation of surface water quality in the Taihu Basin

Fuzzy synthetic evaluation with the entropy method for weight determination was used to assess the surface water quality in the Taihu Basin during the studied period. Figure 5 illustrated intra-annual variability of surface water quality in the Taihu Basin between 2001 and 2002.

As shown in Fig. 5a, most of the samples in the west of Lake Taihu belong to class III water quality in both wet and dry season. Some sites, such as Wangjiaqiao, Jiuxian, Lianjiangqiao, and Panjiaba displayed class V water quality during the dry season, because most of them are located in the agricultural region and received nutrients from farmland. These sites showed obvious seasonal variances due to agricultural activities. As mentioned above, Changzhengqiao was the heaviest polluted site due to the industrial pollution and domestic wastewater.

There are only a few sites with relatively good water quality belonging to class II through the year, i.e. Hushanqiao, Tanjingcun, Jiangbianzha and Dagangqiao. Two

Fig. 5 Intra-annual variability of surface water quality in 2001 (**a**) and 2002 (**b**) in the Taihu Basin based on fuzzy synthetic analysis. The *red* and *blue* dots represent the monthly water quality in the catchment. The position of the dot relative to the centre of the diagram expresses time (clockwise from January to December). The ranking of the surface water quality is expressed by the distance from the centre of the diagram—the maximum is depicted at the external *circle*. The position of the monitoring stations, are indicated by the *small crosses* (+)
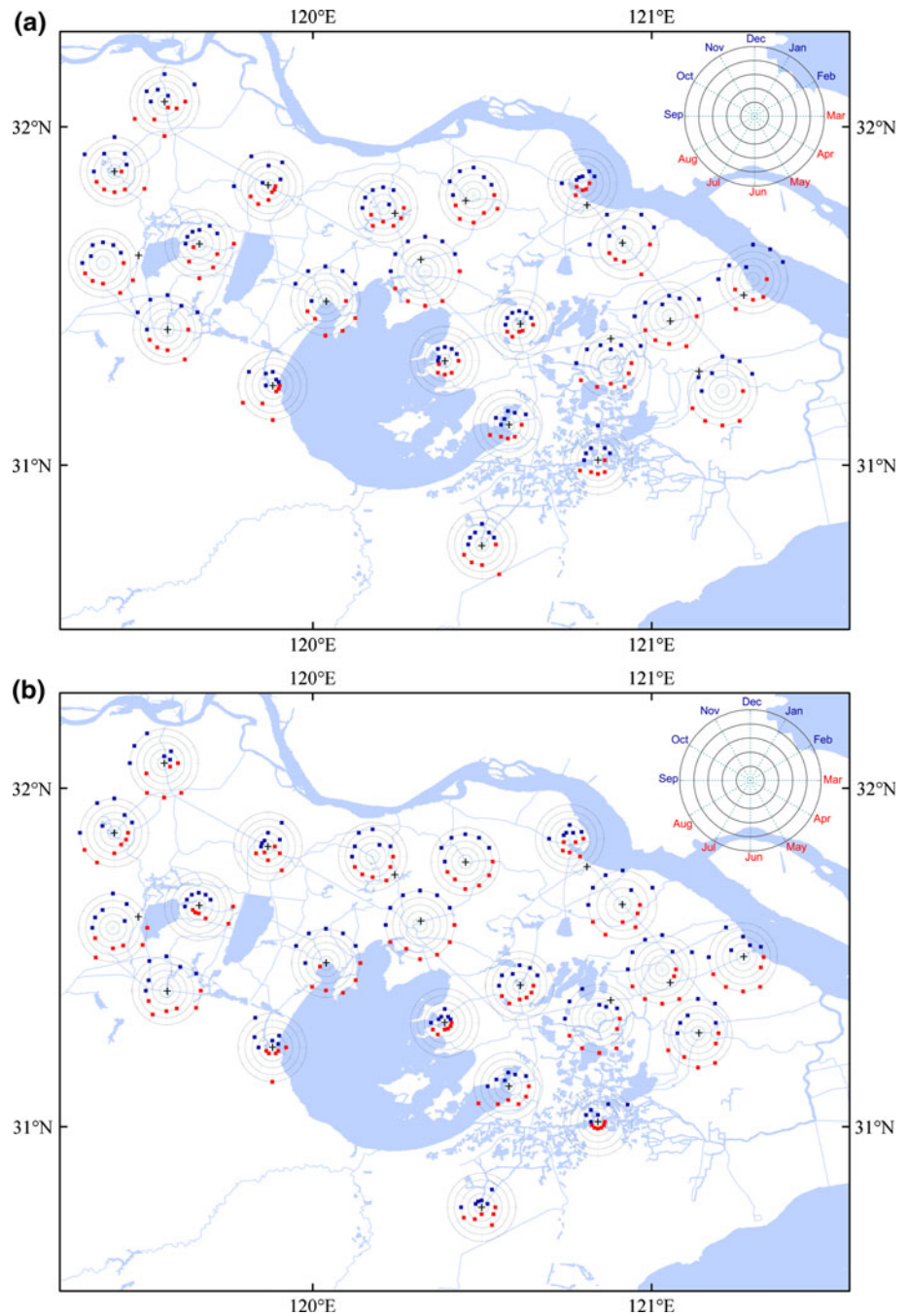
sites: Tanjingcun and Hushanqiao, located downstream of Lake Taihu were not heavily polluted by the wastewater or industry. The sites Jiangbianzha connected with the Yangtze River may be diluted by the water interaction.

