J. Limnol., 2015; 74(2): 234-247
DOI: 10.4081/jlimnol.2014.906

ORIGINAL ARTICLE

Using integrated multivariate statistics to assess the hydrochemistry of surface water quality, Lake Taihu basin, China

Xiangyu MU,[1] James C. BROWER,[1] Donald I. SIEGEL,[1*] Anthony J. FIORENTINO II,[1] Shuqing AN,[2] Ying CAI,[2] Delin XU,[2] Hao JIANG[2]

[1]Department of Earth Sciences, Syracuse University, Syracuse, NY 13244, USA; [2]Department of Life Science, Nanjing University, Nanjing 210093, China
*Corresponding author: disiegel@syr.edu

ABSTRACT

We investigated the hydrochemical setting of Lake Taihu (eastern China) to determine how different land use types influence the variability of surface water chemistry in different water sources to the lake. Major water types within the watershed range from calium-magnesium bicaxrbonate water, typical of relatively pristine water, highly ocntaminated water characterized by more sulfate, sodium, chloride and nutrients. Principal components analysis produced three principal components that explained 78% of the variance in the water quality and reflect three major types of water chemistry. Agricultural land use is associated with greater concentrations of nutrients; urban areas with high concentrations of sodium, chloride, sulfate, fluoride and potassium; and natural weathering with calcium, magnesium and bicarbonate. Discriminant analysis and hierarchical cluster analysis produce complementary and similar results. Broadly speaking, future remediation to reduce nutrient loadings to the lake or industrial contamination can now be focused on specific land use practices, which are readily identifiable by using statistics in conjunction with GIS.

Key words: Principal component analysis, discriminant analysis, hierarchical cluster analysis, geographical information system (gis), surface water quality, water-rock interaction, land use types.

Received: November 2013. Accepted: September 2014.

INTRODUCTION

The demand for freshwater needed for increasing crop production and industrialization occurs almost everywhere in China and these conflicting needs have led to widespread water contamination. Taihu Lake (eastern China) notably suffers periodic hyper-eutrophication and drinking water deterioration, from heavy nutrient loadings from all of these sources This pollution has led to shortages of freshwater for the City of Wuxi and other nearby cities. Taihu Lake, the third largest freshwater body in China, has historically been considered a cultural treasure of China, and has supported long-term fisheries. However, remediation of its water cannot be effectively done without first characterizing the broad nature of the non-point source pollution.

Both natural and anthropogenic activities influence the dissolved chemical composition of surface water (Subramani *et al*., 2005; Shi *et al*., 2011; Stonestrom *et al*., 2009). Natural factors include the mineral solubility in bedrock and soils, geomorphology, and climate. Human modification of unaltered natural landscapes due to agricultural and urban uses results in losses of forested and wetland areas, increasing storm runoff and anthropogenic chemical and waste water inputs (Wayland *et al*., 2003; Fitzpatrick *et al*., 2007; Cortecci *et al*., 2009; Tang *et al*., 2005). The sources for nutrients and other contaminants can be evaluated by analysing the chemical difference between the various surface waters (Deocampo, 2004; Poinke and DeWalle, 1994; Mason *et al*., 1999) by using a broad range of approaches including descriptive graphics, statistical methods, and isotopic analysis (Hounslow, 1995; Hem, 1992).

Of these a combination of descriptive graphics and empirical multivariate statistical analyses prove most useful within the context of characterizing contamination in watersheds with complicated land use types and histories (Barros Grace *et al*., 2008; Belkhiri *et al*., 2010; Panno *et al*., 2006; Reeve *et al*., 1996), particularly for data sets on surface water chemistry that are widely distributed in space with little temporal record. For example, principal components analysis (PCA), hierarchical cluster analysis (HCA) and discriminant analysis (DA) have been extensively used in environmental hydrogeology to characterize differences in surface water chemistry, including contamination in large watersheds (Cloutier *et al*., 2008; Farnham *et al*., 2003; Alberto *et al*., 2001; Cortecci *et al*., 2009; Belkhiri *et al*., 2010). Graphical approaches such as Piper diagrams (Piper, 1944) geochemically classify surface water quality into hydro-geochemical facies of water (Back, 1966; Frey *et al*., 2007). Coupled with information from geographic information systems (GIS), the combination of multivariate statistics and geochemical plotting approaches can relate

various land use types to water quality at the watershed scale (Howarth *et al*., 2010; Meador and Goldstein, 2003; Fitzpatrick *et al*., 2007). This combination has been used to evaluate changes of water chemistry resulting from complex changes and differences in land use patterns (De Carlo *et al*., 2004; Steuer *et al*., 1997).

We assess herein the hydrochemistry of Lake Taihu, the third largest fresh water lake in China within a broad context to place its serious water quality deterioration in regional context by using multivariate statistics and graphical methods on a large synoptic chemical data set from Lake Taihu to better understand the complex interdependence of major ions, nutrients, and trace metals and their sources within the lake and its watershed.

Study area

Lake Taihu has a surface area of 2340 km^2 within a catchment area of 36,500 km^2, and lies in the east China Taihu plain on the south side of the Yangtze River delta (Yuan *et al*., 2010). The Lake Taihu basin is located in a subtropical monsoonal climate area, with an annual average temperature of 15°C to 17°C, and annual average precipitation of about 1180 mm. Lake Taihu is situated at the center of a geomorphic depression. The mean water level of Lake Taihu lies about 3 m above sea level, and the lake's maximum and mean depth are only 2.6 m and 1.9 m, respectively. Water in the lake has a residence time of about 300 days (Cao *et al*., 2012).

Upstream water from the western half of the basin flows in streams from low hills (6~12 m in elevation) to the lake and then diffuses northward to the Yangtze River and southeastward to the China East Sea through a complex network of canals and manmade conveyances. Toward the east, the lake basin consists of a dense water web of canals and flat landforms; these form a complex hydrosystem that is interlaced with over a hundred streams and canals and dotted with small depression lakes and ponds of different sizes, including Lakes Tao and Ge (Fig. 1) (Yu *et al*., 2012).

Limnological history

Lake Taihu was one of the many lakes that formed on the Taihu alluvial plain, which consists of loess deposited during the Last Glacial Maximum Period (Sun and Huang, 1993). The limnological history of Lake Taihu remains uncertain after many decades of study. Shatter cones, shock metamorphosed quartz and microtekties in Devonian age rocks at the lake suggest that the unusual circular shape of the lake basin reflects an impact crater that could have formed many millions of years ago (Wang *et al*., 2002) to as recently as 4500 years ago (Xie *et al*., 2008). Sediment cores show that wind blown loess covered the landscape about 16,000 years ago during continental glaciation. After the loess was deposited and compacted, the landscape could have become waterlogged to create Lake Taihu and other lakes (Zhang *et al*., 2004; Sun *et al*., 1987). But, microfossils in Lake Tai sediments show the landscape became inundated with sea- water from the East China Sea as sea level rose between 10 and 15 thousand years ago during deglaciation. About 5000 years ago, the marine water in the Tai region became brackish, and some workers hypothesize that sediment deposited as spits and barrier bars at the mouths of the ancient Yangtze and other rivers coalesced to make a lagoon, which was largely removed from the ocean. Water flowing into the lagoon could then have displaced the brackish water to form the fresh water lake seen today (Qin *et al*., 2007; Chen *et al*., 1959).

Overall water quality

The Lake Taihu watershed now mostly is used for agriculture and partial aquaculture (~51%) and urban land use (~23%), with only ~13% remaining as forested lands and barren areas. Landsat images from 1990 to 2008, show agricultural land within the watershed has decreased from 64% to 51% whereas urban areas have increased from 10% to 23% (Xu *et al*., 2009).

