Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Environmental Sciences 13 (2012) 1213 – 1226
Procedia Environmental Sciences

The 18th Biennial Conference of International Society for Ecological Modelling

Assessment of water quality in Baiyangdian Lake using multivariate statistical techniques

Y.Zhao , X.H. Xia, Z.F. Yang*, F. Wang

State Key Laboratory of Water Environment Simulation, School of Environment, Beijing Normal University,Beijing 100875, China

Abstract

Water quality can be considered a key contributor to both health and disease for humans. This study involved the evaluation and interpretation of complex water quality data and the sources of pollution in Baiyangdian Lake (China). It also allowed us to obtain more advanced information about water quality, and to design a monitoring network for this study area. Multivariate statistical techniques, including principal component analysis (PCA) and hierarchical cluster analysis (CA) were applied to evaluate water quality of the lake. The 21 physicochemical parameters focused on in the study were analyzed in water samples collected monthly over a two-year period from 13 different sites located in and around the lake. Exploratory analysis of experimental data involved use of PCA and CA in an attempt to discriminate sources of variation measured in the samples. PCA was used to identify a reduced number of five principle components, demonstrating up to 92% of both temporal and spatial changes. CA classified similar water quality stations into 5 clusters based on the PCA scores. The results showed that cluster 5 (site 2) was characterized as the most heavily polluted site, a result that can be attributed to the pollution from the nearby Fuhe River (a upstream river that receiving almost all of the domestic sewage and some industrial wastewater from Baoding City). Cluster 1 (sites 3, 4, 5, 6 and 7) and cluster 4 (site 1) were identified as moderately polluted in association with the both domestic and agricultural sewage, as well as fishery-related pollution in the lake. Cluster 2 (sites 11, 12 and 13) and cluster 3 (sites 8, 9 and 10) were less polluted, which suggests that the water quality was better in the eastern and central portions of the lake.

Keywords: Water quality; Physical-chemical parameters; Principle component analysis; Cluster analysis; Baiyangdian Lake

* Corresponding author. Tel.: +86 10 58807596; fax: +86 10 58807596.
E-mail address:zfyang@bnu.edu.cn.

 doi:10.1016/j.proenv.2012.01.115

1. Introduction

Water quality is considered to be a key contributor to both health and the state of disease for humans. Surface water quality in a region is largely influenced by both natural processes and by anthropogenic inputs [1]. Niemi et al. [2] reported that human activities mainly impact surface water quality through atmospheric pollution, effluent discharges, the use of agricultural chemicals, in addition to the increased exploitation of water resources. This has generated great pressure on aquatic ecosystems, resulting in a decrease of water quality and biodiversity, loss of critical habitats, and an overall decrease in quality of life for local inhabitants [3]. It is therefore essential to prevent and control water pollution and to implement regular monitoring programs.

Conventional water quality regulations contain quality classes based on crisp sets, with the limits between different classes having inherent imprecision [4]. A parameter being close to or far from the limit has equal importance in evaluating the concentration of the parameter. Besides that, not all of the water quality parameters may be included in a single class. Having various quality classes in one sampling location may cause confusion in the quality definition of the sampling site [5]. In the recent years, the application of different multivariate statistical techniques, such as cluster analysis (CA), principal component analysis (PCA), factor analysis (FA) and discriminate analysis (DA) has helped to identify possible sources that influence water systems, and have offered a valuable tool in the reliable management of water resources [6,7]. These techniques have been effectively employed to classify water quality data and detect similarities among samples and/or variables in many research studies [8-16].

Baiyangdian Lake (115º45′–116º07′E, 38º43′–39º02′N) is the largest shallow freshwater lake in northern China (Fig.1), and plays an important role in sustaining agriculture, climate regulation, and flood control. It consists of 143 small and shallow lakes linked by thousands of ditches, with a mean surface area of 366km^2 and a water depth of less than 2 meters [17]. There are 39 types of aquatic plants in the lake, of which *Ceratophyllum demersum* and *Phragmites australis* are the dominant submerged and emergent plants, respectively. As a famous tourist resort in China, Baiyangdian Lake receives more than 850,000 tourists every year. However, the water quality of the lake has continuously worsened as a result of increasing industrial activities and land use [18]. In addition, persistent water withdrawals associated with irrigation needs and periods of severe drought have contributed to this decline in water quantity. Because serious water pollution is a threat to human health, periodical monitoring and assessment of the water quality of Baiyangdian Lake is imperative. One particular problem with water quality monitoring is the complexity associated with the analysis of a large number of variables, each of which contains rich information on the characteristics of the water resources at hand [19]. It can be more easily and accurately handled through the application of multivariate methods and exploratory data analysis [20].

In this study, PCA and CA have been used to evaluate data and draw conclusions about the similarities and dissimilarities existing among the various water quality parameters, as well as to identify variables specific to studying spatial dissimilarity in Baiyangdian Lake. Finally, the influence of the pollution sources on the water quality parameters was ascertained.

2. Material and methods

2.1 Sample collection and analytical procedures

Water samples from 13 sites were collected at 0.5m (a depth representative of the mixed water columns) on the basis of bimonthly intervals between January 2007 and December 2008 in Baiyangdian Lake. The samples were kept in 2 L polyethylene plastic bottles which had been previously cleaned with metal free soap, rinsed repeatedly with distilled water, soaked in 10% nitric acid for 24 h, and finally rinsed with ultrapure water. All water samples were maintained at 4°C first during transportation to the laboratory, and then later for processing and analysis.

