ASSESSMENT OF WATER QUALITY IN TAIHU LAKE USING A RADIAL BASIS FUNCTION NETWORK, STRUCTURE INDEX, AND PRINCIPAL COMPONENT ANALYSIS

HANG, X. – GAO, H.* – JIA, S.

College of Information and Electrical Engineering, China Agriculture University
Beijing 100083, China
(phone: +86-10-6273-8830; fax: +86-10-6273-6746)

**Corresponding author*
e-mail: hjgao@cau.edu.cn

(Received 3rd Jun 2019; accepted 11th Oct 2019)

Abstract. A novel integrated approach was used to assess water quality in Taihu Lake in real-time using a radial basis function (RBF) network, structure index (SI) and principal component analysis (PCA). A total of eight sampling points covering an area of 2,338 km^2 were sampled once weekly from 2000 to 2005, except for 2003 and 2004. An RBF network was developed with 23 water quality parameters used as inputs. Then SI was applied to investigate main parameters. Parameters affecting water quality were also determined using PCA. The water quality of Taihu Lake was estimated by synthesizing the results of SI and PCA. Total nitrogen, with a weight of 0.34 for the first component, had the greatest negative effect on Taihu Lake water quality, being slighter than that of ammonia nitrogen, which had a weight of 0.324 for the first component. Conductivity, total phosphorus and alkalinity had weights of 0.322 0.292 and 0.267 for the first component, respectively. It is possible to identify potential sources of water pollution and the results indicated that the water quality of Taihu Lake in the study period was generally graded as Class V according to Environmental quality standards for surface water (GB 3838-2002).
Keywords: *PCA, RBF neural network, SI, water quality assessment, Taihu*

Introduction

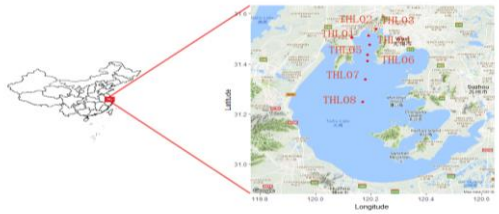
Environmental sustainability is among the most serious global issues being addressed by planners, policy analysts, political scientists and others (Gleeson, 2001). Water pollution is an increasing concern, threatening the ecological integrity and sustainability of some of the world's largest water bodies (Havens et al., 2001). Lake water pollution is of particular concern because of its impact on fish habitats (Scavia et al., 2014) and human and animal health (Utah, 2016), and the fact that these are areas where there are complex interconnections between anthropogenic and climatic factors (Richards, 2010; Brooks, 2016). Harmful cyanobacterial blooms are one of the most significant markers of impaired lake water quality, especially in eastern China, a region with a rapidly growing economy (Paerl, 2011).

Taihu Lake (*Fig. 1*), situated in the Yangtze delta with an area of 2,338 km^2, is the most industrialized area in China, having a high population density and significant urbanization and economic development (Qin et al., 2002). The lake supports 10 million people working in tourism, fisheries, shipping and other industries. Taihu Lake also collects waste from urban, agricultural and industrial areas of surrounding cities, helping rapid growth in local economies (Qin et al., 1999; Qin, 2008). However, the increase of nutrients coming from urban and agricultural development in the watershed, has accelerated eutrophication (James et al., 2009; Duan et al., 2009). Cyanobacterial blooms, indicators of advanced eutrophication, in freshwater lakes impact fish,

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1424114258

recreational lake users, and the drinking water supply (Fogg, 1969; Paeral,1988). The water quality of the lake is closely related to public health in the local area.

Prior to the 1970s, average water quality of Taihu Lake was classified as Grade II according to GB3838-88 (Groundwater Environmental Quality Standard. China: GB. 2003). Since the 1980s, discharge of industrial waste water and domestic sewage has increased yearly while the water quality of Taihu Lake has dropped, on average, by one grade every 10 years. In the 1980s, the water quality was graded between the second and third levels; in the 1990s, the water quality decreased to the fifth level, and eutrophication had increased. Since the late 1990s, water quality deterioration has slowed and some key indicators, including TP, ammonia nitrogen and chemical oxygen demand of permanganate (Permanganate index), have improved (China Taihu Ecosystem Positioning Observation Data Query System. Taihu Lake Water Quality Inquiry System. Accessed 1990-2018, http://lake.data.ac.cn/taihu).



The name of observation station	East longitude (°)	North latitude (°)
THL01	120.13117	31.50383
THL02	120.19067	31.51317
THL03	120.21944	31.53968
THL04	120.19433	31.47633
THL05	120.18796	31.43609
THL06	120.18733	31.41117
THL07	120.18017	31.33833
THL08	120.17062	31.24816

Figure 1. *Location of the study area*

Statistical techniques, visual modeling, prediction algorithms, and time-series analyses have been used to evaluate and monitor water quality, and inform policy-makers (Junli et al., 2019; Jun et al.,2019; Taufiq et al., 2019; Dutta et al., 2018; Jabbar et al., 2018; Meng et al.,2018; Tong et al.,2018; Drakard et al., 2018; Shengli et al., 2018; Omar et al., 2018; Putri et al., 2018; Kellner et al., 2018; Islam et al., 2018; Skowron et al., 2018; Masindi et al., 2018; Meifang et al., 2018; Guilin et al., 2018; Zhi-Qiang et al., 2018; Ju et al., 2018; Cecconello et al., 2018). Methods that have been proposed to assess water quality include Roveda's water quality index (Roveda et al., 2013), EFER (Aminravan et al., 2013) and use of ships with deliberative control architecture (Halal et al., 2014). A RBF network algorithm is commonly used because of its ability to learn, generalize, and adapt (Garrett, 1994). The RBF network algorithm used in this study combines several lake water quality parameters, enabling the most important parameters to be identified based on analysis of the neighboring nodes. It can efficiently describe the non-linear relationships among complex water quality datasets (Adel et al., 2018; Zheng et al., 2017; Alizamir et al., 2017; Asadollahfardi et al., 2017; Wu et al., 2016; Zounemat-Kermani et al., 2016; Shiau et al., 2016; Bagheri et al., 2016).

Combining RBF with SI and PCA (Dutta et al., 2018; Prusty, 2018; Sun et al., 2018; Xiaohu et al., 2018; Wang et al., 2019; Xiao et al., 2019; Zhaoxue et al., 2018) yields more accurate results and provides sufficient evidence to characterize lake water

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1424114258

quality. In this study, RBF, SI and PCA were applied to identify the main parameters affecting water quality. The primary objectives of this study were to assess water quality in Taihu Lake, identify the main parameters that damage the water quality, and investigate possible sources of pollution.

Material and methods

This study focused on Taihu Lake, which is located in the southern border of the Yangtze River Delta (*Fig. 1*). Taihu Lake is the third largest freshwater lake in China, covering an area of 2,338 km^2. It has a length of 68.5 km from north to south, an average length of 34 km from east to west, and a coastline of 405 km.

Taihu Lake has a complicated river and channel network, with a mean depth of 1.9 m and a maximum depth of 2.6 m. Pen fish is the main form of aquaculture in Taihu Lake. Marsh development occurs primarily along the eastern shoreline. Taihu Lake has many functions, including supplying drinking water, supporting fisheries, tourism and shipping, and retaining flood water.

To evaluate the water quality of Taihu Lake, monthly data from eight sampling points in Taihu Lake were collected for the period 2000 to 2005, except for 2003 and 2004. The data used in this study are from "Lake-Watershed Science Data Center, National Earth System Science Data Sharing Infrastructure, National Science & Technology Infrastructure of China" (http://lake.geodata.cn). Parameters evaluated included pH, total nitrogen, total phosphorus, Chlorophyll a, Demagnesium chlorophyl, permanganate index, dissolved oxygen, ammonia nitrogen, nitrite nitrogen, nitrate, phosphate, alkalinity, potassium, sodium ion, calcium ion, magnesium ion, fluorine ion, chloride, sulfate, silicate, Dissolved total nitrogen, Dissolved total phosphorus, and conductivity. The abbreviations of words used in this paper are shown in *Table 1*.