Water quality in Lake Taihu was good during the 1960's. However, by the 1981, water quality had deteriorated considerably. For example, total inorganic nitrogen (TIN) in the lake water increased 18 times higher than in 1960 (Qin, 2008). Agriculture and urban areas contribute most nutrients to the lake as well as heavy metals (Wilhelm *et al*., 2011). The combination of agricultural, industrial and human waste disposal has caused enhanced eutrophication and frequent phytoplankton blooms (Lin *et al*., 2006). In the early 1970's, blue-green algal blooms first appeared in Wuli Bay of Wuxi City, and subsequently their scale and frequency increased greatly. In the 2000's, algal blooms occurred more than five times every year between May and October in the north-western part of Lake Taihu including Meiliang Bay and even the center of Lake Taihu has also suffered green-blue algal blooms which adversely affected water quality and its water supply to the local cities (Lin *et al*., 2006; Yuan *et al*., 2010; Zeng *et al*., 2012).

Our study focuses on the high plains region south of the Yangtze River and the north-western part of the Lake Taihu watershed, which has experienced the worst water quality and most severe blue algal blooms. There are two major cities within the study site: Changzhou City and Wuxi City whereas the remaining areas are mostly used for agriculture, aquaculture and residences (Fig. 1).

METHODS

Sample and data collection

We synoptically sampled surface water in the northwestern area of Lake Taihu Basin from May 22 to June 10,

2010 and during September 12 to 19, 2010 at 87 sampling sites on rivers and canals and 23 sampling sites in the lake. These sites were then plotted on a Digital Elevation Model (cell size 100 m ×100 m at 1:100,000 scale) and layered on a land-use map at 1:100,000 scale (both accessed from the Institute of Remote Sensing Application Chinese Academy of Sciences). We used ArcGIS 10.1 to analyse the land use maps and extract DEM map information (Fig. 2; Supplementary Tabs. 1 and 2). Agricultural land use constitutes about ~59% of the study area, urban and industry area ~23% coverage, and forestry and water source conservation area about ~9%. The remainder of the area consists of sur-

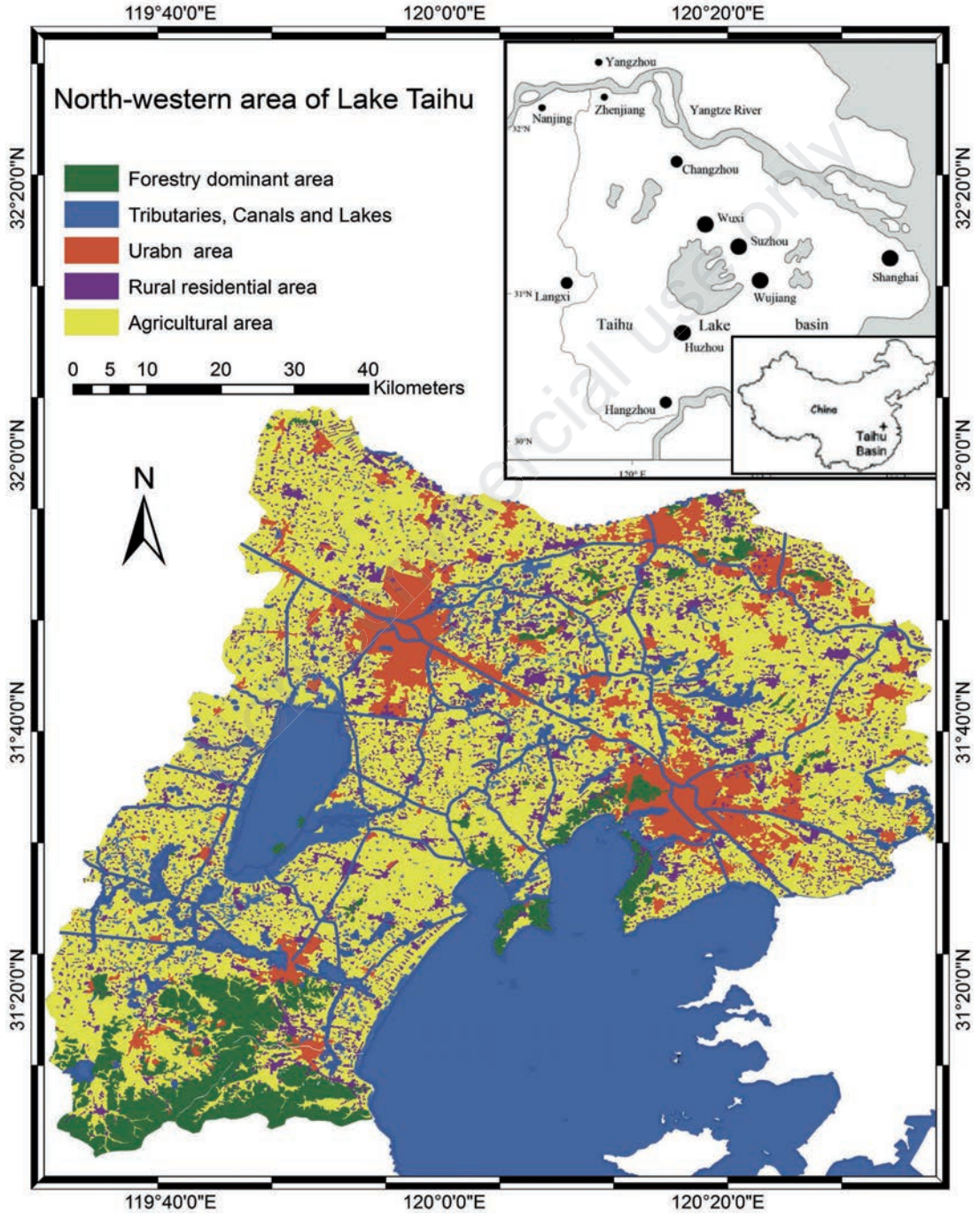

Fig. 1. Study area and land use in the northwestern part of Lake Taihu basin.

face water and undeveloped land. If the surface water depth at sampling sites was less than 2 m, we collected water sample at depths of 0.5 m and 1.5 m depth and mixed them to make a single composite sample per site; if water depth ranged from 2~3 m, water samples were collected at depths of 0.5 m, 1.5 m, 2.5 m and then averaged.

Chemical analysis

Immediately after sampling, we measured suspended sediment (SS), temperature, electric conductivity (EC), dissolved oxygen (DO), pH, specific conductance (SC), total dissolved solutes (TDS), water depth and turbidity onsite using a multi-parameter water quality meter (CyberScan