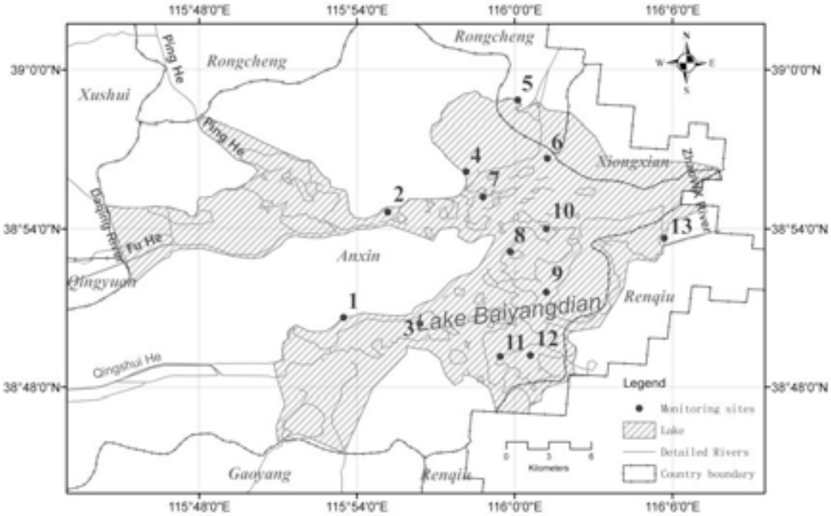

Fig. 1. Location and sampling sites in Baiyangdian Lake

Although 23 hydrochemical and physical variables were monitored, the concentrations of Cr and Pb were reported at less than detection levels (0.004 mg/L for Cr and 0.01 mg/L for Pb). As such, 21 variables were actually analyzed and reported on in this study. As shown in Table 1, the measurements of water quality parameters are summarized on the basis of standard methods established for surface water monitoring in China [21]. The temperature, pH, electrical conductivity (EC), and DO of each water sample were measured in situ by field instruments including a mercury thermometer, digital pH, EC and DO meter, respectively. All water samples were analyzed for different physico-chemical parameters within 48 h of collection, with chemical oxygen demand (COD_{Mn}) determined on the same sampling day while biological oxygen demand (BOD) is also determined promptly to avoid time-induced changes in bacterial concentration. TSS was determined gravimetrically at a temperature of 105-110 °C. Total hardness was measured by EDTA complexometry titration. Total alkalinity was determined by acid titration using methyl-orange as endpoint, and chloride by silver nitrate ($AgNO_3$) titration using potassium chromate (K_2CrO_4) solution as an indicator. The BOD was determined by the dilution and seeding method, while COD_{Mn} was determined by the acid titration method. SO_4 was determined spectrophotometrically by the barium sulfate turbidity method, and NH_4-N was measured with Nessler's reagent. NO_3-N and NO_2-N were analyzed by phenol disulfonic acid colorimetry and N-(1-naphthyl)-ethylenediamine colorimetry, respectively. TN and TP were analyzed by absorption spectrophotometry after decomposition with potassium peroxodisulfate ($K_2S_2O_8$).

The acid-treated water samples were analyzed for the determination of major cations, Ca, Na, and K were measured by flame photometry, while Mg was determined by the flame atomic absorption spectrometer (FAAS). For trace and toxic elements, the volume of water samples was reduced by heating at 60 ℃ on an electric hot plate. As was determined using hydride generation atomic absorption spectrometer method (HGAAS). Pb was analyzed by using the electrothermal atomic absorption spectrometer (ETAAS), and Cr by using the dinitrodiphenyl carbazide spectrophotometric method.

Table 1 Water quality parameters and their associated abbreviations, units, and analytical methods used

Variables	Abbreviations	Units	Analytical methods
Water temperature	T	℃	Thermometer
pH	pH	pH unit	pH-meter
Electrical conductivity	EC	mS/m	Electrometric
Total suspended solids	TSS	mg/L	Gravimetric
Calcium	Ca	mg/L	Flame photometer
Magnesium	Mg	mg/L	FAAS
Potassium	K	mg/L	Flame photometer
Sodium	Na	mg/L	Flame photometer
Chloride	Cl	mg/L	Titrimetric
Sulphate	SO_4	mg/L	Spectrophotometric
Total hardness	T-Hard	mg/L	Titrimetric
Total alkalinity	T-Alk	mg/L	Titrimetric
Dissolved oxygen	DO	mg/L	Probe method
Ammoniacal nitrogen	NH_4-N	mg/L	Spectrophotometric
Nitrite nitrogen	NO_2-N	mg/L	Spectrophotometric
Nitrate nitrogen	NO_3-N	mg/L	Spectrophotometric
Permanganate index	COD_{Mn}	mg/L	Acid titration
Biochemical oxygen demand	BOD	mg/L	Dilution and seeding
Arsenic	As	ug/L	Hydride generation AAS
Chromium	Cr	mg/L	Spectrophotometric
Lead	Pb	mg/L	ETAAS
Total phosphorus	TP	mg/L	Spectrophotometric
Total nitrogen	TN	mg/L	Spectrophotometric

2.2 Multivariate statistical analysis

All mathematical and statistical computations were made using SPSS 13.0 (SPSS Inc., Chicago IL., USA). Multivariate analysis of the lake water quality data set was performed using principal component and cluster analysis techniques.

- Principal components analysis (PCA)

PCA is a pattern recognition technique that attempts to interpret the variance within a large set of intercorrelated variables by converting them into a smaller set of independent variables [7]. It provides information on the most significant parameters used to describe the entire data set, data reduction, and to summarize the statistical correlation among constituents in the water with a minimum loss of original information [22, 23]. PCA has been used on a correlation matrix of rearranged data to explain the structure of the underlying dataset and to identify the unobservable, latent pollution sources.

- Cluster analysis (CA)

Cluster analysis is defined as the classification of similar objects into groups where the number of groups as well as their forms are unknown [24], with the primary purpose being the assembly of objects based on the characteristics they possess. Hierarchical agglomerative clustering is the most common approach, which provides instinctive similarity relationships between any one sample and the entire data set, is typically illustrated by a dendrogram (tree diagram). The dendrogram presents a picture of the groups and their proximity to one another, with a dramatic reduction in the dimensionality of the original data [22]. In this study, hierarchical agglomerative CA was carried out on the normalized data by means of Ward's method, using squared Euclidean distances as a measure of similarity [25].

In this research, PCA was applied to summarize the statistical correlation among water quality parameters. The concentrations of physico-chemical parameters tend to differ greatly; as such, the statistical results should be highly biased by any parameter having a high concentration. Thus, each water quality parameter was standardized (z-scale) before PCA the analysis was performed in order to minimize the influence of different variables and their respective units of measurements. The calculations were performed based on the correlation matrix of chemical components, and the PCA scores were obtained from the standardized analytical data. CA was used to detect spatial similarity for grouping sampling sites located within the monitoring network.