This study combined RBF, SI and PCA to derive accurate results. Many parameters were included in our assessment of Taihu Lake water quality. Changes in parameter values were captured and analyzed in the MATLAB environment (MathWorks, Natick, MA, USA) using RBF algorithms to derive an accurate results. A novel three-layer RBF network was developed for this study. A confusion matrix was used to assess the performance of the RBF algorithm.

After the RBF network was built, the weights of and links between nodes were ascertained. SI was used to determine the parameters contributing most significantly to the output neurons. SI shows a high value when a parameter makes a great contribution to the RBF network, even if it has a less weight.

Following the above steps, the main parameters affecting each output neuron were determined. SPSS software (ver. 21.0; IBM Corp., Armonk, NY, USA) was used to compute the principal components, that is to seek projection directions with maximal variances which occupy bigger quantity (Flury, 1990; Joliffe, 2011). By investigating the principal components, the most important parameters affecting water quality can be identified using the RBF algorithm. A flow diagram illustrating the analysis process is shown in *Figure 2*.

Assessment model of water quality based on RBF algorithm was used to estimate the percentage amounts of each measured parameter in the lake. All data were pre-processed and separated into three data sets: training, test and verification sets. MATLAB was used for construction and analysis of the RBF network. There were 23 input layer neurons and 1 neuron in the output layer was used to represent the actual

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1424114258

water quality shown in *Figure 3*. There can be up to 71 hidden layers in this network. The nodes in the input and output layers are arranged in one line. A confusion matrix was used to assess the performance of the network and indicated that most of the predicted data fit, or nearly fit, the actual results.

The SI for the whole network was calculated as

$$\sum_{j}^{n} \sum_{k=1}^{j=1} \frac{|Wij - Wik|}{Djk} \quad \text{(Eq.1)}$$

where W_{ij} and W_{ik} are the connection weights of input variable i in neurons, and j and k of the RBF network, respectively; $||D_{jk}||$ is the topological distance between neurons j and k; and N is the total number of output neurons in the RBF network.

Table 1. *The abbreviations of words used in this paper*

Words	Abbreviation	Unit
pH	pH	/
Total nitrogen	TN	mg/L
Total phosphorus	TP	mg/L
Chlorophyll a	CHLA	μg/L
Demagnesium chlorophyl	DC	μg/L
Permanganate index	PI	mg/L
Dissolved oxygen	DO	mg/L
Ammonia nitrogen	NH3-N	mg/L
Nitrite nitrogen	NN	mg/L
Nitrate	NNA	mg/L
Phosphate	PO	mg/L
Alkalinity	ALKY	mmol/L
Potassium ion	K	mg/L
Sodium ion	Na	mg/L
Calcium ion	Ca	mg/L
Magnesium ion	Mg	mg/L
Fluorine ion	F	mg/L
Chloride	Cl-	mg/L
Sulfate	SO4	mg/L
Silicate	SiO	μmol/L
Dissolved total nitrogen	Dissolved TN	mg/L
Dissolved total phosphorus	Dissolved TP	mg/L
Conductivity	EC	μS/cm

The SI of each input node is presented in *Table 2*. Input layer 14, which represents Na, was the parameter most critical to the performance of the whole network. Dissolved TP and F were the second and third most important parameters, respectively. Ca, with an SI of 21.437 was next, followed by TN, PI and DO. Dissolved TN, Mg, NH3-N,

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1424114258

NNA, NN, pH, EC, ALKY, and TP had SIs above 5; SiO, Cl-, CHLA, PO, DC, K, and SO4 had SIs below 5. Garson's Algorithm (Garson, 1991) and Connection weights approach (Olden, 2002b) are also presented as methods to study the relative importance of inputs on the output. The results are displayed in *Table 2*. Seven factors affecting water quality were identified by developing the network and analyzing the weight of each node by SI as well as synthesizing the results of Garson and Connection weights which was in the order of: TP, SiO, Ca, SO4, Na, K, and F.

Table 2. *The weight of each input node by different methods*

Input layer(s)	Parameter(s)	Weight(s) by SI algorithm	Weight(s) by Garson's algorithm	Weight(s) by connection weights approach
1	pH	6.556	589.59	0.801
2	TN	20.016	247.61	0.300
3	TP	5.145	9.663	0.012
4	CHLA	3.521	1661.71	2.028
5	DC	1.291	374.09	0.459
6	PI	16.219	400.97	0.523
7	DO	13.787	608.08	0.847
8	NH3-N	8.251	80.472	0.086
9	NN	6.644	5.041	0.006
10	NNA	7.739	59.17	0.073
11	PO	3.050	137.11	0.181
12	ALKY	5.658	2558.9	3.323
13	K	1.025	2160.1	2.897
14	Na	43.750	603.89	0.809
15	Ca	21.437	39.15	0.053
16	Mg	8.386	3094.08	4.049
17	F	23.735	4489.51	5.887
18	Cl-	4.632	4050.9	5.185
19	SO4	0.362	32.35	0.044
20	SiO	4.880	2609.58	3.581
21	Dissolved TN	9.444	0.933	0.001
22	Dissolved TP	25.069	314.82	0.418
23	EC	5.959	30309	39.439

PCA was conducted to investigate the impacts of TN, NH3-N, PI, and ALKY. TN, TP, NN, K, SO4, SiO and PI were used to assess lake water quality according to *Tables 3* and *4*.

Table 3. *Relative importance of the studied components to water quality*

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
Proportion of variance	0.3009486	0.1968562	0.1087038	0.07997529	0.06654579	0.0451578	0.03621469	0.03031488	0.02527567	0.02249223
Cumulative proportion	0.3009486	0.4978047	0.6065085	0.68648383	0.75302961	0.7981874	0.83440210	0.86471699	0.88999266	0.91248488

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1424114258

Table 4. *Component matrix*

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
pH	0.179	0	0.41	0	-0.22	0	0.252	0.114
TN	-0.34	-0.124	0	0	0.167	0	0	0
TP	-0.292	0	0	0	0	-0.332	0	0.218
CHLA	0	-0.421	0.37	0	-0.104	0	0	0
DC	-0.101	-0.381	0.223	0	-0.158	0.111	-0.28	0.108
PI	-0.199	-0.345	0.216	0	0	-0.18	0	0.158
DO	0.173	0	0.225	-0.419	0	0.202	0	0.238
NH3-N	-0.324	0	-0.151	0	0.102	-0.125	0.131	0
NN	-0.201	-0.258	0	0.156	0.129	0.259	0	-0.257
NNA	-0.213	0	0.111	0	0.318	-0.204	-0.288	-0.288
PO	-0.23	0	-0.173	-0.105	-0.388	0.564	0.232	0
ALKY	-0.267	0	-0.138	0	0	-0.105	0	0
K	-0.241	0.292	0.263	0.159	-0.146	0	0	-0.104
Na	-0.247	0.276	0.255	0.103	0	0	0	0
Ca	-0.177	0.296	0	0.11	-0.347	0.136	-0.333	0.263
Mg	-0.217	0.377	0.253	0	-0.142	0	0	-0.1
F	0	0	0	0.363	0.396	0.339	0.314	0.656
Cl-	-0.27	0.122	0	-0.467	0	0	0.113	0
SO4	-0.258	0.109	0	-0.486	0.127	0	0	0
SiO	-0.104	0	-0.383	0	-0.237	0	-0.475	0.382
Dissolved TN	0	-0.104	-0.173	-0.105	-0.338	0.564	0.232	0
Dissolved TP	0	0	-0.232	0	-0.435	0.137	0.491	-0.132
EC	-0.322	0	0	0	0	0	0	-0.274