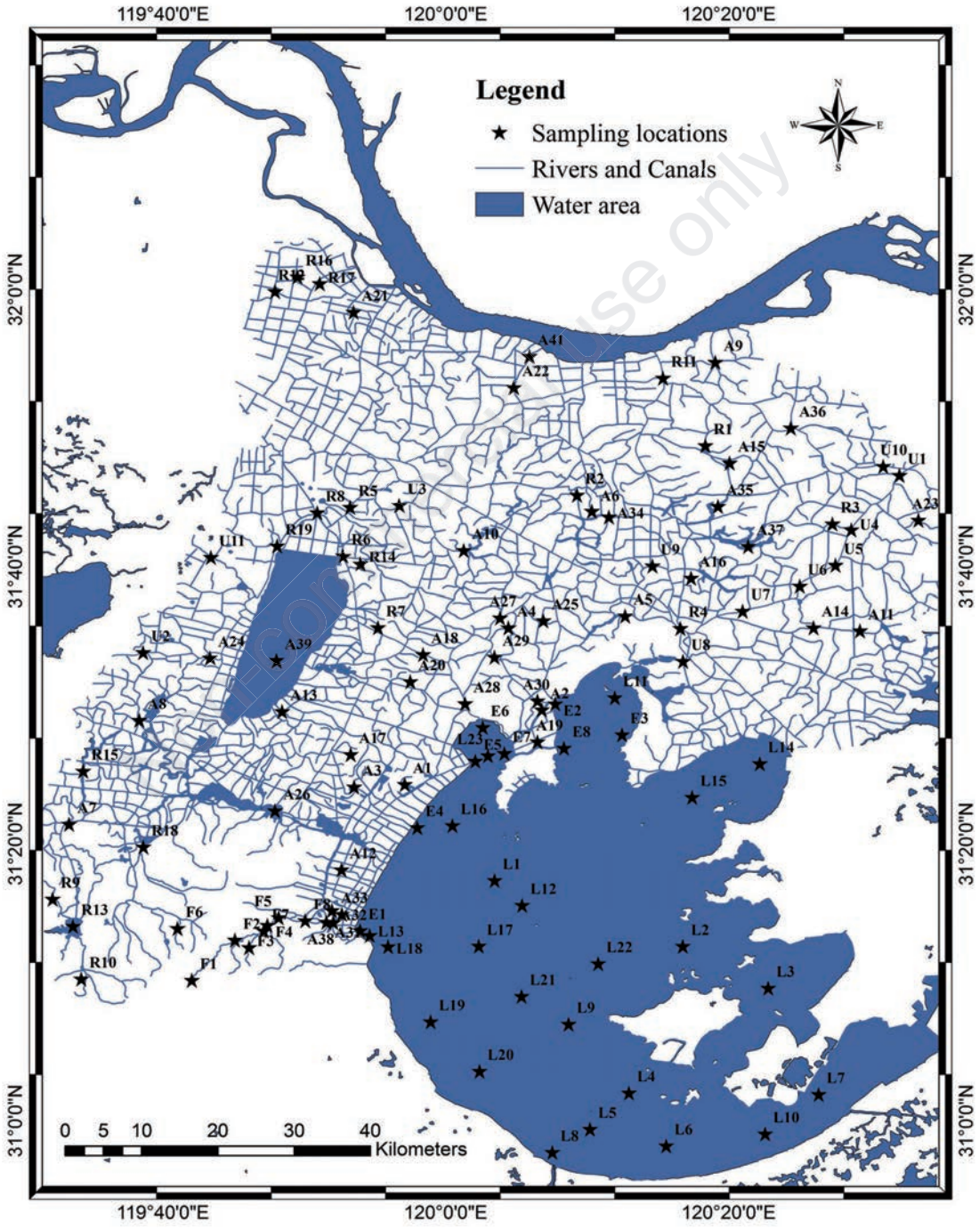

Fig. 2. Sampling locations for lake water, streams, and canals.

PCD6500). We then filtered all samples through 0.45 μm Millipore filters by vacuum filtration to remove suspended sediments prior to subsequent analysis. We placed the filtered water samples into two 100 ml polypropylene containers and one was immediately acidified to pH <2 by adding ultrapure nitric acid for later cation analysis. All samples were stored at temperature of <4°C until analyses.

Samples were analysed for Ca, Mg, Fe, Cr, Sr, Na, K, Mn, Si, F, Cl, Br, NO_3, SO_4, PO_4, TN and TP. Bicarbonate was calculated by charge balance. We measured dissolved metals using an Elan DRC-e Inductively Coupled Plasma Mass Spectrophotometer (ICP-MS) at the State University of New York College of Science and Forestry (Syracuse) and measured anions using a Dionex Ion Chromatograph (IC) at the Heroy Hydrogeology Laboratory of Syracuse University. The TN and TP were measured in the chemical analysis laboratory of the Biology Department of Nanjing University. By alkaline-persulfate digestion and acid persulfate methods respectively. Standards, blanks and replicate samples were run every 10 samples; precision and error generally was within 5% for all analyses. Specific conductance measured in the field was corrected for water temperature.

Statistical data analysis

We used principal components analysis (PCA), hierarchical cluster analysis (HCA), and discriminant analysis (DA) for multiple groups to analyse the surface water chemistry data. Davis provides a clear explanation of PCA as used in this paper (Davis, 2002). The original data were initially transformed in two steps. First, 1.0 was added to the data in order to eliminate values of zero. Note that the addition of a constant does not affect the structure of the data because this procedure constitutes a simple translation which can be reversed at later stages in the analysis. Secondly, these data were then transformed to their logarithms (base 10), because the logarithms are linearly distributed whereas the original data were mostly curvilinear. Finally, the transformed data were converted to standardized z-scores with means of zero and standard deviations of unity such that all variables are expressed in identical units.

We then calculated a Pearson product moment correlation matrix for the major chemical variables. Eigenvalues and their corresponding unrotated principal components were then extracted from the correlation matrix. The coefficients or loadings for the variables in the unrotated eigenvetors or principal components were then normalized to their corresponding eigenvalues and the loadings display the major correlation patterns in the data. The principal components are orthogonal to one another and so they extract independent sources of variance or correlation. The eigenvalues represent the variances of the orthogonal projections of the samples onto the unrotated principal components and the amount of variance involved in each principal component. Inspection of a scree plot of the eigenvalues *versus* their rank order discloses the number of eigenvalues and principal components that are required to explain the major sources of correlation within the variables of the data set.

The retained principal components were then rotated by the varimax criterion. Varimax is an orthogonal rotation that maximizes or minimizes the coefficients or loadings of the variables onto the principal components. Varimax rotation is a common orthogonal rotation that can identify clusters of variables, which aids in interpretation. The rotated principal component scores show the relations of the samples to the sources of correlation in each principal component, and these scores were computed by the regression weight method. Contour plots of the principal component scores reveal the spatial distribution of the various chemical signatures in the data.

Ward's sum of squares method was used for cluster analysis of the samples (Davis JC, 2002; Ward, 1963; Johnson and Wichern, 2007). The basic property of Ward's technique is that it minimizes the error sum of squares or error variance at each step of clustering. Inasmuch as the error sum of squares or the error variance is a squared criterion, the method tends to produce tight clusters resulting in a clearer interpretation. Ordination techniques such as principal components and clustering algorithms like Ward's sum of squares are complementary. Ordination methods preserve large-scale patterns at the expense of small scale distortion. On the other hand, cluster analysis is effective at grouping similar items together but overall patterns may be lost or obscured.

The water samples from Lake Taihu were obtained from areas that are characterized by a variety of land uses. Initially, the samples were assigned to groups according to the major type of land use in the area immediately surrounding each sampling site. The chemical differences between the multivariate means of the waters from the different land use types were then investigated by discriminant analysis for multiple groups (Davis, 2002; Hastie and Tibshirani, 1996; Johnson and Wichern 2007; McLachlan, 2004). The Wilk's Lambda criterion provided the basic significance test. Wilk's Lambda equals the ratio of the determinants of the within groups sum of squares and cross products matrix (W), and the total sum of squares and cross products matrix (T). Smaller ratios indicate greater differences among the groups. The between groups sum of squares and cross products matrix (B), measures the differences between the means of the various groups and is derived by subtracting T from W.

Chi-square and F-ratio statistics test the statistical significance of Wilk's Lambda. If the differences between the various groups are insignificant, the calculations are terminated and the groups are treated as homogeneous. Conversely, if significant differences occur between the

multivariate means of the groups, the analysis continues. The multiple group discriminant functions or canonical variables are given by the eigenvalues and eigenvectors of the matrix $W^{-1}B$. This matrix is asymmetrical so its eigenvectors may be correlated. If so, the eigenvectors are rotated so that they are orthogonal and uncorrelated, and their coefficients are standardized. Plots of the scores for the canonical variables display the main relationships between the samples in the different land use groups.

A classification or identification matrix reveals the nature of overlap between the water samples originally placed in the various groups of land use. Each water sample is then identified or classified into a land use group based on its minimum Mahlanobis squared distance from the multivariate mean of each group. The within groups covariance matrix for the statistics is taken from the original data rather than the canonical variables. The assignments of the water samples to the closest group assume equal probabilities of *a priori* group memberships. A matrix of Mahlanobis squared distances is also computed for the means of all of the land use groups, and its values demonstrate the statistical significance of the differences between the pairs of groups.

Our analytical strategy is straight forward. The rotated principal components and the cluster analysis are intended to explore the main structure of the data in the form of relationships within and between the chemical variables and the samples. The probabilistic significance of these relationships is then subsequently tested by discriminant analysis on the groups of samples that are characterized by different chemical compositions.