2.3 Comprehensive evaluation of water quality in the lake

A comprehensive pollution index method has been applied to evaluate water quality qualitatively in many existing studies [26, 27, 28]. The detail of the comprehensive pollution index is below:

$$P = \frac{1}{n}\sum_{i=1}^{n} C_i / S_i \qquad (1)$$

where *P* is comprehensive pollution index, C_i is the measured concentration of the pollutant (mg/L), S_i represents the limits allowed by the State Environmental Protection Administration (SEPA) of China for water quality, and n is the number of selected pollutants. Ultimately, the values determined for *P* could be used to classify the water quality level of the lake (Table 2).

Table 2 Standard of surface water quality classification

Comprehensive pollution index (*P*)	Water quality level
≤0.20	Ⅰ Cleanness
0.21-0.40	Ⅱ Sub-cleanness
0.41-1.00	Ⅲ Slight pollution
1.01-2.0	Ⅳ Moderate pollution
≥2.01	Ⅴ Severe pollution

3. Results and discussion

3.1 The Physical-chemical characteristics of water in Baiyangdian Lake

The basic statistics calculated for the lake's water quality are summarized in Table 3, which presents the range, mean, and standard deviation of the results for each of the 21 parameters.

Water temperatures showed a characteristic annual cycle, with higher values during the summer (29.0-33.2℃) and lower values in the winter season (0.5-2℃). The pH values of collected water samples ranged from 7.6-8.6, within the limit range of 6-9 allowed by the SEPA for water quality [21]. The TDS and EC during the annual season cycle showed significant variations, with values of 1-230 mg/L and ranging from

764-2030 mS/m, respectively. The major cations (Na, Ca, Mg, and K) were found to be present in lower concentrations than levels measured in several other surficial water bodies throughout the world [1, 29, 30]. The anions Cl and SO_4 were measured in lower concentrations than SEPA limits, ranging from 108.15-152.23 mg/L and 140.1-188.7 mg/L, respectively.

It was noted that the average nutrient concentrations (TN, TP and NH_4^+) were measured in higher concentrations than the guide levels, with the values measured at 1.18-19.03mg/L, 0.07-1.88 mg/L and 0.36-15.72mg/L, respectively. The water nutrient contamination mainly resulted from domestic and industrial sewage introduced by the upstream Fuhe River, which is considered to be a major pollution source for the lake based on its receipt of agricultural runoff, and domestic and industrial sewage. Locally, the impacts of domestic wastes and animal husbandry also represent pollution sources for the lake. These results are consistent with a study conducted in Lake Pandu Bodhan, where high nutrient concentrations were documented due to anthropogenic activities such as organic pollutants releases and the discharge of domestic sewage [31].

The monitoring of oxygen concentrations in aquatic system is an important aspect of this study [32], as the biological, chemical, and physical processes in the lake are numerous and complex; as such, that there is no model can be used to study these concentrations without a careful analysis of local characteristics. COD_{Mn} is widely used to determine waste concentrations, and is primarily applied to pollutant mixtures such as domestic sewage, agricultural, and industrial waste. The higher concentrations of BOD are attributed to local anthropogenic pollution associated with fisheries and domestic wastes. In Baiyangdian Lake, the highest values of COD_{Mn} and BOD were recorded at sites 1 and 2 (13.48 mg/L for COD_{Mn} and 21.38 mg/L for BOD), while the lowest values were measured at site 10 (7.50mg/L for COD_{Mn} and 3.04 mg/L for BOD). Although high levels of COD_{Mn} and BOD were observed in all study sites, DO values were still found to be within the permissible limits of Chinese water quality standards. These results differ from other water bodies characterized by higher COD_{Mn} and BOD concentrations and lower DO values [1, 31].

The dissolved metal concentrations measured in Baiyangdian Lake were minor. The concentrations of Cr and Pb were measured below the detected limits. Although average As concentrations were also below the limits, they were close to the permissible limit of WHO (10ug/L) for sampling sites located at the confluence of Fuhe River and the lake [33]. Thus, the industrial and domestic sewage from Fuhe River were determined to be major pollution source.

Table 3 Range, mean and standard deviation of water quality parameters at sampling locations in Baiyangdian Lake during 2007-08