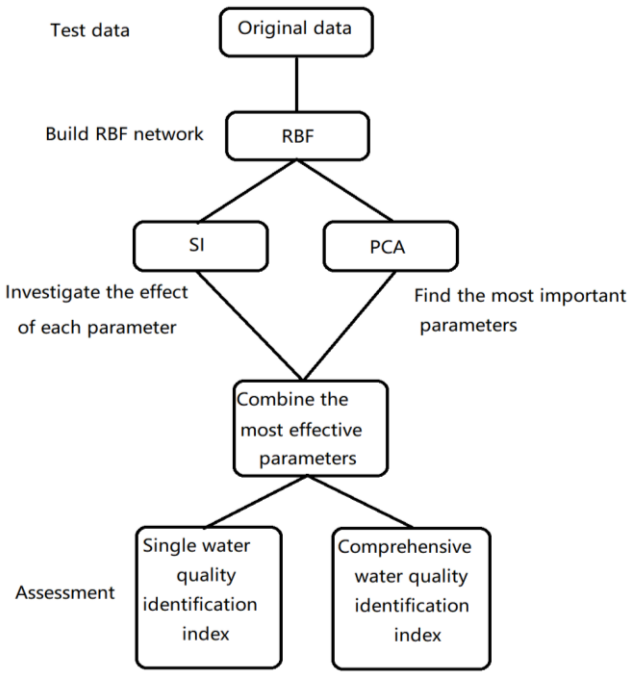


Figure 2. *Process flow diagram*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1424114258

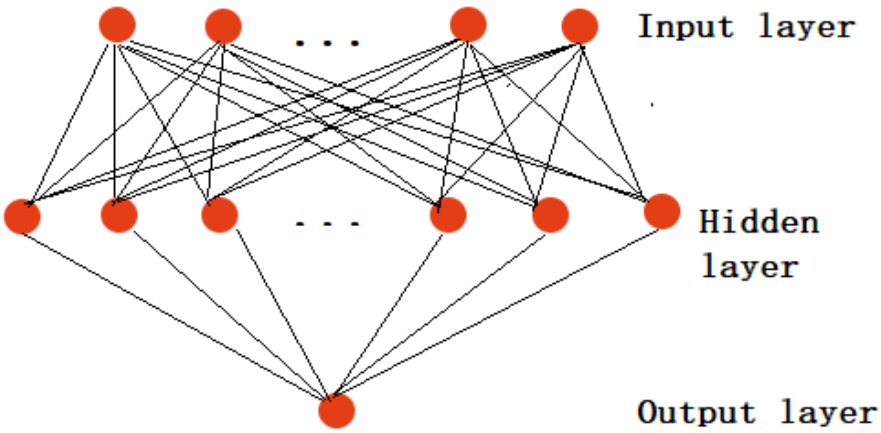


Figure 3. *Basic network structure*

Some criteria were used to evaluate the performance of the RBF model. The sum-squared error which can be calculated in *Equation 2* show the total epochs (iterations) required for the function to converge on parameters. Fitness value, indicators used to measure the quality of individuals in a population, can be calculated in *Equation 3.* Mean square error calculated in *Equation 4* was used to assess performance of each model.

$$\text{Sum - squared - error} = \sum (\text{actual value} - \text{predicted value})^2 \quad \text{(Eq.2)}$$

$$\text{Fittting} = \frac{1}{\text{Sum - squared - error}} \quad \text{(Eq.3)}$$

$$\text{mean square error} = \frac{\text{Sum - squared - error}}{n} \quad \text{(Eq.4)}$$

Results

Analysis of the main parameters affecting water quality

During the study period, TP concentrations in the samples ranged from 0.013 to 2.133 mg/L, with an average of 0.145 mg/L (*Table 5*). The peak concentrations in the third quarter of 2000 and 2002 may be due to the construction of new factories and the associated increase in the quantity of polluted water entering Taihu Lake during this period.

Increasing temperatures in spring promote the release of TP and increase its concentration in water. Throughout the summer, the concentration of TP in water may decrease because aquatic plants consume nutrients (*Fig. 4-1a*). TP peaks in 2000 and 2005 occurred in February, while the peaks in 2001 and 2002 occurred in July.

The dissolution of SiO was relatively stable, ranging from 0 to 157 mg/L and with an average of 77.27 mg/L (*Fig. 4-1b*). The low SiO concentration observed in the fourth quarter of 2000 may have been related to diatom activity. However, the variation in SiO concentrations seen in 2002 and 2005 indicates that diatom activity is likely not the

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1424114258

main factor affecting SiO concentrations; values were high all year round, particularly in the autumn and winter. Organic pollution of the water body is another possible factor influencing SiO concentrations, which were at an acceptable level for drinking water.

Table 5. *Statistical analysis of the main parameters affecting water quality in Taihu Lake*

Parameter	Max	Min	Avg	Var
pH	9.44	7.25	8.20	6.43
TN	13.34	0.39	6.32	309.40
TP	2.133	0.013	0.145	0.35
CHLA	521.73	0	33.77	16455.58
DC	58.03	0	5.60	505.23
PI	13.6	2.69	6.10	54.23
DO	14.74	0.47	6.72	241.88
NH3-N	6.95	0.001	3.33	230.21
NN	10.54	0.002	0.10	0.13
NNA	4.28	0.001	1.22	12.11
PO	0.162	0	0.015679688	0.000690782
ALKY	3.35	0.147	2.41	10.63
K	10.33	1.73	5.72	65.93
Na	210	4.4	77.65	16111
Ca	125.4	14.3	32.14	2860
Mg	49	2.2	10.58	159.4
F	10.37	0	1.09	81.51
Cl-	114.2	12	44.15463542	241.4987941
SO4	174.6	24.1	58.12	10915.91
SiO	157	0	77.27	8390.89
Dissolved TN	11.31	0.2	1.673229167	3.581032365
Dissolved TP	0.245	0.003	0.035744792	0.0015636
EC	1100	195	195	121.74

Ca concentrations ranged from 14.3 to 125.4 mg/L; Na concentrations ranged from 4.4 to 210 mg/L; K concentrations ranged from 1.73 to 10.33 mg/L; and Flu concentrations ranged from 0 to 10.73 mg/L.

Ca concentrations are expected to show opposite trends to SiO concentrations; this was observed in 2000 and 2001 in *Figure 4-1c*. The decrease in Ca concentration in 2003 may have been caused by polluted water emanating from factories and a nearby residential development. By 2005, the effect of human and diatom activity had reached a balance, and minimal variation Ca concentrations was observed.

SO4 is toxic to human health even at low concentrations. Over the study period, the SO4 concentration in Taihu Lake ranged from 24.1 to 174.6 mg/L, with an average value of 58.12 mg/L. SO4 concentrations declined through the middle part of the year, except in 2005 (*Fig. 4-1d*). The high SO4 concentration during the winter and spring from 2000 to 2002 may have been due to coal mine emissions. The peak SO4 concentration was observed from 2000 to 2002, with the appearance of the peak varied from January to April. SO4 concentrations were consistent with SiO concentrations (*Fig. 4-1b* and *d*, respectively).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1424114258