The computer programs for the various statistical analyses consist of: Statistical Package for the Social Sciences (SPSS) ver. 13.0 for correlations, principal components and the discriminant analysis along with a computer program for discriminant analysis classifications written in the APL programming language by J.C. Brower, ArcGIS 10.1 software for contour plots of rotated principal component scores, and the Matlab (2009a) program for hierarchical cluster analysis.

Graphical geochemical methods

We used Piper Diagrams (Hounslow, 1995) to independently describe surface water chemical composition from our statistical characterization. A Piper diagram consists of two ternary diagrams showing the percentages of equivalent charges for major dissolved cations and major anions, and a central diamond- shaped figure between them. Points on the anion and cation ternary diagrams are projected upward to where they intersect on the diamond.

Four major water quality types occur as end-members of possible hydrochemical facies: calcium-bicarbonate (right hand apex of diamond field), calcium-chloride or calcium-sulfate (upper apex), sodium-chloride (right hand apex) and sodium-bicarbonate facies (lower apex). Most waters actually reflect either mixtures of these end members or geochemical evolution from one end member to another. In these cases, data points would plot on linear trends from one end-member to the next. For example, were sea water (mostly Na and Cl) mixing with water associated with calcium carbonate (mostly Ca and HCO3 in water), the data points would plot on linear trends from; the Ca corner of the lefthand triangular diagram to the Na corner, the HCO3 to Cl corners on the lower triangular diagram, and from the left side to the right side of the diamond between the triangular figures.

RESULTS AND DISCUSSION

Major solutes composition

The major solute compositions of Lake Taihu water and north-western watershed surface water are shown in Tab. 1. We ordered the major cations in a sequence from the highest to the lowest average concentrations: Na^+, Ca^{2+}, Mg^{2+}, K^+. Their mean values and standard deviations are: 8.7±2.34 mg L^{-1} (Na); 48.7±30.9 mg L^{-1} (Ca), 42.3±10.5 mg L^{-1} (Mg), 5.00±2.2 mg L^{-1} (K). For major anions, mean values and standard deviations are: 159.3±65.1 mg L^{-1} (HCO_3), 77.1C±47.9 mg L^{-1} (SO_4), 53.2±29.3 mg L^{-1} (Cl). The average TDS value 361 mg L^{-1}, ranging from 95 to 736 mg L^{-1}. Generally speaking, the north-western watershed surface water of Taihu Basin can be described by the cations Ca^{2+} and Na^+ and anions HCO_3^-, Cl^- and SO_4^{2-}.

Graphical representation of hydrochemical data

The results of plotting the geochemical data on a Piper diagram are shown in Fig. 3. The Piper plot shows that Ca^{2+}, Na^+, Mg^{2+} and the anionic species bicarbonate and chloride largely control the surface-water classification. The majority of water samples belong to the Ca-Mg-HCO_3 hydrochemical facies. Lake Taihu water ranges from Ca-Mg-HCO_3 to Ca-Mg-Cl_2 types. The water samples collected from tributary mouths discharging into Lake Taihu are dominated by Ca-Mg-HCO_3 type waters. Water samples collected in headwater streams are scattered on the diamonds, but most lie within the Ca-Mg-HCO_3 and Ca-Mg-Cl_2 hydrochemical facies. Waters in a few rivers were Na-Cl and Na-HCO_3 types. However, none of the data fell on a trend between seawater and other water types. Such a trend would end near the extreme right hand apex of the diamond field where seawater would plot. We saw no evidence for seawater intrusion in our data set.

Principal component analysis

The positive correlations among Ca, Mg, K, Na, Sr, Cl, F, SO_4, HCO_3, NO_3, TN and TP, comprise a major theme in the data. Tab. 2 and the scree plot of Fig. 4 sug-

Tab. 1. Major ionic composition of 110 surface water samples (in mg L^{-1}). Soluted expressed in standard chemical nomenclature.

Chemical composition	Minimum	Maximum	Mean	Std. deviation
Ca	20.88	70.15	42.27	10.47
K	0.83	16.71	5.04	2.19
Mg	3.00	16.28	8.68	2.34
Mn	ND	0.48	0.07	0.09
Na	4.40	167.46	48.68	30.85
Si	0.76	6.70	2.77	1.03
Sr	0.07	0.32	0.21	0.05
F	0.05	1.06	0.52	0.20
Cl	4.67	172.66	53.22	29.30
Br	0.03	0.61	0.16	0.10
NO_3	0.22	45.93	8.24	7.39
SO_4	7.34	298.65	77.10	47.91
HCO_3	22.07	323.50	159.27	65.12
TN	0.46	8.77	2.52	1.36
TP	0.02	66.77	0.94	6.27
DO	0.69	11.88	5.74	2.37
SC (ms cm^{-1})	0.15	1.15	0.56	0.2
TDS	95	736	361	124
pH (units)	4.73	9.47	7.42	0.80

TN, total nitrogen; TP, total phosphorus; DO, dissolved oxygen; SC, specific conductance; TDS, total dissolved solids.

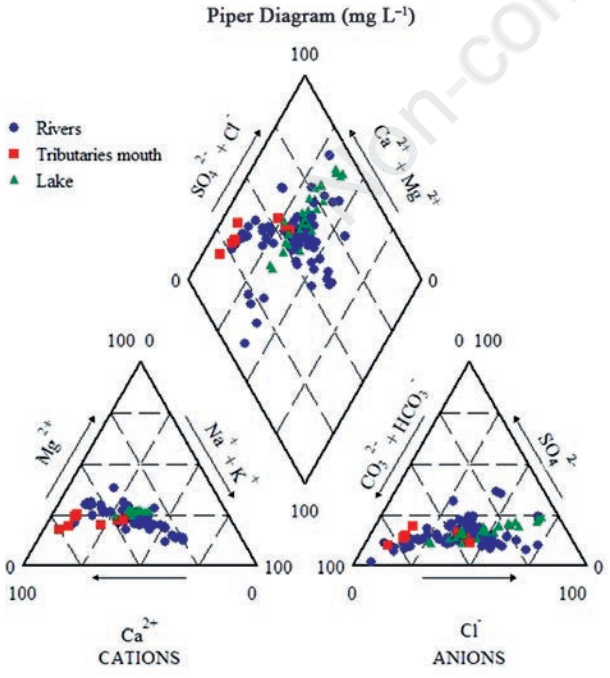

Fig. 3. Piper diagram showing hydrochemical facies of waters in LakeTaihu basin.

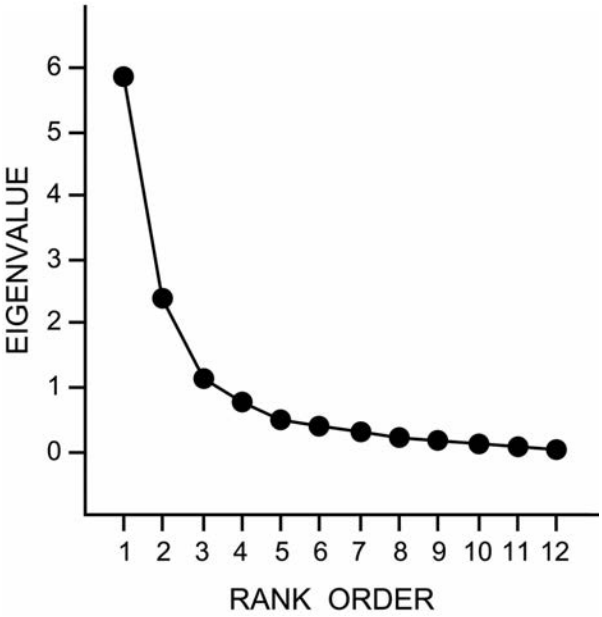

Fig. 4. Plot of eigenvalues *versus* rank order for the geochemistry of Lake Taihu waters.

gests that first three principal components explains the major pattern of correlation in the data, inasmuch as the rate of decline of the eigenvalues greatly decreases after the three largest ones are extracted from the correlation matrix. In addition, only three eigenvalues are greater than 1.0. An eigenvalue less than unity would be associated with less variance than that explained by one of the original variables. Together, the first three principal components account for almost 78% of the variation in the data. Consequently, the first three principal components were rotated to their varimax configuration in Tabs. 2 and 3.