Parameters		Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7
TEMP	R	1.2-29.0	0-33.2	1.8-30.5	2.0-31.0	1.6-29.5	1.5-30.6	1.9-29.8
(℃)	M±Sd.	15.14±9.96	14.63±11.24	14.88±10.83	16.04±10.08	15.72±10.08	16.24±10.25	16.16±10.05
pH	R	7.6-8.2	7.7-8.9	7.9-8.6	8.0-8.5	7.8-8.4	7.8-8.4	7.9-8.3
	M±Sd.	8.00±0.18	8.03±0.33	8.09±0.19	8.11±0.16	8.09±0.19	8.08±0.18	8.07±0.13
EC	R	864-1069	902.0-2030.0	867.0-1091.0	764.0-1279.0	867.0-1101.0	877.0-1102.0	852.0-1168.0
(mS/m)	M±Sd.	977.33±80.05	1341.75±312.9	972.18±73.01	1048.90±164.58	1009.89±84.61	1004.20±82.12	1010.70±103.61
TSS	R	4.0-25.0	7.0-179.0	1.0-44.0	9.0-79.0	1.0-69.0	1.0-37.0	4.0-70.0
(mg/L)	M±Sd.	10.00±7.55	43.00±35.69	15.09±12.45	28.20±25.14	19.78±15.7	15.40±10.70	26.00±18.90
Ca	R	41.3-75.1	48.7-89.7	44.0-68.8	47.1-72.5	14.0-64.6	13.2-65.7	40.1-66.4
(mg/L)	M±Sd.	59.94±12.19	74.75±15.22	59.66±11.70	60.57±9.92	54.98±20.12	55.13±20.61	56.47±12.25
Mg	R	21.9-44.0	27.9-42.5	33.1-43.9	26.6-45.0	28.4-62.6	27.0-59.6	27.6-51.9
(mg/L)	M±Sd.	32.64±8.24	34.45±6.09	36.92±4.24	33.45±6.82	41.17±11.70	38.17±11.69	35.98±9.04
K	R	4.8-8.8	8.5-14.1	5.2-9.9	6.2-12.3	5.5-11.6	5.2-10.6	5.3-12.3
(mg/L)	M±Sd.	6.97±1.66	12.27±2.22	7.37±1.87	9.27±2.07	8.07±2.01	8.27±1.85	9.54±2.60
Na	R	84.4-116.0	80.0-173.0	84.5-98.1	78.4-129.0	82.0-117.0	75.8-115.0	79.4-122.0
(mg/L)	M±Sd.	98.32±11.72	133.00±33.22	93.26±5.21	102.85±17.45	100.65±13.34	97.82±13.49	102.50±14.89
Cl	R	90.2-123.0	99.8-263.0	94.6-122.0	74.9-143.0	89.6-130.0	94.4-129.0	87.6-129.0
(mg/L)	M±Sd.	110.10±10.27	152.23±43.34	109.69±8.47	112.52±19.05	111.14±14.41	109.75±11.91	110.29±14.12
SO_4	R	105.0-117.0	103.0-256.0	119.0-184.0	74.9-195.0	83.5-192.0	79.0-173.0	64.4-190.0
(mg/L)	M±Sd.	140.11±22.62	171.25±40.14	151.90±16.89	149.10±41.72	143.50±36.29	143.70±30.55	141.54±41.01
T-Hard	R	150.0-187.0	133.0-218.0	153.0-199.0	127.0-192.0	155.0-185.0	154.0-192.0	133.0-183.0
(mg/L)	M±Sd.	164.56±10.98	186.33±24.32	169.18±13.29	169.10±18.81	173.00±10.23	171.20±12.14	167.70±15.04
T-Alk	R	102.0-121.0	136.0-245.0	96.2-126.0	112.0-149.0	117.0-145.0	111.0-142.0	112.0-144.0
(mg/L)	M±Sd.	113.20±7.98	182.17±39.20	115.04±12.73	130.00±15.80	130.50±11.50	128.00±11.76	129.00±14.00
DO	R	2.0-10.6	3.2-17.9	3.4-15.8	2.6-17.0	3.1-16.7	3.1-22.2	2.0-18.5
(mg/L)	M±Sd.	5.99±4.40	7.97±5.17	9.92±3.43	9.38±4.19	9.23±4.14	10.42±5.25	10.06±5.46
NH_4-N	R	0.2-0.7	0.3-43.7	0.25-0.58	0.4-9.6	0.2-4.1	0.3-3.2	0.2-10.4
(mg/L)	M±Sd.	0.44±0.14	15.71±12.86	0.37±0.13	2.97±1.66	0.87±0.20	1.05±0.98	2.35±1.54
NO_2-N	R	0.003-0.02	0.005-0.6	0.004-0.034	0.003-0.104	0.003-0.078	0.003-0.103	0.003-0.109
(mg/L)	M±Sd.	0.004±0.002	0.157±0.09	0.010±0.008	0.033±0.028	0.012±0.005	0.021±0.005	0.027±0.015
NO_3-N	R	0.1-1.3	0.11-12.8	0.2-2.6	0.2-2.4	0.19-1.8	0.11-1.75	0.1-2.3
(mg/L)	M±Sd.	0.61±0.41	2.97±2.0	0.7±50.21	0.97±0.62	0.63±0.57	0.74±0.57	0.94±0.73
CODMn	R	6.0-16.7	7.8-23.6	5.1-17.8	5.8-21.2	6.6-12.1	5.2-15.8	4.4-15.3
(mg/L)	M±Sd.	9.94±3.74	13.48±5.32	9.16±3.83	11.17±5.21	9.41±2.08	9.79±2.95	9.27±2.88
BOD	R	4.1-73.6	3.6-31.4	4.3-15.3	2.1-14.3	2.0-15.6	2.1-21.5	2.0-14.2
(mg/L)	M±Sd.	21.38±12.22	16.33±10.12	6.38±3.70	7.26±3.58	5.62±2.01	6.89±2.01	7.52±4.77
As	R	1.9-8.1	1.6-40.1	1.0-9.0	1.7-8.3	1.3-5.9	1.2-8.1	1.3-11.0
(ug/L)	M±Sd.	3.55±2.60	8.15±2.11	3.10±2.40	4.60±2.30	4.10±1.64	4.62±2.61	5.09±3.07
TP	R	0.05-0.67	0.16-3.01	0.04-2.89	0.07-0.96	0.04-0.35	0.04-0.52	0.11-0.88
(mg/L)	M±Sd	0.251±0.225	1.879±0.840	0.420±0.221	.440±0.361	0.214±0.111	0.283±0.178	0.415±0.279
TN	R	0.8-1.9	1.7-45.3	0.9-4.0	1.0-11.3	0.9-6.2	0.9-6.9	1.2-9.0
(mg/L)	M±Sd.	1.18±0.35	19.03±15.74	1.58±0.83	4.61±3.88	1.83±1.66	2.28±1.83	3.54±2.02