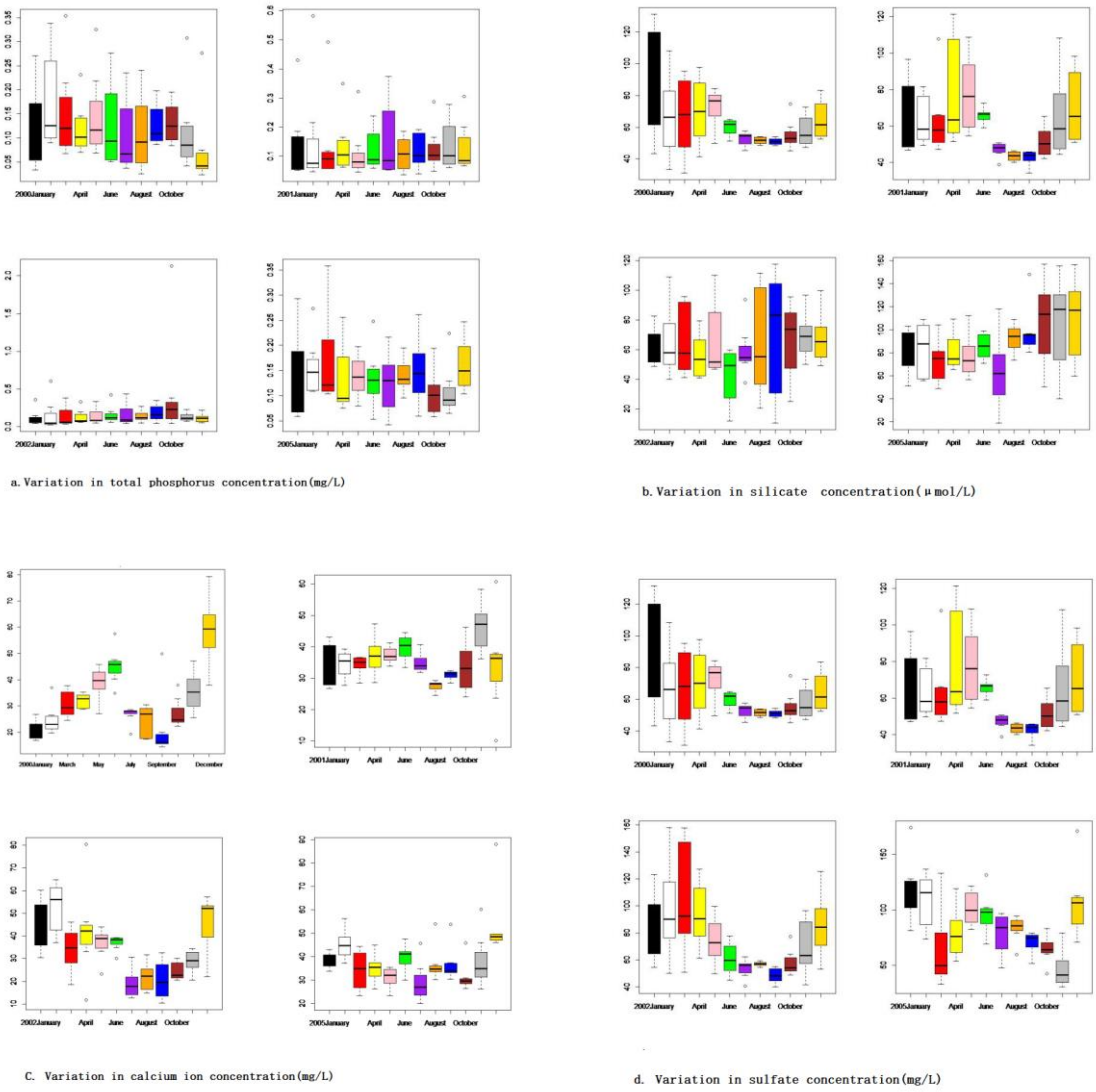


Figure 4-1. *Variation in different parameters' values by study year*

Nas are a common component of natural waterways. The most important characteristic of Nas in natural waterways is the wide variation in concentration under different conditions. *Figure 4-2e* shows that Na concentrations were low during the wet seasons. In the Earth's crust, K and Na are similarly abundant (2.60% and 2.64%, respectively). Although they have similar chemical properties, K concentrations are generally much lower than Na concentrations in natural waterways. K and Na concentrations decreased from 10% to 4% from 2000 to 2005 (*Fig. 4-2e* and *f*, respectively). The K to Na quality ratio was 0.1028 in 2005, while the ratio was 0.118, 0.120, 0.128 in 2003, 2001 and 2000 respectively. Ks are more mobile than Nas due to soil colloid adsorption, but are absorbed by plants. Increases in Na concentrations were mainly due to the drainage of sewage into the basin.

An acceptable F concentration in drinking water is 0.5 - 1.0 mg/L; the F concentration in Taihu Lake was within this range (*Fig. 4-2g*). The peak high and low points in 2000 and 2001 were similar, although the peak occurred later in 2001. The F concentration trend in 2002 was opposite to that in 2001.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1424114258

TN and SO4 concentration trends were opposite which are illustrated in *Figures 4-2h* and *4-1d*. TN concentration was highest in the middle part of 2000, while in the following years there were low TN concentrations in the middle of the year.

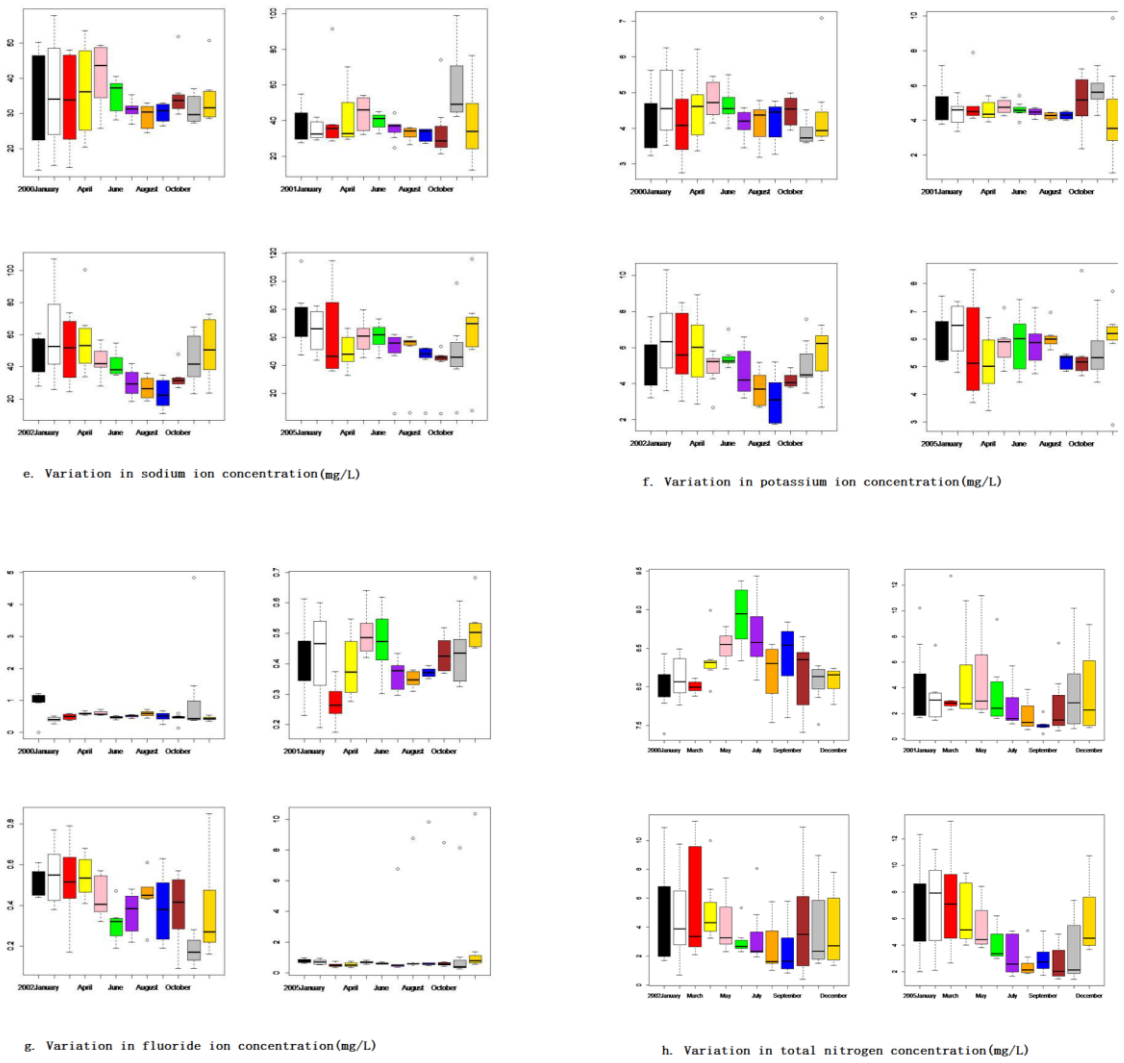


Figure 4-2. *Variation in different parameters' values by study year*

Formation of NH3-N is due to a lack of oxygen during the conversion of ammonia to nitrate. Over the study period, the NH3-N concentration varied from 0.001 mg/L to 6.95 mg/L, with an average of 0.10 mg/L. The NH3-N concentration was high during autumn and winter (*Fig. 4-3i*) and has increased in recent years in Taihu Lake; this is directly related to the lack of oxygen in the water caused by eutrophication of the water body.