Rotated principal component 1 extracts about 35% of the variance in the correlation matrix and strong positive loadings occur among Na, K, F, Cl, and SO_4. Samples with high principal component scores are characterized by large amounts of Na, K, F, Cl, and SO_4 whereas those with low scores are depleted in these variables. These solutes commonly relate to anthropogenic contamination derived from human activities in the watershed, since evaporate minerals, such as halite and gypsum, do not occur or are rare in the study area.

The second rotated principal component is associated with roughly 28% of the information in the correlation matrix. Strong positive contributions occur among Ca, Mg, Sr and alkalinity. High scores for the second principal component occur for samples that are rich in Ca, Mg, Sr and HCO_3. This component reflects chemistries naturally obtained by dissolution of carbonate rich sediments, common in the Lake Taihu area.

The last rotated principal component explains nearly 16% of the trace of the correlation matrix. The component is dominated by positive loadings for NO_3 and total nitrogen (TN) and total phosphorus (TP), consistent with that expected from discharges of agricultural fertilizer, pesticide, animal waste, septic material, and effluents containing nitrogen and phosphorus from industries and residents. Samples with high scores on the third principal component contain abundant discharges of one or more of these types.

The three dimensional plot of the rotated loadings for the variables of the first principal components shows clusters of variables that can be related to these three groups, namely Na, K, F, Cl, and SO_4 in one sector, Ca, Mg, Sr, and HCO_3 in a second region, and NO_3, TN and TP in a third area (Fig. 5).

Each sample can be assigned to a land use category based on the main type of land use in the area in the immediate vicinity of the sample. The land use data are taken from field notes, areal photographs, and GIS maps of the Lake Taihu region. Six categories of land use have been identified and will be discussed throughout the paper:

i) Agriculture and aquaculture; some small factories are also located in these areas.
ii) Mouths of tributaries into Lake Taihu.

Tab. 2. Data for eigenvalues and variance contributions of principal components.

Principal component	Eigenvalues of correlation matrix	Percent of variance	Cumulative percent of variance
1	5.871	48.927	48.927
2	2.413	20.11	69.037
3	1.124	9.369	78.406
4	0.779	6.491	84.898
5	0.506	4.215	89.113
6	0.413	3.443	92.556
7	0.305	2.539	95.095
8	0.206	1.713	96.808
9	0.165	1.376	98.184
10	0.108	0.903	99.087
11	0.072	0.601	99.687
12	0.038	0.313	100
Sum of rotated loadings	**Percent of variance**	**Cumulative % of variance**	
4.163	34.691	34.691	
3.328	27.733	62.423	
1.918	15.983	78.406	

Tab. 3. Matrix of three principal components rotated to varimax configuration. Underlined numbers show major components.

	Principal component 1	Principal component 2	Principal component 3
Na	<u>0.796</u>	0.420	0.024
K	<u>0.748</u>	0.550	0.065
F	<u>0.908</u>	0.025	-0.058
Cl	<u>0.931</u>	0.147	-0.027
SO_4	<u>0.862</u>	0.151	0.157
Ca	0.177	<u>0.854</u>	0.321
Mg	0.570	<u>0.698</u>	0.039
Sr	0.369	<u>0.801</u>	0.174
HCO_3	0.052	<u>0.886</u>	0.153
NO_3	-0.138	0.309	<u>0.771</u>
TN	0.002	0.255	<u>0.812</u>
TP	0.147	-0.032	<u>0.687</u>
Percent of variance	34.691	27.733	15.983
Cumulative percent of variance	34.691	62.423	78.406

TN, total nitrogen; TP, total phosphorus.

iii) Water conservation and forestry areas.
iv) Lake Taihu basin.
v) Rural residential and village areas.
vi) Urban and industrial areas.

A plot of the rotated principal component scores of these samples on the first two principal components is shown in Fig. 6. Although there is some overlap, the two principal components tend to separate samples with different land use patterns into different regions of the plot. This indicates that the various types of land use are characterized by somewhat different water chemistries. Three different sources of pollution can be identified by the principal components:

i) Sewage water discharged from urban areas and flow of chemical solutes and industrial effluents along with rainfall.
ii) Hydro-geochemical processes due to rock weathering and water-rock interaction and other types of mineral dissolution.
iii) Non-point sources of pollution discharged from agricultural irrigation using fertilizer, human household and animal wastes, and fish farming which could bring nutrients and salts to the watershed along with river flow and rain wash.

Maps of the scores for the samples onto the rotated principal components show the spatial distribution of these correlated sources of the various chemical components (Fig. 7). The map of the scores for the first principal component indicates that samples with the highest scores are concentrated in the northeastern area of Lake Taihu compared to the more western and southern regions of the watershed. These samples are characterized by higher concentrations of Na, K, F, Cl, and SO_4 in and near the dense urban areas of Wuxi, Changzhou, and Suzhou. The Na and Cl can be associated with human wastes. The SO_4 could constitute an oxidation product of sulphide pollution discharged from industry or municipal landfill leachates. Wuxi and Suzhou are highly developed cities with dense populations, different kinds of industries, and significant amounts of tourism.

In contrast, the largest scores for the second rotated principal component are located in the northern and western parts of the watershed. These areas of higher elevations with rolling hills have less development and little urbanization. Since 2002, the Chinese government has delivered Yangtze water to the lake and some of the chemistry in this part of the watershed may reflect this river water. In addition, the underlying sediments mainly consist of carbonates and carbonates are common in the soils.

Three areas with high scores and high concentrations of NO_3, TN, and TP can be identified on the map for the third principal component. In these samples, the dominant land use is a combination of rural agriculture and residential areas, where people, livestock, and aquaculture pro-

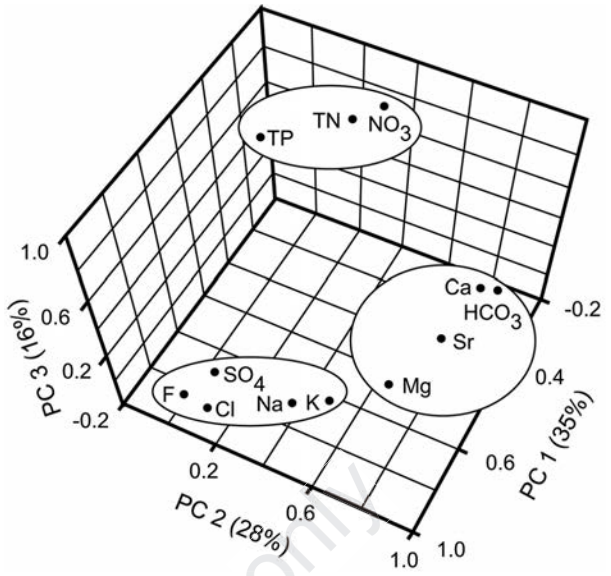

Fig. 5. Graphical representation of principal component scores in three dimensional space showing major grouping of water quality parameters.