R: Range; M: Mean; Sd: Standard deviation

Table3 continued

Parameters		Site 8	Site 9	Site 10	Site 11	Site 12	Site 13
TEMP	R	2.0-32.0	1.3-31.5	1.8-31.0	1.5-30.5	1.0-31.0	0.5-30.0
(℃)	M±Sd.	16.98±10.04	16.76±10.29	16.60±10.33	15.87±9.89	15.88±10.22	16.10±10.29
pH	R	7.8-8.3	7.9-8.3	7.7-8.2	7.8-8.1	7.9-8.3	7.7-8.4
	M±Sd.	8.08±0.14	8.05±0.14	7.98±0.18	7.96±0.14	8.06±0.12	8.01±0.19
EC	R	933.0-1292.0	895.0-1125.0	925.0-1454.0	884.0-1067.0	884.0-1010.0	800.0-1155.0
(mS/m)	M±Sd.	1090.10±113.4	1017.60±73.23	1100.10±185.3	989.50±54.68	963.12±43.24	980.82±129.31
TSS	R	6.0-59.0	1.0-50.0	1.0-76.0	1.0-47.0	2.0-66.0	1.0-230.0
(mg/L)	M±Sd.	16.70±16.42	16.90±11.28	20.90±15.24	13.80±10.22	20.18±15.24	34.73±30.01
Ca	R	25.9-59.0	38.4-72.7	28.7-113.0	21.5-71.0	28.1-70.7	19.8-65.7
(mg/L)	M±Sd.	55.80±20.19	57.48±12.98	61.07±29.03	50.50±18.15	47.72±17.02	45.72±19.33
Mg	R	33.9-70.5	30.2-53.0	34.3-76.1	29.4-52.9	31.0-49.7	30.1-60.2
(mg/L)	M±Sd.	48.18±15.78	41.73±9.12	51.45±17.35	38.45±10.57	37.78±17.02	43.78±11.71
K	R	4.2-6.9	4.2-8.1	4.3-6.7	4.0-9.1	4.1-8.3	5.0-8.0
(mg/L)	M±Sd.	6.05±0.96	6.40±1.28	5.68±0.86	6.32±1.80	6.59±1.50	6.37±1.13
Na	R	94.0-123.0	90.0-111.0	88.4-125.0	82.8-111.0	85.0-120.0	47.6-113.0
(mg/L)	M±Sd.	106.03±10.18	100.60±7.75	105.73±13.90	99.98±10.00	103.18±13.00	91.87±24.14
Cl	R	98.4-121.0	96.3-120.0	96.4-123.0	103.0-128.0	105.0-128.0	65.4-139.0
(mg/L)	M±Sd.	112.14±6.79	111.33±6.74	112.24±7.63	117.30±9.11	115.45±8.38	108.15±21.15
SO4	R	116.0-252.0	116.0-196.0	105.0-355.0	106.0-180.0	106.0-192.0	111.0-190.0
(mg/L)	M±Sd.	179.10±50.45	157.30±27.74	188.70±81.81	140.11±26.89	145.55±25.61	155.64±29.79
T-Hard	R	158.0-256.0	149.0-216.0	155.0-337.0	137.0-180.0	125.0-183.0	138.0-215.0
(mg/L)	M±Sd.	189.13±6.77	176.00±25.52	205.60±67.31	159.80±14.35	156.73±18.46	167.27±24.96
T-Alk	R	103.0-164.0	101.0-143.0	104.0-184.0	104.0-121.0	91.0-114.0	83.2-135.0
(mg/L)	M±Sd.	129.17±24.40	117.00±17.62	135.17±29.69	113.33±7.03	102.45±8.26	109.53±17.42
DO	R	3.9-11.8	5.4-14.7	5.3-11.4	2.2-14.2	6.7-19.9	4.7-18.0
(mg/L)	M±Sd.	8.91±2.37	9.64±2.46	8.34±2.04	7.70±3.61	11.05±3.88	9.98±4.22
NH_4-N	R	0.2-0.7	0.2-0.7	0.2-0.6	0.2-1.0	0.2-0.6	0.2-1.2
(mg/L)	M±Sd.	0.39±0.15	0.38±0.16	0.38±0.11	0.51±0.23	0.36±0.14	0.46±0.27
NO_2-N	R	0.005-0.031	0.004-0.040	0.003-0.029	0.003-0.037	0.004-0.053	0.003-0.054
(mg/L)	M±Sd.	0.012±0.005	0.013±0.004	0.007±0.003	0.011±0.004	0.016±0.02	0.009±0.005
NO_3-N	R	0.1-1.6	0.2-1.7	0.1-1.7	0.3-3.0	0.2-2.5	0.2-1.7
(mg/L)	M±Sd.	0.84±0.52	0.78±0.54	0.76±0.51	0.90±0.51	1.05±0.85	0.65±0.51
COD_{Mn}	R	6.1-15.7	4.9-12.1	6.1-9.7	6.0-19.5	5.5-9.9	6.4-13.9
(mg/L)	M±Sd.	8.76±3.44	7.62±2.47	7.57±1.45	9.14±4.41	7.50±1.29	9.75±2.24
BOD	R	2.4-6.2	2.3-9.2	2.6-4.7	3.1-6.2	2.0-10.7	2.2-10.7
(mg/L)	M±Sd.	3.94±1.15	5.24±2.39	3.04±1.27	3.85±1.56	4.32±2.72	4.52±2.87
As	R	0.7-5.1	0.8-6.1	1.0-4.9	0.8-6.3	1.1-4.0	0.6-4.1
(ug/L)	M±Sd.	3.09±1.47	2.85±1.80	2.53±1.21	2.54±1.67	2.10±0.90	1.63±0.95
TP	R	0.03-0.17	0.03-0.16	0.01-0.20	0.03-0.21	0.03-0.61	0.02-0.33
(mg/L)	M±Sd	0.080±0.040	0.084±0.038	0.067±0.049	0.085±0.064	0.121±0.045	0.113±0.004
TN	R	1.1-2.4	1.0-2.5	0.6-2.6	0.7-3.6	0.8-4.9	0.7-4.1
(mg/L)	M±Sd.	1.66±0.51	1.49±0.47	1.31±0.54	1.70±0.81	1.98±1.24	1.54±0.92

R: Range; M: Mean; Sd: Standard deviation

3.2 Principle component analysis

The sampling sites were the grouping (dependent) variables, while the measured parameters constituted the independent variables. The result of the PCA analysis base on the correlation matrix of chemical components is expressed in Table 4.

Table 4 Eigenvector and eigenvalues on the correlation matrixes of concentration of physico-chemical parameters in Baiyangdian Lake

Parameters	PC1	PC2	PC3	PC4
TEMP	-0.044	0.161	0.125	0.066
pH	0.001	-0.065	0.357	0.505
EC	0.072	0.111	-0.010	0.009
TSS	0.055	0.036	0.181	-0.343
Ca	0.066	0.036	-0.095	0.384
Mg	-0.029	0.248	0.010	-0.059
K	0.069	-0.086	0.133	0.105
Na	0.072	0.073	-0.007	-0.051
Cl	0.074	0.021	-0.027	-0.225
SO4	0.021	0.253	-0.029	0.057
T-Hard	0.024	0.242	-0.053	0.275
T-Alk	0.073	0.078	0.025	0.154
DO	-0.022	-0.003	0.445	-0.095
NH_4-N	0.079	0.000	0.026	-0.103
NO_2-N	0.078	0.002	0.051	-0.119
NO_3-N	0.075	0.015	0.031	-0.195
COD_{Mn}	0.067	-0.087	-0.005	0.021
BOD	0.044	-0.126	-0.264	0.193
As	0.072	-0.048	0.062	0.255
TP	0.078	-0.032	0.027	-0.023
TN	0.078	0.000	0.038	-0.116
Eigenvalue	12.603	3.627	2.061	1.131
Variability %	60.012	17.269	9.813	5.387
Cumulative %	60.012	77.281	87.094	92.482