The FSE indicates that water quality at most of the sampling sites vary from class III to V. Figure 5b shows the assessment results in 2002 which are very similar to that in 2001. The water quality at Zhushagangkou and Luxudaqiao located in the southeast of Lake Taihu is better than the northern sites. This is attributed to the water retention time of Lake Taihu. The average retention time of

Lake Taihu is about 5 months, but it is shorter in the south, since most runoff water is discharged via the Taipu River, in the southeast. Water quality, therefore, is better in the south than in the north (Qin et al. 2007).

Conclusions

Water quality analysis of the samples from the Taihu Basin indicates that pH values are in the range of natural surface water quality. The concentration of COD_{Mn}, BOD, COD,

TP, NH_4^+–N, and TN at most monitoring sites in the Taihu Basin indicates serious pollution. Spatial distributions of COD, NH_4^+–N, TP and TN concentrations show that higher values appeared in water samples near large cities.

1. The results from PCA suggested that nutrient pollution, organic pollution, and agricultural runoff were potential pollution sources. The extracted significant factor (PC1), explaining 73.72% of the total variance, is correlated with DO, BOD, COD, COD_{Mn}, NH_4^+–N, TP, and TN. The second factor accounting for 12.47% of the total variance is associated with pH.

2. The results of CA are in agreement with that from PCA. The CA divided 22 sampling sites into three groups: site 16 (Changzhengqiao) was the most polluted sample, and sites 1, 3, 6, 7, 11–13, 17–22 are moderately polluted points, the water quality of remaining sites is comparably better.

3. The FSE showed that there is no significant variance of water quality between 2001 and 2002 among all the sites. Seasonal changes could be detected at some sites in the west of Lake Taihu within the agricultural region. Changzhengqiao was found to be the heaviest polluted site due to large amounts of industrial pollutants, and domestic wastewater discharged into the rivers directly. The evaluation results indicate that water quality at this site belongs to class V during the whole period. The water quality at four sites (i.e. Hushanqiao, Tanqingcun, Jiangbianzha and Dagangqiao) showed good water quality ranging from class II to class IV.

In this study both the multivariate statistical analysis (PCA and CA) and FSE method were applied to assess the water quality in the Taihu Basin. The water quality assessment results indicate that the variables responsible for water quality variations are mainly related to nutrients (non-point sources: agricultural activities) in the west of Lake Taihu and organic pollution (point source: urbanization, industry and domestic wastewater) in the east area of the Taihu Basin. This study demonstrates the utility of multivariate statistical techniques and FSE method for analysis and interpretation of complex data set in environmental science, identification of pollution sources, and understanding spatial and temporal variations in water quality, which may provide a useful tool for water quality management, as well as guide the wastewater treatment and management to some extent. However, an updated long-term continuous data set would be necessary to improve the results.