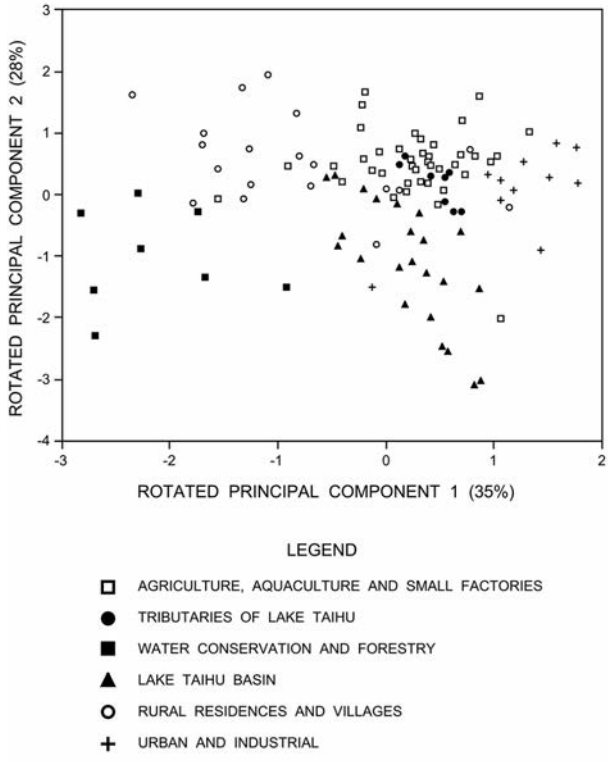

Fig. 6. Rotated principal components 1 and 2.

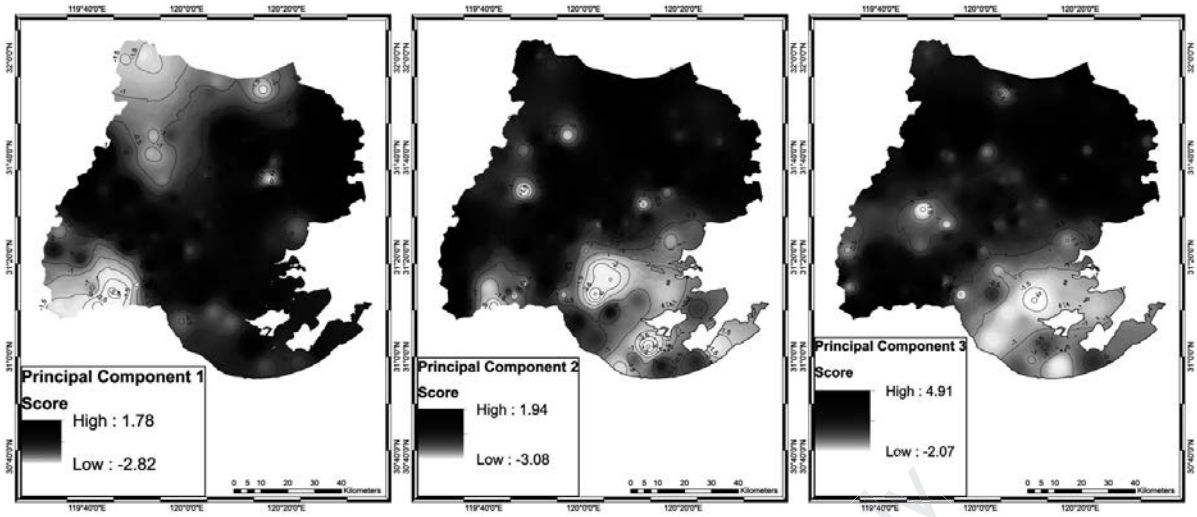

Fig. 7. Maps showing the basin variability in principal components 1-3.

duce discharges of N and P into rivers and streams. This is consistent with the fact that these regions of the lake have the lowest quality of water and the most severe amount of eutrophication in all years.

Hierarchial cluster analysis

The principal components provide the data for the cluster analysis. The main structure of the dendrogram consists of seven clusters that are seen at an error sum of squares level of approximately 5.0 (Fig. 8). The statistical significance of the differences between the seven clusters will be accessed by discriminant analysis in the next section. The water samples in each major cluster associate with a single type of land use as follows in Tab. 4: Cluster 1, Agriculture and aquaculture; Cluster 2, Urban and industry; Cluster 3, Rural and villages and urban and industry; Cluster 4, Agriculture and aquaculture; Cluster 5, Lake Taihu basin waters; Cluster 6, Rural and villages; and Cluster 7, Water conservation and forestry. The oc-

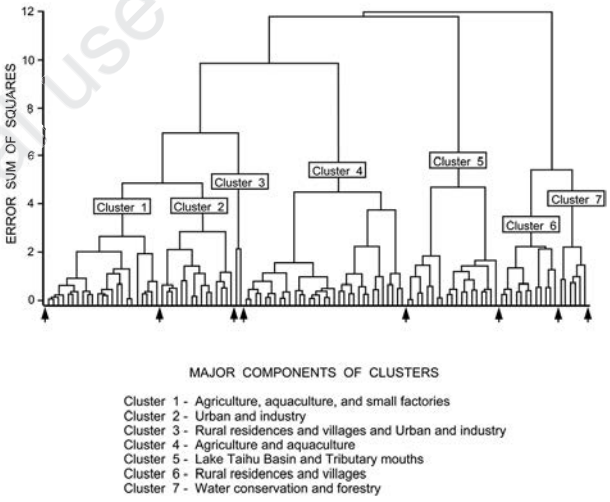

Fig. 8. Cluster analysis results of water quality in Lake Taihu basin. Distinct clusters fall out in similar pattern as found from PCA analysis, each reflecting different land uses.

Tab. 4. Distributions of samples in clusters.

Cluster number	Predominant land use						Predominant land use
	Agriculture, aquaculture, and minor factories	Tributary mouths	Water conservation and forestry	Lake Taihu basin	Rural residences and villages	Urban and industry	
1	16	3	0	0	4	0	Agriculture, aquaculture and minor factories in small areas
2	3	0	0	0	2	10	Urban and industry
3	0	0	0	0	1	1	Rural residences and villages and Urban and industry
4	20	5	0	5	3	0	Agriculture and aquaculture in large areas
5	1	0	0	18	0	0	Lake Taihu basin
6	1	0	0	0	9	2	Rural residences and villages
7	0	0	6	0	0	0	Water conservation and forestry

currence of two major clusters of water samples dominated by agricultural and aquacultural activities is rather surprising. The farms of Cluster 1 are generally small and sometimes occur with small factories. On the other hand, the agriculture and aquaculture areas of Cluster 4 cover much larger areas.

Cluster analysis effectively preserves small scale features and groups similar items together. Such relationships are found within the main clusters. Inspection of Tab. 4 illustrates a series of overlaps between waters in the various land use types as follows: the tributary mouth waters tend to cluster with agricultural and aquacultural samples. Most samples of lake water are distinct, but some link to samples from agricultural and aquacultural areas. The water samples in water conservation and forestry areas are mostly separate but a few cluster with rural and village regions. Waters from agricultural and aquacultural areas commonly group with those from rural and village lands. Some waters from urban and industrial regions are found in clusters with agricultural and aquacultural or rural and village areas. These overlaps between the waters in areas of different types of land uses can also be seen in the plot of the scores for the principal components (Fig. 6). We expected some overlap in the multivariate statistical analyses because of the complexity of the landscape and the fact that some land use types form small subsets within larger areas characterized by another type of land use. Nevertheless, the fact that the principal component and cluster analyses can broadly partition the water quality types into the various landscape use categories demonstrates the effectiveness of using a combination of these statistical methods on this and similar data sets.

Discriminant analysis

Discriminant analysis explores the statistical significance of the differences in water chemistry between the samples derived from areas subject to various types of land use and complements information previously derived from the structural methods of principal components and hierarchical cluster analysis. Two discriminant analyses were done on the water chemistry data, both of which produced very similar results. One analysis treated the seven clusters seen in the dendrogram (Fig. 8). The second analysis was done on the chemistry of the water samples from the six land use groups outlined previously from the principal components and cluster analyses. The results discussed here are obtained from the second analysis. The Chi-square value for the differences between all six land use groups equals 164 with 60 degrees of freedom which is significant at any reasonable probability level.