Four components of PCA analysis showed 92.5% of the variance in the data set, as the eigenvectors classified the 21 physico-chemical parameters into four groups. The first component (PC1) included nutrient parameters COD_{Mn} and Cl, which accounted for over 60% of the total variance in the data set. In other words, the nutrient parameters, COD_{Mn} and Cl account for similar patterns seen in lake water samples. This group of nutrient parameters also reflected the degree of eutrophication and organic pollution of the lake, suggesting that the anthropogenic pollution mainly stems from the discharge of domestic and industrial sewage. The second component (PC2) included Mg, Na, SO_4, temperature, and electrical conductivity. This component accounted for 17.3% of the total variance measured that demonstrated strong positive loadings for major ions. The third and fourth components (PC3 and PC4) included the physical parameters, K, Ca, and As, which demonstrated 9.8% and 5.4% of the total variance, respectively.

3.3 Cluster analysis

A dendrogram of sampling sites obtained by Ward's method is shown in Figure 2. 13 sampling sites were divided into five groups. Cluster 1 corresponded to sites 3, 4, 5, 6 and 7, which were located in the north lake. Cluster 2 included sites 11, 12, and 13, which were located in the eastern portion of the lake. Cluster 3 contained sites 8, 9 and 10, which were in the lake center. Clusters 4 and 5 corresponded to sites 1 and site 2, respectively, and were located in the western part of the lake. The cluster analysis revealed different properties at each site with respect to physical and chemical variables. The five groups vary according to natural backgrounds features. Additionally, the water quality measured at these sites appeared to be affected by different pollutant sources. A comprehensive pollution index was applied to further demonstrate the results of CA analysis (Table 5).

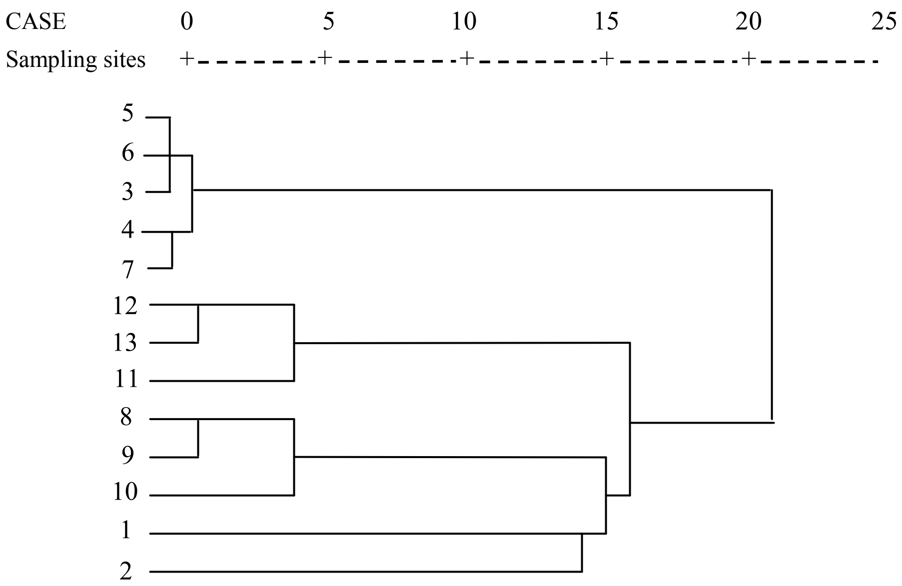

Fig. 2. Dendogram based for agglomerative hierarchical clustering (wards method) based on the PCA scores

Table 5 Single pollution index and comprehensive pollution index of 13 sampling sites

Sites	P_{DO}	P_{NH4}	P_{NO3}	P_{CODMn}	P_{BOD}	P_{As}	P_{TP}	P_{TN}	P
Site 1	1.20	0.44	0.06	1.66	5.34	0.07	5.02	1.18	1.87
Site 2	1.59	15.72	0.30	2.25	4.08	0.16	37.58	19.03	10.09
Site 3	1.98	0.37	0.08	1.53	1.60	0.06	8.40	1.58	1.95
Site 4	1.88	2.97	0.10	1.86	1.82	0.09	8.80	4.61	2.77
Site 5	1.85	0.87	0.06	1.57	1.41	0.08	4.29	1.83	1.49
Site 6	2.08	1.05	0.07	1.63	1.72	0.09	5.66	2.28	1.82
Site 7	2.01	2.35	0.09	1.55	1.88	0.10	8.30	3.54	2.48
Site 8	1.78	0.39	0.08	1.46	0.99	0.06	1.60	1.67	1.00
Site 9	1.93	0.38	0.08	1.27	1.31	0.06	1.68	1.49	1.02
Site10	1.67	0.38	0.08	1.26	0.76	0.05	1.34	1.31	0.86
Site11	1.54	0.51	0.09	1.52	0.96	0.05	1.70	1.70	1.01
Site12	2.21	0.36	0.10	1.25	1.08	0.04	2.42	1.98	1.18
Site13	2.00	0.46	0.06	1.62	1.13	0.03	2.27	1.54	1.14

Group 1 (northern portion of the lake), the sampling locations were determined to have mainly been subjected to effluents from non-point sources. The values of the comprehensive pollution index were 1.9, 2.8, 1.5, 1.8 and 2.5 for sites 3, 4, 5, 6 and 7, respectively, which demonstrated moderate or severe pollution. Water quality in site 3 was determined to have been influenced by sewage received from Tanghe River reservoir. Sampling locations from site 4 are subjected to many tourist attractions, with diesel boats being the primary means of transportation means around the lake. The pollutants in sites 5, 6 and 7 were determined to be from agricultural and domestic sewage, particularly the dispersed and unsettled wastewaters from local villages. Group 2 (east of the lake) showed lower pollutant levels than Group 1, with comprehensive pollution indexes measured at 1.0, 1.2, and 1.1 for sites 11, 12 and 13, respectively. The water quality at Group 2 sampling locations was determined to be only slightly polluted, with domestic wastes being the major source of pollutants. Group 3, where samples were collected from the lake center, was determined to be the least polluted area. The comprehensive pollution indexes for Group 3 were measured at 1.0, 1.0 and, 0.8 for sites 8, 9 and 10, respectively. Group 4, (site 1 in the western portion of the lake) was determined to be moderately polluted, with a comprehensive pollution index value of 1.9. Higher concentrations of BOD and COD_{Mn} were measured in this area. Group 5 (site 2 also in the western portion of the lake) sampling locations were located near the entrance of Fuhe River into Baiyangdian Lake; as such, these were the most heavily polluted sampling sites. The comprehensive pollution index was 10.1, determined to be four times higher than the limit value established for severe pollutant levels.