References

Bowen RE, Depledge MH (2006) Rapid assessment of marine pollution (RAMP). Mar Pollut Bull 53:631–639. doi:10.1016/j.marpolbul.2006.09.002

Boyle TP, Fraleigh HD (2003) Natural and anthropogenic factors affecting the structure of the benthic macroinvertebrate community in an effluent-dominated reach of the Santa Cruz River, AZ. Ecol Indic 3:93–117. doi:10.1016/S1470-160X(03)00014-1

Bu HM, Tan X, Li SY, Zhang QF (2009) Water quality assessment of the Jinshui River (China) using multivariate statistical techniques. Environ Earth Sci. doi:10.1007/s12665-009-0297-9

Cattell RB, Jaspers J (1967) A general plasmode No. 20-10-5-2 for factor analytic exercises and research. Multivariate Behav Res Monogr 2(3):1–212

Chen YW, Fan CX, Teubner K, Dokullil M (2003) Changes of nutrients and phytoplankton chlorophyll-a in a large shallow lake, Taihu, China: an 8-year investigation. Hydrobiologia 506:273–279. doi:10.1023/B:HYDR.0000008604.09751.01

Chinese Environmental Protection Agency (2002) National surface water environmental quality standards of China (GB3838-2002). China Standards Press, Beijing (in Chinese)

Cude CG (2001) Oregon water quality index: a tool for evaluating water quality management effectiveness. J Am Water Resour Assoc 37:125–137. doi:10.1111/j.1752-1688.2001.tb05480.x

Dahiya S, Singh B, Gaur S, Garg VK, Kushwaha HS (2007) Analysis of groundwater quality using fuzzy synthetic evaluation. J Hazard Mater 147:938–946. doi:10.1016/j.jhazmat.2007.01.119

Emery WJ, Thomson RE (1997) Data analysis methods in physical oceanography. Pergamon Press, New York, p 634

Gao C, Zhu JG, Zhu JY, Gao X, Dou YJ, Hosen Y (2004) Nitrogen export from an agriculture watershed in the Taihu Lake area, China. Environ Geochem Health 26:199–207. doi:10.1023/B:EGAH.0000039582.68882.7f

Guo HY, Wang XR, Zhu JG (2004) Quantification and index of non-point source pollution in Lake Taihu region with GIS. Environ Geochem Health 26:147–156. doi:10.1023/B:EGAH.0000039577.67508.76

Huang WY, Xu G, Wu QL, Qin BQ (2004) The pollutant budget in Lake Taihu. In: Qin BQ, Hu WP, Chen WM (eds) Process and mechanism of environmental changes of Lake Taihu. Science Press, Beijing, pp 21–28 (in Chinese)

Icaga Y (2007) Fuzzy evaluation of water quality classification. Ecol Indic 7:710–718. doi:10.1016/j.ecolind.2006.08.002

Jolliffe IT (2002) Principal component analysis, 2nd edn. Springer, New York, p 478

Lu XW, Li LY, Lei K, Wang LJ, Zhai YX, Zhai M (2009) Water quality assessment of WeiRiver, China using fuzzy synthetic evaluation. Environ Earth Sci. doi:10.1007/s12665-009-0303-2

Luo J, Pang Y (2005) Study on flux of pollutants discharged into Taihu Lake through main inflow river channels. Hohai Univ 33:131–135

Maria JB, Graca C (2006) Identification of similar environmental areas in Tagus Estuary by using multivariate analysis. Ecol Indic 6:508–515. doi:10.1016/j.ecolind.2005.07.001

Milovanovic M (2007) Water quality assessment and determination of pollution sources along the Axios/Vardar River, Southeastern Europe. Desalination 213:159–173. doi:10.1016/j.desal.2006.06.022