PI is commonly used as a comprehensive indicator of the degree of contamination of surface water by organic and inorganic matter. The PI values ranged from 2.69 to 13.6 over the study period. In 2000, the PI concentration was low but exhibited relatively large variation in the following years (*Fig. 4-3j*). In addition, in 2005, the PI concentrations were relatively high.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1424114258

ALKY indicates that a water body contains a substance that can accept hydrogen ions. The ALKY of Lake Taihu in summer and autumn was lower than in winter and spring (*Fig. 4-3k*). Over the study period, the minimum value was 0.147, with an overall upward trend in recent years and a peak value of 3.35. The temperature in the Taihu Lake basin is relatively high in summer and autumn. When there are large numbers of plants, microorganisms will produce acidic substances that neutralize some alkaline substances. Additionally, the use of coal for heating in the winter and spring seasons has greatly increased.

EC trends were consistent from 2000 to 2005 (*Fig. 4-3l*). In the spring and summer, as temperature increased, so too did EC. During the wet season, EC decreased due to density loss.

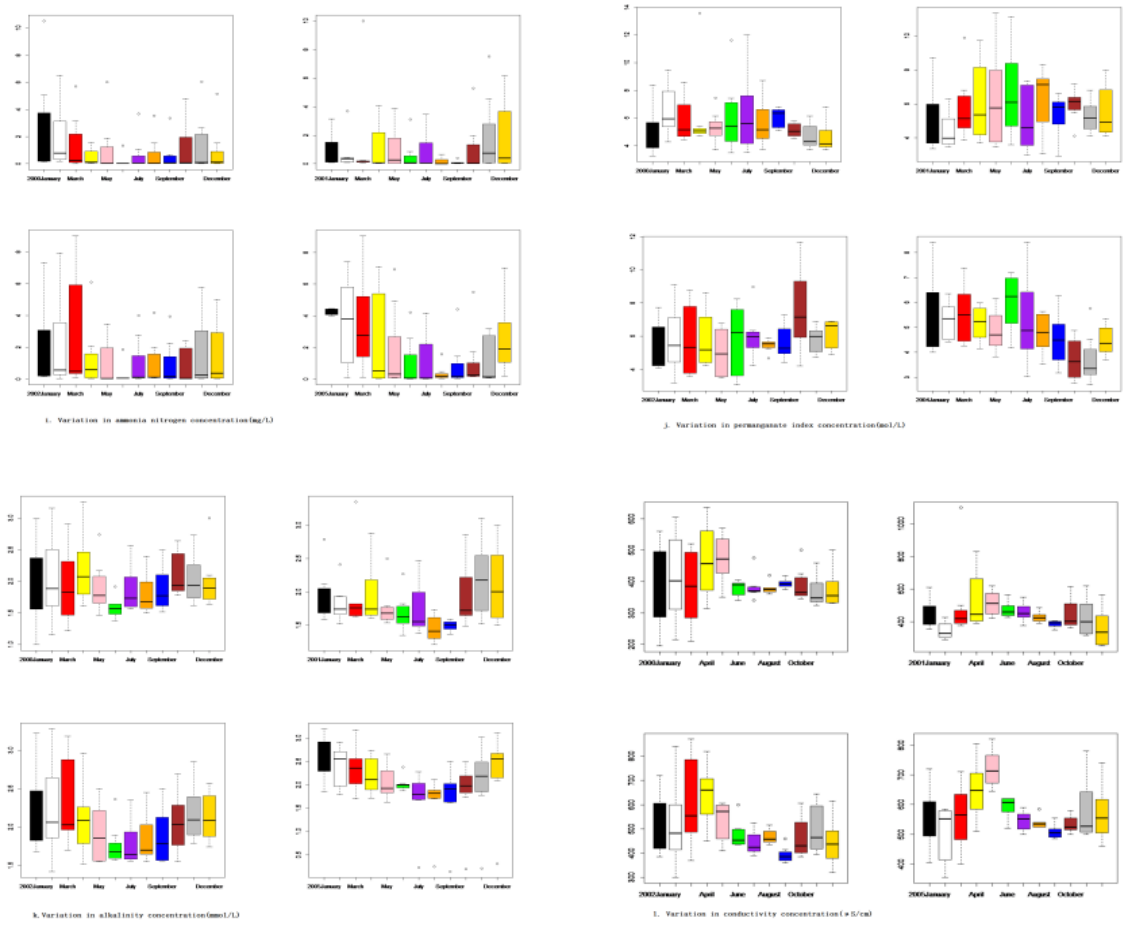


Figure 4-3. *Variation in different parameters' values by study year*

Difference in the water quality index over time

Differences in water quality parameters in Taihu Lake from 2000 to 2005, except for 2003 and 2004

Of the 23 chosen water quality indicators, 4 exhibited little variation over the study period, while 4 exhibited significant variation (*Table 5*). The factors affecting water quality showed similarities as well as differences; the difficulty of assessing water quality can be reduced by focusing only on the factors exhibiting significant changes.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1424114258

Water quality index

The main factors influencing water quality, as calculated by the RBF network algorithm, PCA and SI algorithm, were included in the water quality index formula. This yielded a simple water quality index based on single water quality parameters, and a comprehensive index - water quality identification index of water samples which is the worst water quality index (*Table 6*).

Table 6. *Water quality index*

Model number	Feature(s)	Evaluation result
1	Single index – TP	IV
2	Single index – K	V
3	Single index – SO4	V
4	Single index – SiO	V
5	Single index – TN	V
6	Single index NN	IV
7	Single index – PI	II
8	Comprehensive index - water quality identification index	V

Discussion

The sum-squared error graphs (*Fig. 5*) show the total epochs (iterations) required for the function to converge on TP, SO4, SiO and Ca. The red line indicates the ideal objective function evolution curve, and the blue line represents the objective function curve. Graphs of TP, total SO4, SiO and K were produced using BP-GA. 40 epochs (iterations) were required for TP, SO4, SiO and Ca to converge (*Fig. 5*). In *Figure 6,* the blue line indicates the ideal fitness evolution curve, and the other line indicates the actual fitness evolution curve. The fitness of individuals tends to plateau after 70 generations, indicating that they have reached fitness. *Figure 7* shows the performance of each model. The goal is set to 1e-028. All parameters reach their goals at 5 epochs (iterations).

The error rate of the RBF algorithm used to assess the water quality of Taihu Lake in this study was compared to those of other prediction methods (Random Forest and LASSO) applied to the same data sets. The RBF network data were more accurate than those of the other methods. Under identical conditions and using the same data sets, the error rate was 0.032 for the RBF network, 0.093 for Random Forest, and 0.352 for LASSO. Random Forest is commonly used to make predictions and its results are generally accurate; however, the RBF method used in this study was significantly more accurate (*Table 7*).