In addition, the standardized analytical data set was used to compare the variation of physico-chemical parameters measured at different sampling sites, as shown in Fig. 3. With the exception of BOD and Mg, all other parameters were found to be high at sampling site 2, showing significant pollution at this site; as such, it is considered to be the main contributing source of pollution in the lake. Most physico-chemical parameters demonstrate low normalized values at sites 8, 9, 10, 11, 12 and 13 (below 0), whereas the values of these parameters are less than 1 at sites 3, 4, 5, 6 and 7.

In summary, Clusters 2 and 3 demonstrated the lowest pollution levels, Clusters 1 and 4 demonstrated moderate pollution, and Cluster 5 demonstrated the highest levels of pollution. Accordingly, spatial variation of water quality in Baiyangdian Lake showed that water quality was better in center and eastern portion than in western and northern areas. This implies that, for a rapid assessment of water quality, only one site in each cluster presents a useful spatial assessment of the water quality for the entire network. It is evident that the CA technique is useful in offering reliable classification of surface waters throughout a whole region, and will make it possible to accurately perform spatial assessment in an optimal manner. Thus, the number of sampling sites and associated cost in operating the monitoring network can be reduced without compromising the study.

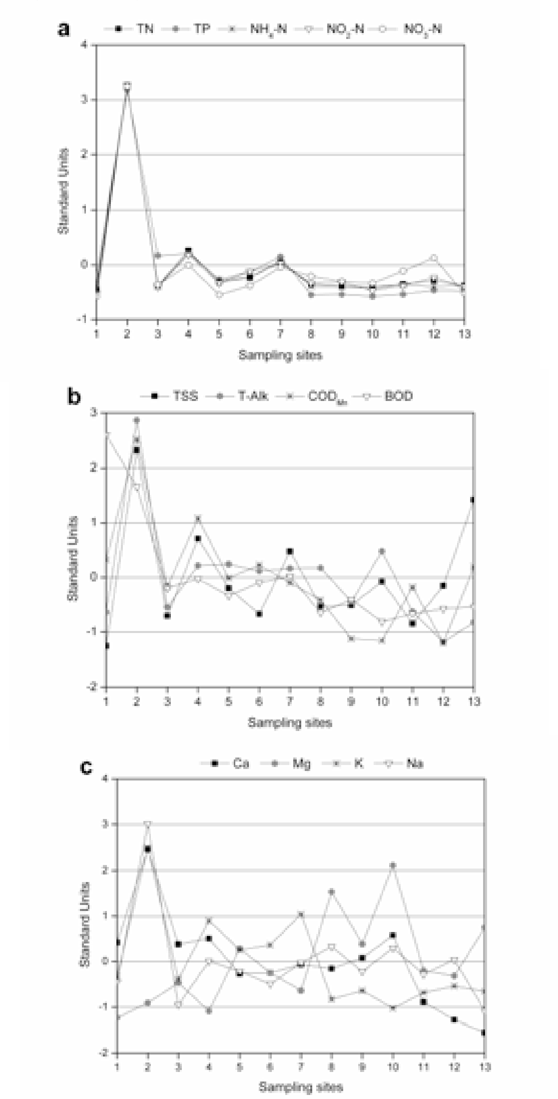

Fig. 3. Standard units of chemical concentrations measured in Baiyangdian Lake. The standard unit is defined as z=(x-u)/S, where x is the raw concentrated data, u represents the mean values, and S is the standard deviation. (a) TN, TP, NH_4-N, NO_2-N and NO_3-N; (b) TSS, T-Alk, CODMn, BOD; (c) Ca, Mg, K and Na.

4. Conclusion

In this study, different multivariate statistical techniques were used to evaluate variations in surface water quality of Baiyangdian Lake. Cluster analysis grouped 13 sampling sites into five clusters according to similar water quality characteristics. Based on information obtained from this study, it is possible to design an optimal sampling strategy, resulting in a reduced number of sampling sites and lower associated costs. PCA helped to identify the factors or sources responsible for water quality variations. The main cause of degradation to the lake is determined to be the discharge of industrial and agricultural wastes, domestic sewage w from the upstream Fuhe River, and pollution from local villages around the lake. This study illustrates the usefulness of multivariate statistical techniques in the analysis and interpretation of complex data sets, in identifying pollutant sources, and in understanding variations in water quality for effective lake water management. Measures should be taken to reduce anthropogenic discharges in the lake; otherwise, high levels of pollution have the potential to influence the population and contribute to socio-economic disaster. These results should be considered for future planning and management of the lake.

Acknowledgements

Financial support from the National Water Pollution Control and Treatment Project in China (No. 2008ZX07209-009), the Major State Basic Research Development Program (No.2010CB951104), and the Fundamental Research Funds for the Central Universities are highly appreciated.

References

[1] Kazi TG, Arain MB, Jamali MK, Jalbani N, Afridi HI, Sarfraz RA, Baiga JA, Shaha, Abdul Q. Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. *Ecotox Environ Safe* 2009; **72**: 301-9

[2] Niemi GJ, Devore P, Detenbeck N, Taylor D, Lima A. Overview of case studies on recovery of aquatic systems from disturbance. *Environl Manage* 1990; **14**: 571 – 87

[3] Herrera-Silveira JA, Morales-Ojeda SM. Evaluation of the health status of a coastal ecosystem in southeast Mexico: assessment of water quality, phytoplankton and submerged aquatic vegetation. *Mar Pollut Bull* 2009; **59**:72 – 86

[4] Silvert W. Fuzzy indices of environmental conditions. In: INDEX Workshop on Environmental Indicators and Indices, St. Petersburg, Russia; 1998.