Qin BQ, Xu PZ, Wu QL, Luo LC, Zhang YL (2007) Environmental issues of Lake Taihu, China. Hydrobiologia 581:3–14. doi:10.1007/s10750-006-0521-5

Ross TJ (2004) Fuzzy logic with engineering applications. Wiley, New York

Shen G, Lu Y, Wang M, Sun Y (2005) Status and fuzzy comprehensive assessment of combined heavy metal and organochlorine pesticide pollution in the Taihu Lake region of China. J Environ Manage 76:355–362. doi:10.1016/j.jenvman.2005.02.011

Shrestha S, Kazama F (2007) Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. Environ Modell Softw 22:464–475. doi:10.1016/j.envsoft.2006.02.001

Singh KP, 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 Res 38:3980. doi:10.1016/j.watres.2004.06.011

Vega M, Pardo R, Barrado E, Deban L (1998) Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Res 32:3581–3592. doi:10.1016/S0043-1354(98)00138-9

Wall ME, Rechtsteiner A, Rocha LM (2003) Singular value decomposition and principal component analysis. In: Berrar DP, Dubitzky W, Granzow M (eds) A practical approach to microarray data analysis. Kluwer, Norwell, pp 91–109

Wang XJ, Zhang W, Huang YN, Li SJ (2004) Modeling and simulation of point-non-point source effluent trading in Taihu Lake area: perspective of non-point sources control in China. Sci Total Environ 325:39–50. doi:10.1016/j.scitotenv.2004.01.001

Wang QG, Gu G, Higano Y (2006) Toward intergraded environmental management for challenges in water environmental protection of Lake Taihu Basin in China. Environ Manage 37:579–588. doi:10.1007/s00267-004-0347-8

Wang XL, Lu YL, Han JY, He GZ, Wang TY (2007) Identification of anthropogenic influences on water quality of rivers in Taihu watershed. J Environ Sci 19:475–481. doi:10.1016/S1001-0742(07)60080-1

Ward JH (1963) Hierarchical Grouping to optimize an objective function. J Am Stat Assoc 58:236–244

Wunderlin DA, Diaz MDP, Ame MV, Pesce SF, Hued AC, Bistoni MD (2001) Pattern recognition techniques for the evaluation of spatial and temporal variation in water quality. A case study: Suquia river basin (Cordoba Argentina). Water Res 35:2881–2894. doi:10.1016/S0043-1354(00)00592-3

Xie YX, Xiong ZQ, Xing GX, Sun GQ, Zhu ZL (2007) Assessment of nitrogen pollutant sources in surface waters of Taihu Lake region. Pedosphere 17(2):200–208. doi:10.1016/S1002-0160(07)60026-5

Xu PZ, Qin BQ (2005) Water quantity and pollutant fluxes of the surrounding rivers of Lake Taihu during the hydrological year of 2001–2002. J Lake Sci 17(3):213–218 (in Chinese)

Xu H, Yang LZ, Zhao GM, Jiao JG, Yin SX, Liu ZP (2009) Anthropogenic impact on surface water quality in Taihu Lake Region, China. Pedosphere 19:765–778. doi:10.1016/S1002-0160(09)60172-7

Yen J, Langari R (1999) Fuzzy logic: intelligence, control and information. Prentice-Hall, New Jersey

Yidana SM, Ophori D, Banoeng-Yakubo B (2008) A multivariate statistical analysis of surface water chemistry data—the Ankobra Basin, Ghana. J Environ Manage 86:80–87. doi:10.1016/j.jenvman.2006.11.023

Zhang HC, Cao ZH, Wang GP, Zhang HA, Wong MH (2003) Winter runoff losses of phosphorus from paddy soils in the Taihu Lake Region of South China. Chemosphere 52:1461–1466. doi:10.1016/S0045-6535(03)00483-1

Zou ZH, Yun Y, Sun JN (2006) Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment. J Environ Sci 18:1020–1023. doi:10.1016/S1001-0742(06)60032-6