Clearly, differences exist between the chemistries of the water samples from the six land use types. But despite the overall differences between the waters of the six groups of land use, many of the tests between pairs of land use groups are not significant, and only seven of the 15 possible tests were significant at a probability of 0.05 or 0.01 (Tab. 5). The discriminant analysis only was able to identify or classify 61.8 percent of the samples correctly (Tab. 6). This lack of agreement indicates a large amount of overlap between the waters occur in the different land use categories. Nevertheless, the percent of correct classifications is significantly greater than chance or random allocation. With six groups, random roll would only allocate about 17% of the samples correctly. Inspection of the F-ratios and the classification or identification matrix in Tabs. 5 and 6 reveals some interesting patterns.

The Lake waters of Group 4 closely resemble and are not significantly different from those in the Tributary mouths in Group 2 which suggests that these can be visualized as the ambient or background waters for Lake Taihu. The different types of land use modify this water in various ways. he only significant differences for the Lake Taihu waters and the Tributary mouths involve the rural residences and villages of Group 5 and the urban and industrial samples in Group 6. This suggests that these types of land use exert the strongest effects on the background waters although they do so in very different ways. The addition of fertilizer, animal and human wastes, and chemical reactions between the water and the limestone bedrock are important in the rural areas; in addition Group 5 samples are typically lower in Na, K, F, Cl, and SO_4 than the background waters, and the rural areas seem to lack natural sources of these constituents. On the other hand, the waters in the Urban and industrial areas of Group 6 can be enriched in Na, K, F, Cl, and SO_4, and some samples are extremely high in TP and the nitrogen compounds stemming from human anthropogenic and industrial contributions. As expected, the rural samples in Group 5 and the urban and industrial ones in Group 6 are significantly different and none of the observed samples overlap between the two groups.

The waters in the agricultural and aquacultural areas of Group 1 are not statistically different from the background waters in Lake Taihu and the Tributary mouths of Groups 2 and 4 and the principal component scores (Fig. 6) and the identification matrix of Tab. 6 reveals considerable overlap between the water samples in these groups. Overall the agricultural water samples are slightly more variable with respect to Na, K, F, Cl, and SO_4 (principal component 1) and enriched in the carbonate (Ca, Mg, Sr, and HCO_3 of principal component 2) and nitrogen and phosphorus (principal component 3) compared with background water. The mean compositions of the Agricultural and aquacultural waters of Group 1 are significantly different from those in the rural areas of Group 5 and the urban and industrial areas of Group 6. The samples of the rural and agricultural groups overlap as seen in the classification or identification matrix. Interestingly, the agricultural and aquacultural waters of

Group 1 are roughly intermediate between the rural and village samples of Group 5 and the urban and industrial areas of Group 6 with respect to Na, K, F, Cl, and SO_4 (principal component 1).

Although only represented by eight water samples, the Water conservation and forestry areas of Group 3 are somewhat unique. The only statistically significant difference is with the Agricultural and aquacultural samples of Group 1, but this may be partly due to the small sample size of the group. For a small group, the samples are characterized by a wide range of variation. Compared to the background water, these samples are low in Na, K, F, Cl,

Tab. 5. F-ratios comparing pairs of groups.

	Group 1 Agriculture, aquaculture, and minor factories	Group 2 Tributary mouths and forestry	Group 3 Water conservation	Group 4 Lake Taihu basin	Group 5 Rural residences and villages	Group 6 Urban and industry
Agriculture, aquaculture and small factories	-	2.02	2.82**	1.62	3.02**	4.12**
Tributary mouths		-	0.852	1.91	2.74*	1.12
Water conservation and forestry			-	1.3	2.39	0.718
Lake Taihu basin				-	3.71**	2.45*
Rural residences and villages					-	2.93*
Urban and industry						-

Degrees of freedom for denominator of F-ratio

	Group 1 Agriculture, aquaculture, and minor factories	Group 2 Tributary mouths and forestry	Group 3 Water conservation	Group 4 Lake Taihu basin	Group 5 Rural residences and villages	Group 6 Urban and industry
Agriculture, aquaculture and small factories	-	36	36	51	47	39
Tributary mouths		-	3	18	14	6
Water conservation and forestry			-	18	14	6
Lake Taihu basin				-	29	21
Rural residences and villages					-	17
Urban and industry						-

*The degrees of freedom for the numerators of all F-ratios equal 12. *P=0.05; **P=0.01.*

Tab. 6. Classification matrix for 110 samples.

Group		Predicted group membership						Total
		1	2	3	4	5	6	
Count	1	23	5	1	9	3	0	41
	2	3	5	0	0	0	0	8
	3	0	0	5	2	0	1	8
	4	7	0	0	14	0	2	23
	5	2	1	0	1	15	0	19
	6	3	0	2	0	0	6	11
%	1	56.1	12.2	2.44	21.95	7.32	0	100
	2	37.5	62.5	0	0	0	0	100
	3	0	0	62.5	25	0	12.5	100
	4	30.43	0	0	60.87	0	8.7	100
	5	10.53	5.26	0	5.26	78.95	0	100
	6	27.27	0	18.18	0	0	54.55	100

61.8% of original grouped cases correctly classified.

and SO_4 (principal component 1), probably because natural sources of these components are not present. In addition the Group 3 waters seem to be low in the carbonate compounds Ca, Mg, Sr, and HCO_3, perhaps due to their soil composition.

CONCLUSIONS

The dissolved solutes in the Lake Taihu watershed derive from three major sources: water-rock interaction, atmospheric precipitation, and anthropogenic activity. Lake water generally has lower concentrations of TDS than surface waters because Lake Taihu has a larger volume of water than in tributaries and inflow solutes have been mixed and diluted by precipitation on the lake. Those water samples from upland streams with lower concentrations of TDS than lake water probably have less anthropogenic contamination than other areas. Waters with higher TDS reflect higher concentrations of Na and Cl that are common in human wastes and sewage water. The Lake Taihu basin is underlain by carbonate rocks or soils that are rich in carbonates, which explains the dominance of Ca-Mg-HCO_3 waters. In fact, Lake Taihu is famous for its ornamental limestone.

The combination of our statistical and geochemical analysis all suggest that Na, K, F, Cl^- and SO_4 in surface waters probably relate to anthropogenic contamination from sewage water, human waste, and industries; Ca, Mg, Sr and alkalinity (HCO_3) relate to uncontaminated water conditioned by the dissolution of carbonate-rich soils and rock. Finally, nutrients (NO_3 and total nitrogen TN), are related to agricultural or aquaculture fertilizer, animal waste, and septic discharge.

Given the complexity of the study area, many waters mix and classifications overlap. Nonetheless, we were able to assign virtually all of the water samples into coherent groups from the combination of our methods. The spatial distributions of three extracted principal components clearly relate to solute concentrations that are consistent with their respective locations which are impacted by different land use patterns. The combination of our integrated multivariate statistical approaches coupled with simple geochemical classification was able to assess how and where the major water quality differences in the Lake Taihu area occur and how they are influenced by different land use types.

ACKNOWLEDGMENTS

This research has been supported by the Crucial Special Project: National Water Pollution Control and Management Science (2008ZX07526-001), State Key Development Program for Basic Research of China (2008CB418201 and 2008 CB 418004). Special thanks to the colleagues in School of Life Sciences of Nanjing University, Ying Cai, Delin Xu, and Hao Jiang for the collaboration in field work.

REFERENCES

Alberto WD, Del Pillar DM, Valeria AM, Fabiana PS, Cecilia HA, De Los Angeles BM, 2001. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia River Basin (Cordoba-Argentina).Water Res. 35:2881-2894.

Back W, 1966. Hydrochemical facies and groundwater flow patterns in the northern part of the Atlantic Coastal Plain. Available from: http://pubs.usgs.gov/pp/0498a/report.pdf

Barros Grace V, Mas-Pla J, Oliveira Novais T, Sacchi E, Zuppi GM, 2008. Hydrological mixing and geochemical processes characterization in an estuarine/mangrove system using environmental tracers in Babitonga Bay (Santa Catarina, Brazil). Cont. Shelf. Res. 28:682-695.