Table 7. *RBF model error rate compared with other methods*

Method	RBF network	Random forest	LASSO
Error rate	0.032	0.093	0.352

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1424114258

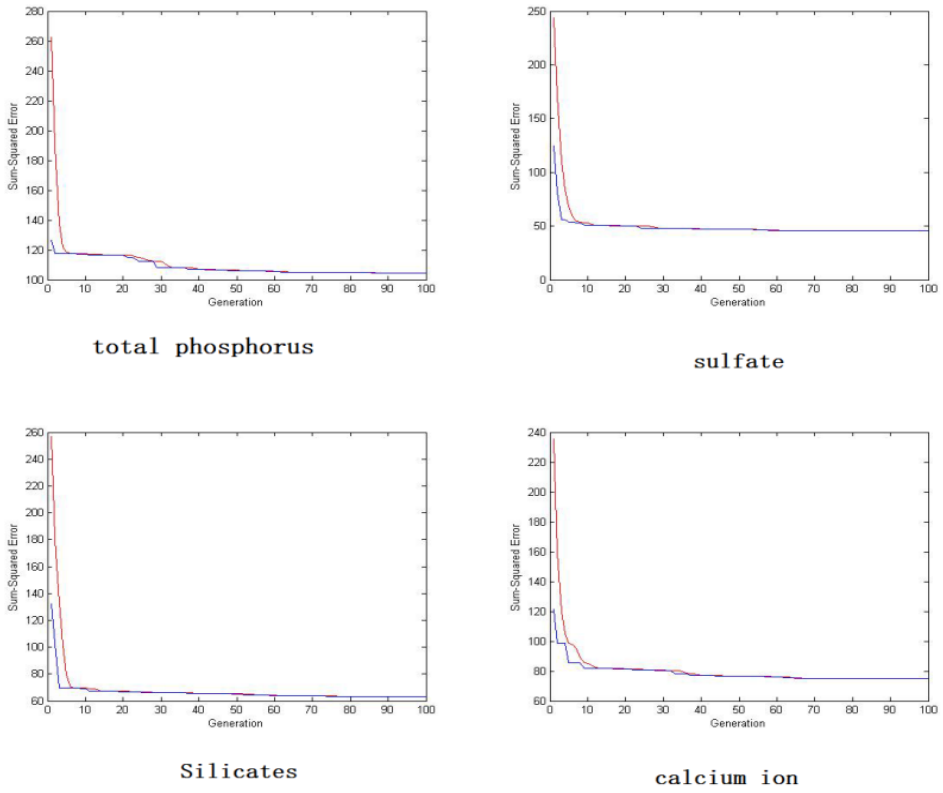


Figure 5. *Sum-squared error of TP, SO4, SiO and Ca affecting water quality*

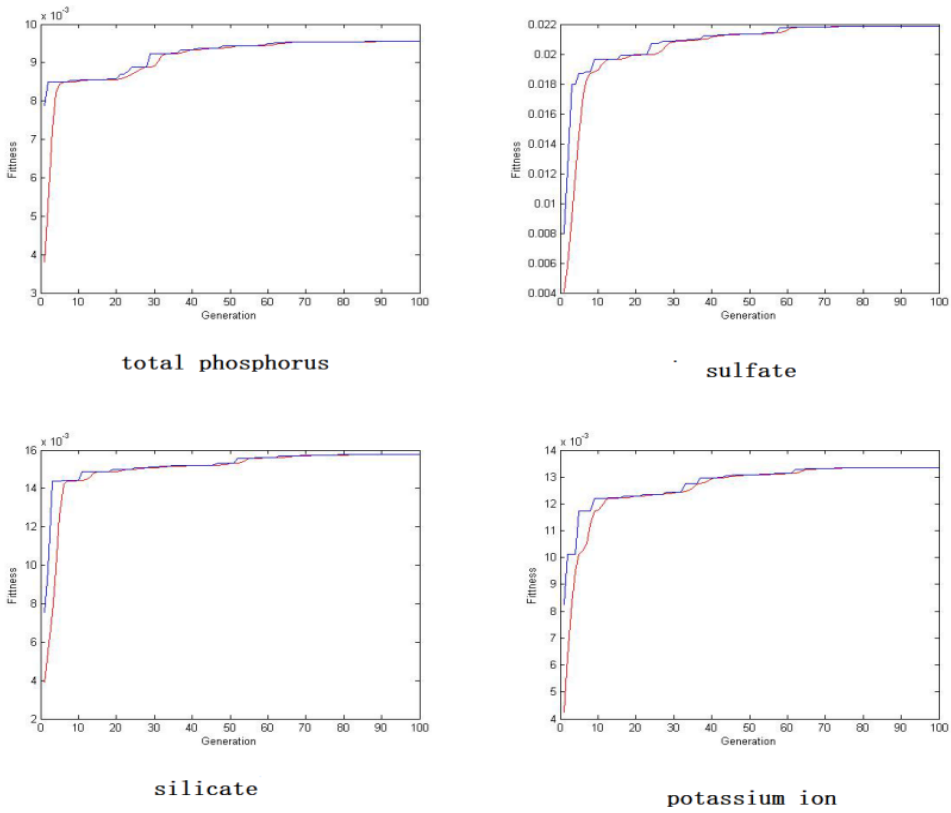


Figure 6. *Fitness of TP, SO4, SiO and Ca affecting water quality*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1424114258

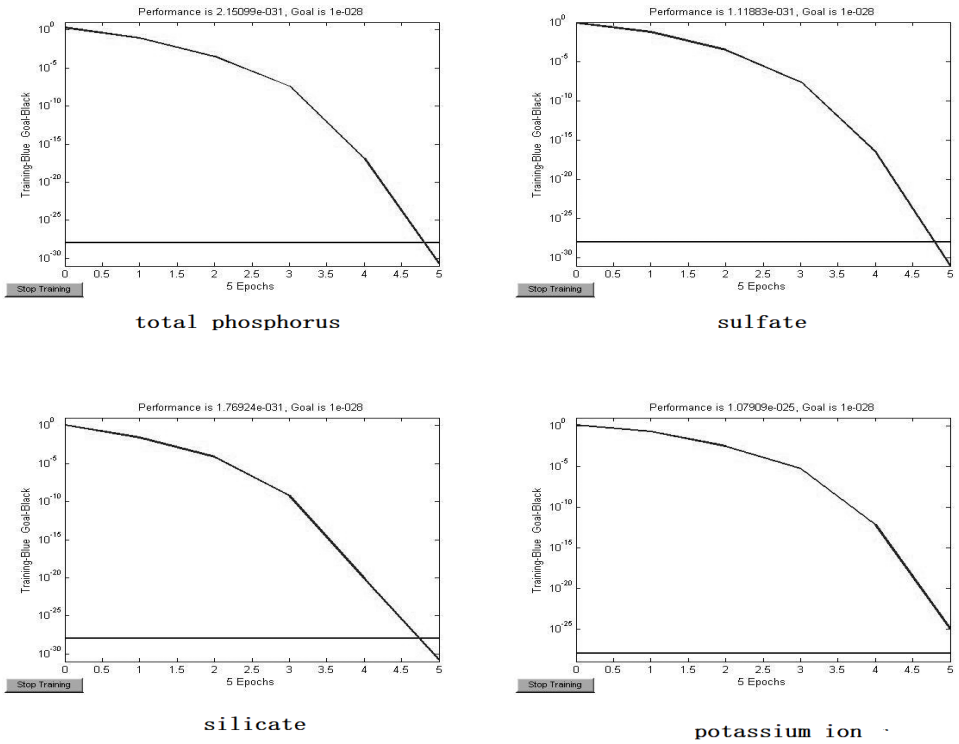


Figure 7. *Performance and goal for TP, SO4, SiO and Ca*

Conclusions

This study proposed a lake water quality assessment method based on an RBF network, SI and PCA. The method is suitable for water quality assessment of large areas over a long period and can identify the most important factors damaging water quality via a comprehensive evaluation. Water quality can be evaluated over time and space and evaluation indexes comprising single and multiple factors can be generated, which avoids over-generalization and better reflects the overall water quality.

The water quality of Taihu Lake for the period 2000 to 2005, except for 2003 and 2004 was between Grade 4 and Grade 5. Thus, water quality satisfied established standards based on our evaluation results. Future water quality control and protection efforts directed toward Taihu Lake should focus on improving management and supervision, as well as water treatment technology; such treatments are expected to improve the water quality of Taihu Lake.

The major environmental issue affecting Taihu Lake concerns control of pollution sources, particularly of Cl-, sulfides, etc. Control of pollution sources is a prerequisite for managing eutrophication in the lake. Ensuring safe drinking water requires point and non-point source pollution control, including a reduction in pollution from sewage. And adjustment of industrial practices and environmental management policies are necessary.

In this study, we investigated the factors affecting the quality of Taihu Lake in the view of data miner. But it is really important to find the very pollution sources, the process in which pollution sources affect the water quality and the cause of pollution sources.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1424114258

Acknowledgements. This work was supported by the National Natural Science Foundation of China (grant number 31371531). The authors acknowledge data support from "Lake-Watershed Science Data Center, National Earth System Science Data Sharing Infrastructure, National Science & Technology Infrastructure of China." (http://lake.geodata.cn).