[5] Icaga Y. Fuzzy evaluation of water quality classification. *Ecol Indic* 2007; **7**: 710 – 18

[6] Reghunath R, Murthy TRS, Raghavan BR. The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. *Water Res* 2002; **36**: 2437 – 42

[7] Simeonova P, Simeonov V, Andreev G. Water quality study of the Struma River Basin, Bulgaria (1989–1998). *Cent Eur J Chem* 2003; **1**: 136–212

[8] Kotti ME, Vlessidis AG, Thanasoulias NC, Evmiridis NP. Assessment of river water quality in Northwestern Greece. *Water Resour Manag* 2005; **19**: 77 – 94

[9] Singh KP, Malik A, Mohan D, Sinha, S. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)-a case study. *Water Res* 2004; **38**: 3980-92

[10] Kuppusamy MR, Giridhar VV. Factor analysis of water quality characteristics including trace metal speciation in the coastal environmental system of Chennai Ennore. *Environ Int* 2006; **32**: 174 – 9.

[11] Ocampo-Duque W, Ferre-Huguet N, Domingo JL, Schuhmacher M. Assessing water quality in rivers with fuzzy inference systems: a case study. *Environ Int* 2006; **32**: 733 – 42.

[12] Wang YS, Lou ZP, Sun CC. Wu ML, Han SH. Multivariate statistical analysis of water quality and phytoplankton characteristics in Daya Bay, China, from 1999 to 2002. *Oceanologia* 2006; **48**: 193-213

[13] Chau KW, Muttil N. Data mining and multivariate statistical analysis for ecological system in coastal waters. *J Hydroinform* 2007; **9**:305 – 17

[14] Mendiguchía C, Moreno C, García-Vargas M. Evaluation of natural and anthropogenic influences on the Guadalquivir River (Spain) by dissolved heavy metals and nutrients. *Chemosphere* 2007; **69**: 1509-17

[15] Wu ML, Wang Y.S. Using chemometrics to evaluate anthropogenic effects in Daya Bay, China. *Estuar, Coast and Shelf S* 2007; **72**:732 – 42

[16] Nakano T, Tayasu I, Yamada Y, Hosono T, Igetaa A, Hyodo F. Andoa A, SaitohaY, Tanakaa T, Wadad E, Yachia S. Effect of agriculture on water quality of Lake Biwa tributaries, Japan. *Sci Total Environ* 2008; **389**:132 – 48.

[17] Xu M, Zhu J, Huang Y, Gao Y, Zhang S, Tang Y. The ecological degradation and restoration of Baiyangdian Lake, China. *J Freshwater Ecol* 1998; **13**: 433-46.

[18] Hua GC, Luo XJ, Li FC, Dai JY, Guo JY, Chen SJ, Hong C, Mai BX, Xu MQ. Organochlorine compounds and polycyclic aromatic hydrocarbons in surface sediment from Baiyangdian Lake, North China: concentrations, sources profiles and potential risk. *J Environ Sci-China* 2010; **22**: 176–83

[19] Filik Iscen C, Emiroglu̇ö, Ilhan S, Arslan N, Yilmaz V, Ahiska S. Application of multivariate statistical techniques in the assessment of surface water quality in Uluabat Lake, Turkey. *Environ Monit Assess* 2008; **144**: 269–76

[20] Wenning, RJ, Erickson, GA. Interpretation and analysis of complex environmental data using chemometric methods. *Trends Anal Chem* 1994; **13**: 446 – 57

[21] SEPAC. State Environmental Protection Administration of China. *Environmental quality standard for surface water, China (GB3838-2002)*. Beijing: China Environmental Science Press; 2002 (in Chinese).

[22] Alberto WD, Del Pilar DM, Valeria AM, Fabiana PS, Cecilia HA, De Los Angeles BM. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia River basin (Cordoba—Argentina). *Water Res* 2001; **35**: 2881 – 94.

[23] Helena B, Pardo R, Vega M, Barrado E, Fernandez JM, Fernandez L. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis. *Water Res* 2000; **34**: 807 – 16

[24] Kaufman L, Rousseeuw P J. Finding groups in data: an introduction to cluster analysis. New York: Wiley; 1990.

[25] Jarvie HP, Whitton BA, Neal C. Nitrogen and phosphorus in east coast British rivers: Speciation, sources and biological significance. *Sci Total Environ* 1998; **210/211**: 79–109.

[26] Jiang HH, Zhu JP, Liang DH, Wu Z L. The relationship between comprehensive pollution index assessment and water quality type distingquishing.*Environmental Monitoring in China* 1999; **15**: 46-8 (in Chinese).

[27] Liu XY, Wan RR, Zhou BG. Analysis on the water quality changes in Guajinggang River. *Water Conservancy Science and Technology and Economy* 2010; **16**: 173-5 (in Chinese).

[28] Zhang T, Liu JL, Wang XM. Causal analysis of the spatial-temporal variation of water quality in Baiyangdian Lake. *Acta Science Circumstantiae* 2010; **30**: 217-61 (in Chinese).

[29] Silva MAL, Rezende CE. Behavior of selected micro and trace elements and organic matter in sediments of freshwater system in south – east Brazil. *Sci Total Environ* 2002; **292**: 21 – 128.

[30] Solaraj G, Dhanakumar S, Murthy KR, Mohanraj R. Water quality in select regions of Cauvery Delta River basin, southern India, with emphasis on monsoonal variation. *Environt Monit Assess* 2010; **166**:435-44

[31] Solanki VR, Hussain MM, Raja SS. Water quality assessment of Lake Pandu Bodhan, Andhra Pradesh State, India. *Environment and Monitoring Assessment* 2010; **163**: 411-9.

[32] Galal-Gorchev H, Ozolins G, Bonnefoy X. Revision of the WHO guidelines for drinking water quality. Ann. Institutionali. Super Sanit. 1993; **29**: 335-45.

[33] WHO. *Guidelines for Drinking-Water Quality*, third ed. WHO, Geneva,Switzerland; 2004.