Belkhiri L, Boudoukha A, Mouni L, Baouz T, 2010.Application of multivariate statistical methods and inverse geochemical modeling for characterization of groundwater - A case study: Ain Azel plain (Algeria). Geoderma 159:390-398.

Cao Y, Zhang E, Chen X, John AN, Shen J, 2012. Spatial distribution of subfossil Chironomidae in surface sediments of a large, shallow and hypertrophic lake (Taihu, SE China). Hydrobiologia 691:59-70.

Chen J, Yu Z, Yun C, 1959. Morphology of Yangtze River Delta. Acta Geogr. Sinica 25:201-220.

Cloutier V, Lefebvre R, Therrien R, Savard MM, 2008. Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system. J. Hydrol. 353:294-313.

Cortecci G, Boschetti T, Dinelli E, Cidu R, Podda F, Doveri M, 2009.Geochemistry of trace elements in surface waters of the Arno River Basin, northern Tuscany Italy. Appl. Geochem. 24:1005-1022.

Davis JC, 2002. Statistics and data analysis in geology. 3. Wiley: 656 pp.

De Carlo EH, Beltran VL, Tomlinson MS, 2004. Composition of water and suspended sediment in streams of urbanized subtropical watersheds in Hawaii. Appl. Geochem. 19:1011-1037.

Deocampo DM, 2004. Hydrogeochemistry in the Ngorongoro Crater, Tanzania, and implications for land use in a World Heritage Site. Appl. Geochem. 19:755-767.

Farnham IM, Johannesson KH, Singh AK, Hodge VF, Stetzenbach KJ, 2003. Factor analytical approaches for evaluating ground water trace element chemistry data. Anal. Chim. Acta 490:123-138.

Fitzpatrick ML, Long DT, Pijanowski BC, 2007. Exploring the effects of urban and agricultural land use on surface water chemistry, across a regional watershed, using multivariate statistics. Appl. Geochem. 22:1825-1840.

Frey KE, Siegel DI, Smith LC, 2007.Geochemistry of west Siberian streams and their potential response to permafrost degradation.Water Resour. Res. 43:W03406.

Hastie T, Tibshirani R, 1996. Discriminant analysis by Gaussian mixtures. J. R. Statist. Soc. B 58:155-176.

Hem JD, 1992. Study and interpretation of the chemical characteristics of natural water. Available from: http://pubs.usgs.gov/wsp/wsp2254/

Hounslow AW, 1995. Water quality data analysis and interpretation. CRC Press: 416 pp.

Howarth RJ, Garrett RG, 2010. Statistical analysis and data display at the Geochemical Prospecting Research Centre and Applied Geochemistry Research Group, Imperial College, London. Geochem.-Explor. Env. A 10:289-315.

Johnson RA, Wichern DW, 2007. Applied multivariate statistical analysis. 6. Pearson: 800 pp.

Lin L, Wu J, Wang S, 2006. Evidence from isotopic geochemistry as an indicator of eutrophication of Meiliang Bay in Lake Taihu, China. Sci. China Ser. D 49:62-71.

Mason CF, Norton SA, Fernandez IJ, Katz LE, 1999. Deconstruction of the chemical effects of road salt on stream water chemistry. J. Environ. Qual. 28:82-91.

McLachlan GJ, 2004. Discriminant analysis and statistical pattern recognition. Wiley-Interscience: 552 pp.

Meador MR, Goldstein RM, 2003. Assessing water quality at large geographic sacles: relations among land use, water physicochemistry, riparian condition, and fish community structure. Environ. Manage. 31:504-517.

Panno SV, Hackley KC, Hwang HH, Greenberg SE, Krapac IG, Landsberger S, O'Kelly DJ, 2006. Characterization and identification of Na-Cl sources in ground water. Ground Water 44:176-187.

Piper CS, 1944. Manganese deficiency in oats. Nature 153:197.

Poinke HB, DeWalle DR, 1994. Streamflow generation on a small agricultural catchment during autumn recharge: nonstormflow periods. J. Hydrol. 163:1-22.

Qin B, 2008. Lake Taihu, China: dynamics of environmental change. Springer: 339 pp.

Qin B, Xu P, Wu Q, Luo L, Zhang Y, 2007. Eutrophication of shallow lakes with special reference to lake Taihu, China. Hydrobiologia 581:3-14.

Reeve AS, Siegel DI, Glaser PH, 1996.Geochemical controls on peatland pore water from the Hudson Bay Lowland: a multivariate statistical approach. J. Hydrol. 181:285-304.

Shi J, Li G, Wang P, 2011. Anthropogenic influences on the tidal prism and water exchanges in Jiaozhou bay, Qingdao, China. J. Coastal Res. 27:57-72.

Steuer J, Selbig W, Hornewer N, Prey J, 1997. Sources of contamination in an urban basin in marquette, Michigan and an analysis of concentrations, loads, and data quality. Available from: http://pubs.usgs.gov/wri/1997/4242/report.pdf

Stonestrom DA, Scanlon BR, Zhang L, 2009. Introduction to special section on impacts of land use change on water resources. Water Resour. Res. 45:W00A00.

Subramani T, Elango L, Damodarasamy SR, 2005. Groundwater quality and its suitability for drinking and agricultural use in Chithar River Basin, Tamil Nadu, India. Environ. Geol. 47:1099-1110.

Sun S, Wu Y, Dong B, 1987. The bottom configuration and recent deposition of Lake Taihu. Memoirs of Nanjing Institute of Geography and Limnology, Academic Sinica. 4:1-14 pp.

Sun SC, Huang YP, 1993. [Taihu Lake].[Book in Chinese]. Ocean Press: 271 pp.

Tang Z, Engel BA, Pijanowski BC, Lim KJ, 2005. Forecasting land use change and its environmental impact at a watershed scale. Environ. Manage. 76:35-45.

Wang E, Wan Y, Xu S, 2002. Discovery and implication of shock metamorphic unloading microfractures in Devonian Bedrock of Taihu Lake, Sci. China Ser. D 45:459-467.

Ward JH, 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58:236-244.

Wayland KG, Long DT, Hyndman DW, Pijanowski BC, Woodhams SM, Haack SK, 2003. Identifying relationships between baseflow geochemistry and land use with synoptic sampling and R-mode factor analysis. Environ. Qual. 31:180-190.

Wilhelm SW, Farnsley SE, LeCleir GR, Layton AC, Satchwell MF, DeBruyn JM, Boyer GL, Zhu G, Paerl HW, 2011. The relationships between nutrients, cyanobacterial toxins and the microbial community in Taihu (Lake Tai), China. Harmful Algae 10:207-215.

Xie Z, Wang H, Sharp T, Decarli P, 2008. New evidence for an impact origin of Taihu Lake, China: possible trigger of the extinction of Liangchu culture 4500 years ago. Proceedings of the AGU Fall Meeting, Abstract No. MR12A-07.

Xu J, Wang J, Liang T, Tang X, 2009. [Analysis of land use change in Taihu Basin in the past 18 years].[Article in Chinese]. Geospatial Information 7:41-51.

Yu C, Cheng X, Hall J, Evans EP, Wang Y, Hu C, Wu H, Wicks J, Scott M, Sun H, Wang J, Ren M, Xu Z, 2012. A GIS-supported impact assessment of the hierarchical flood-defense systems on the plain areas of the Taihu Basin, China. Int. J. Geogr. Inf. Sci. 26:643-665.

Yuan H, Shen J, Liu E, Wang J, Meng X, 2010. Assessment of nutrients and heavy metals enrichment in surface sediments from Taihu Lake, a eutrophic shallow lake in China. Environ. Geochem. Health 33:67-81.

Zeng C, Wang L, Huang W, Wang W, 2012. Research on influencing factors of water environment in the typical western of Taihu Lake based on the principal component analysis. Adv. Mat. Res. 356-360:924-928.

Zhang Q, Jiang T, Shi Y, Lorenz K, Liu C, Martin M, 2004. Paleo-environmental changes in the Yangtze Delta during past 8000 years. J. Geogr. Sci. 14:105-112.