REFERENCES

[1] Adel, Z. H., Mazen, H., Shady, M. (2018): A comparative study of ANN for predicting nitrate concentration in groundwater wells in the southern area of Gaza Strip. – Applied Artificial Intelligence 32(7-8): 727-744. https://doi.org/10.1080/08839514.2018.1506970.

[2] Alizamir, M., Kisi, O., Zounemat, K. M. (2018): Modelling long-term groundwater fluctuations by extreme learning machine using hydro-climatic data. – Hydrological Sciences Journal/Journal Des Sciences Hydrologiques 63(1): 63-73. https://doi.org/10.1080/02626667.2017.1410891.

[3] Aminravan F., Sadiq R., Hoorfar M., et al. (2013): Enhanced fuzzy evidential reasoning using an optimization approach for water quality monitoring. – IFSA World Congress and NAFIPS Meeting, Edmonton, Canada, pp. 1143-1148. https://doi.org/10.1109/IFSA-NAFIPS.2013.6608561.

[4] Asadollahfardi, G., Zangooei, H., Aria, S. H., et al. (2017): Application of Artificial Neural Networks to Predict Total Dissolved Solids at the Karaj Dam. – Environmental Quality Management 26(3): 55-72. http://doi.org/10.1002/tqem.21493.

[5] Bagheri, M., Mirbagheri, S. A., Kamarkhani, A. M., et al. (2016): Modeling of effluent quality parameters in a submerged membrane bioreactor with simultaneous upward and downward aeration treating municipal wastewater using hybrid models. – Desalination & Water Treatment 57(18): 8068-8089. http://doi.org/10.1080/19443994.2015.1021852.

[6] Brooks, B. W., Lazorchak, J. M., Howard, M. D., et al. (2016): Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems. – Environ. Toxicol. Chem. 35(1): 6. https://doi.org/10.1002/etc.3220.

[7] Cecconello, S. T., Centeno, L. N., Guedes, H. A. S. (2018): Índice de qualidade de água modificado pela análise multivariada: estudo de caso do Arroio Pelotas, RS, Brasil. (Water quality index modified by using multivariate analysis: a case study of Pelotas Stream, RS, Brazil.) – Engenharia Sanitaria e Ambiental 23(5): 973-978. http://dx.doi.org/10.1590/S0103-90162002000100026.

[8] Duan, H., Ma, R., Xu, X., Kong, F., Zhang, S., Kong, W., Hao, J., Shang, L. (2009): Two decode restruction of algal blooms in China's Lake Taihu. – Environ. Sci. Technol. 43(10): 3522-3528. http://pubs.acs.org/doi/abs/10.1021/es8031852.

[9] Dutta, S., Dwivedi, A., Suresh, K. M. (2018): Use of water quality index and multivariate statistical techniques for the assessment of spatial variations in water quality of a small river. – Environmental Monitoring and Assessment 190(12): 718. http://dx.doi.org/10.1007/s10661-018-7100-x.

[10] Drakard, V. F., Lanfranco, S.; Schembri, P. J. (2018): Macroalgal fouling communities as indicators of environmental change: potential applications for water quality monitoring. – Journal of the Marine Biological Association of the United Kingdom 98(7): 1581-1588. https://doi.org/10.1017/S0025315417001102.

[11] Flury, B. (1990): Common Principal Components and Related Multivariate Models. – Journal of the American Statistical Association 85(409): 259-260. https://doi/10.2307/2289562.

[12] Fogg, G. E. (1969): The physiology of an algal nuisance. – Proc. Royal Soc. London, B. 173: 175-189. https://doi.org/10.1098/rspb.1969.0045.

[13] Garrett, J. H. (1994): Where and why artificial RBF networks are applicable in civil engineering. – ASCE, J. Comp. Civ. Engng. Special Issue 8(2): 129-39. https://doi.org/10.1061/(ASCE)0887-3801(1994)8:2(129).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1424114258

[14] Garson, G. D. (1991): Interpreting neural-network connection weights. – Artif. Intell. Expert 6: 47-51. http://doi.org/10.1207/s15327752jpa8502.
[15] Gleeson, B. (2001): Governing for the Environment. – In: Low, N., Gleeson, B. (eds.) The Challenge of Ethical Environmental Governance. Palgrave Macmillan, London, pp. 1-26.
[16] Guilin, L., Linfeng, T., Yuexia, C. (2018): Statistics based study on the seasonal variation of main pollutants in Shahu Lake, Ningxia. – Huanjing Huaxue-Environmental Chemistry 37(9): 2071-2080. https://doi/10.7524/j.issn.0254-6108.2017102602.
[17] Kellner, E., Jason, H., Kirsten, S. (2018): Characterization of sub-watershed-scale stream chemistry regimes in an Appalachian mixed-land-use watershed. – Environmental Monitoring and Assessment 190(10): 586. http://doi/10.1007/s10661-018-6968-9.
[18] Halal, F., Pedrocca, P., Hirose, T., et al. (2014): Remote-sensing based adaptive path planning for an aquatic platform to monitor water quality. – IEEE International Symposium on Robotic and Sensors Environments, Timişoara, Romania, pp.43-48. http://doi.org/.10.1109/ROSE.2014.6952981.
[19] Havens, K. E., Kukushima, T., Xie, P., Iwakuma, T., James, R. T., Takamura, N., Hanazato, T., Yamamoto, T. (2001): Nutrient dynamics and the eutrophication of shallow lakes Kasumigaura (Japan), Donghu (PR China), and Okeechobee (USA). – Environ Pollut. 111(2): 263. https://doi.org/10.1016/s0269-7491(00)00074-9.
[20] Islam, A. R. M. T., Shen, S., Haque, M. A. (2018): Assessing groundwater quality and its sustainability in Joypurhat district of Bangladesh using GIS and multivariate statistical approaches. – Environment Development and Sustainability 20(5): 1935-1959. https://doi.org/10.1007/s10668-017-9971-3.
[21] Jabbar, F. K., Grote, K. (2018): Statistical assessment of nonpoint source pollution in agricultural watersheds in the Lower Grand River watershed, MO, USA. – Environmental science and pollution research international 26(2): 1487-1506. https://doi/10.1007/s11356-018-3682-7.
[22] James, R. T., Havens, K., Zhu, G. W., Qin, B. Q. (2009): Comparative analysis of nutrients, chlorophyll and transparency in two large shallow lakes (Lake Taihu, P. R. China and Lake Okeechobee, USA). – Hydrobiologia 627(1): 211-231.https://doi.org/10.1007/s10750-009-9729-5.
[23] Joliffe, I. (2011): Principal Component Analysis. – Springer, New York, pp. 41-64. https://doi.org/10.1007/0-387-22440-8_7.
[24] Ju, Y., Kaown, D., Lee, K. -K. (2018): A three-pronged approach for identifying source and extent of nitrate contamination in groundwater. – Journal of Soil and Water Conservation 73(5): 493-503. https://doi/10.2489/jswc.73.5.493.
[25] Jun, X., Lingqing, W., Li, D. (2019): Characteristics, sources, water quality and health risk assessment of trace elements in river water and well water in the Chinese Loess Plateau. – Science of the Total Environment 650: 2004-2012. https://doi/ 10.1016/j.scitotenv.2018.09.322.
[26] Junli, W., Zishi, F., Hongxia, Q. (2019): Assessment of eutrophication and water quality in the estuarine area of Lake Wuli, Lake Taihu, China. – Science of The Total Environment 650: 1392-1402. https://doi/10.1016/j.scitotenv.2018.09.137.
[27] Masindi, K., Abiye, T. (2018): Assessment of natural and anthropogenic influences on regional groundwater chemistry in a highly industrialized and urbanized region: a case study of the Vaal River Basin, South Africa. – Environmental Earth Sciences 77(20). https://doi.org/10.1007/s12665-018-7907-3.
[28] Meifang, Z., Huayong, Z., Xuewei, S. (2018): Analyzing the significant environmental factors on the spatial and temporal distribution of water quality utilizing multivariate statistical techniques: a case study in the Balihe Lake, China. – Environmental Science and Pollution Research 25(29): 29418-29432. https://doi/ 10.1007/s11356-018-2943-9.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1424114258

[29] Meng, L., Rui, Z., Jin-sheng, W. (2018): Apportionment and evolution of pollution sources in a typical riverside groundwater resource area using PCA-APCS-MLR model. – Journal of Contaminant Hydrology 28: 70-83. https://doi/10.1016/j.jconhyd.2018.10.005.
[30] Olden, J. D., Jackson, D. A. (2002b). Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. – Ecol. Model 154: 135-150. http://doi.org/10.1016/s0304-3800(02)00064-9.
[31] Omar, R. H., Julia, R., Teresa, M., Fulvio, A. (2018): Physicochemical characterization and sources of the thoracic fraction of road dust in a Latin American megacity. – The Science of the Total Environment 652: 434-446. https://doi.org/10.1016/j.scitotenv.2018.10.214.
[32] Paerl, H. W., Hall, N. S., Calandrino, E. S. (2011): Controlling harmful cyanobacterial blooms in a world experiencing anthropogenic and climatic-induced change. – Sci. Tot. Environ. 409(10): 1739. https://doi.org/10.1016/j.scitotenv.2011.02.001.
[33] Paeral, H. W. (1988): Nuisance phytoplankton blooms in coastal, estuarine, and inland waters. – Limnol. Oceanog. 33: 823-847. https://doi.org/10.4319/lo.1988.33.4_part_2.0823.
[34] Prusty, P., Farooq, S. H., Zimik, H. V., et al. (2018): Assessment of the factors controlling groundwater quality in a coastal aquifer adjacent to the Bay of Bengal, India. – Environ Earth Sci. 77: 762. https://doi.org/10.1007/s12665-018-7943-z.
[35] Putri, M. S. A., Lou, C. H., Syai'in, M. (2018): Long-term river water quality trends and pollution source apportionment in Taiwan. – Water 10(10): 1394. https://doi.org/10.3390/w10101394.
[36] Qin, B., Wu, Q., Gao, J. (2002): Water environmental issues in Taihu Lake of China: problems, causes and management. – J. Nat. Res. 17(2): 221-228. https://doi/10.3321/j.issn: 1000-3037.2002.02.015.
[37] Qin, B. Q. (1999): Hydrodynamics of Lake Taihu, China. – Hydrobiologia 28(8): 669-673. https://www.jstor.org/stable/4314980.
[38] Qin, B. Q. (2008): Lake Taihu, China, Dynamics and Environmental Change. – Springer, New York. https://doi/10.1007/978-1-4020-8555-0.
[39] Tong, Q., Pingheng, Y., Groves, C. (2018): Natural and anthropogenic factors affecting geochemistry of the Jialing and Yangtze Rivers in urban Chongqing, SW China. – Applied Geochemistry 98: 448-458.https://doi.org/10.1016/j.apgeochem.2018.10.009.
[40] Richards, R. P., Baker, D. B., Crumrine, J. P., Stearns, A. M. (2010): Unusually large loads in 2007 from the Maumee and Sandusky Rivers, tributaries to Lake Erie. – J. Soil Water Conserv.,65(6): 450-462. http://doi.org/10.2489/jswc.65.6.450.
[41] Roveda, J. A. F., Arashiro, L. T., Roveda, S. R. M. M., et al. (2013): Fuzzy index for public supply water quality. – IFSA World Congress and NAFIPS Meeting, Edmonton, Canada, pp.1155-1159. http://doi.org/10.1109/IFSA-NAFIPS.2013.6608563.
[42] Scavia, D., Allan, J. D., Arend, K. K., et al. (2014): Assessing and addressing the re-eutrophication of Lake Erie: Central basin hypoxia. – Journal of Great Lakes Research 40(2): 226-246. http://dx.doi.org/10.1016/j.jglr.2014.02.004.
[43] Skowron, P., Skowronska, M., Bronowicka, M. U. (2018): Anthropogenic sources of potassium in surface water: the case study of the Bystrzyca river catchment, Poland. – Agriculture Ecosystems & Environment 265: 454-460. https://doi.org/10.1016/j.agee.2018.07.006.
[44] Shengli, Z., Zheng, W., Tianyi, C. (2018): Transcriptomic analysis of zebrafish (Danio rerio) embryos to assess integrated biotoxicity of Xitiaoxi River waters. – Environmental Pollution 242: 42-53. https://doi// 10.1016/j.envpol.2018.06.060.
[45] Shiau, J. T., Hsu, H. T. (2016): Suitability of ANN-based daily streamflow extension models: a case study of Gaoping River basin, Taiwan. – Water Resources Management 30(4): 1499-1513. http://doi.org/10.1007/s11269-016-1235-8.
[46] Sun, Q., Sun, F., Xia, X., et al. (2018): The comparison of ultrasound-assisted immersion freezing, air freezing and immersion freezing on the muscle quality and physicochemical

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1424114258

properties of common carp (Cyprinus carpio) during freezing storage. – Ultrasonics Sonochemistry 51: 281-291. http://dx.doi.org/10.1016/j.ultsonch.2018.10.006.

[47] Taufiq, A., Effendi, A. J., Iskandar, I. (2019): Controlling factors and driving mechanisms of nitrate contamination in groundwater system of Bandung Basin, Indonesia, deduced by combined use of stable isotope ratios, CFC age dating, and socioeconomic parameters. – Water Research 148: 292-305. https://doi.org/10.1016/j.watres.2018.10.049.

[48] Utah Department of Environmental Quality (2016): Potential health risks force closure of Utah Lake from harmful Algal bloom--Lab tests confirms a high probability of health risks. – https://deq.utah.gov/harmful-algal-blooms/harmful-algal-blooms-home/.

[49] Wang, J., Fu, Z., Qiao, H., et al. (2019): Assessment of eutrophication and water quality in the estuarine area of Lake Wuli, Lake Taihu, China. – Science of The Total Environment 650: 1392-1402. https://doi/10.1016/j.scitotenv.2018.09.137.

[50] Wu, W., Du, W., Zhong, J. (2016): Using Radial Basis Function Neural Networks to identify river water data parameters. – Automatic Control and Computer Sciences 50(4): 285-292. http://doi.org/.10.3103/s0146411616040088.

[51] Xiao, J., Wang, L., Deng, L., et al. (2019): Characteristics, sources, water quality and health risk assessment of trace elements in river water and well water in the Chinese Loess Plateau. – Science of the Total Environment 650: 2004-2012. https://doi/ 10.1016/j.scitotenv.2018.09.322.

[52] Xiaohu, W., Jian, L., Jun, W., et al. (2019): Influence of coastal groundwater salinization on the distribution and risks of heavy metals. – Science of The Total Environment 652(20): 267-277. https://doi.org/10.1016/j.scitotenv.2018.10.250.

[53] Zhaoxue, Z., Yi, L., Haipu, L., et al. (2018): Assessment of heavy metal contamination, distribution and source identification in the sediments from the Zijiang River, China. – Science of The Total Environment 645: 235-243. https://10.1016/j.scitotenv.2018.07.026.

[54] Zheng, J. F., Jiao, J. D., Zhang, S., et al. (2017): An optimization model for water quantity and quality integrated management of an urban lake in a water deficient city. – Urban Water Journal 14(3): 1-8. http://doi.org/10.1080/1573062X.2017.1301500.

[55] Zhi-Qiang, Y., Hiroki, A., Kei, N. (2018): Hydrogeochemical evolution of groundwater in a Quaternary sediment and Cretaceous sandstone unconfined aquifer in Northwestern China. – Environmental Earth Sciences 77: 18. https://doi/ 10.1007/s12665-018-7816-5.

[56] Zounemat, K. M., Özgür Kişi, A. J., et al. (2016): Evaluation of data driven models for river suspended sediment concentration modeling. – Journal of Hydrology 535: 457-472. http://doi.org/10.1016/j.jhydrol.2016.02.012.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1424114258