Environ Monit Assess (2009) 152:123–131
DOI 10.1007/s10661-008-0301-y

Assessment of surface water quality using multivariate statistical techniques in red soil hilly region: a case study of Xiangjiang watershed, China

Qi Zhang · Zhongwu Li · Guangming Zeng · Jianbing Li · Yong Fang · Qingshui Yuan · Yamei Wang · Fangyi Ye

Received: 16 December 2007 / Accepted: 9 April 2008 / Published online: 6 June 2008

Abstract In the study, multivariate statistical methods including factor, principal component and cluster analysis were applied to analyze surface water quality data sets obtained from Xiangjiang watershed, and generated during 7 years (1994–2000) monitoring of 12 parameters at 34 different profiles. Hierarchical cluster analysis grouped 34 sampling sites into three clusters, including relatively less polluted (LP), medium polluted (MP) and highly polluted (HP) sites, and based on the similarity of water quality characteristics, the watershed was divided into three zones. Factor analysis/principal component analysis, applied to analyze the data sets of the three different groups obtained from cluster analysis, resulted in four latent factors accounting for 71.62%, 71.77% and 72.01% of the total variance in water quality data sets of LP, MP and HP areas, respectively. The PCs obtained from factor analysis indicate that the parameters for water quality variations are mainly related to dissolve heavy metals. Thus, these methods are believed to be valuable to help water resources managers understand complex nature of water quality issues and determine the priorities to improve water quality.

Keywords Water quality management · Principle component analysis · Cluster analysis · Xiangjiang watershed · Red soil hilly region

Q. Zhang · Z. Li (✉) · G. Zeng · J. Li ·
Y. Fang · Q. Yuan · Y. Wang · F. Ye
College of Environment Science and Engineering,
Hunan University,
Changsha 410082, People's Republic of China
e-mail: lizhongwu@yahoo.com.cn

J. Li
Environmental Engineering Program,
University of Northern British Columbia,
Prince George, BC V2N 4Z9, Canada

Introduction

The water environment quality issue is a subject of ongoing concerned with the development of economy in any country. Especially, since the reform and open-up in China, the water resources problems related to environmental degradation have increasingly been serious, because of the rapid industrialization and urban sprawl. Due to their roles in transporting domestic and industrial wastewater and non-point source pollutants from agricultural land in their vast drainage basins, rivers are among the most vulnerable water bodies to pollution. Anthropogenic influence (urbanization, industrial and agricultural activities, increasing consumption of water resources) as well as natural process (changes in precipitation inputs, erosion, weathering of crustal materials) degrade surface water quality and impair their use for drinking, industrial, agriculture, recreation or other purposes (Carpenter et al. 1998; Jarvie et al. 1998). In order to effectively manage and research river water environment, obtaining water environment quality parameter data is indispensable. Although the regular measurements needs doing much work, because of spatial and temporal variation of water environment quality, monitoring by regular measurements, which will provide a representative and reliable estimation of surface water quality, is necessary. The long-term monitoring for many profiles in different reach will generate a large and complex database, which needs a good approach to interpret (Chapman 1992).

The application of different multivariate statistical techniques, such as cluster analysis (CA), principal component analysis (PCA), factor analysis (FA), helps in the interpretation of complex data matrices to better understand the water quality and ecological status of the studied systems, allows the identification of possible factors that influence water environment systems and offers a valuable tool for reliable management of water resources (Simeonov et al. 2003; Praus 2005; Shrestha and Kazama 2007). In recent years, many studies related with these methods have been carried out. For instance, Zhou et al. (2007) and Boyacioglu and Boyacioglu (2007) use the PCA and CA to classify the sampling sites and to identify the latent pollution source. Mendiguchía et al. (2004) used the CA to divide a watershed to four zones with different water quality except using PCA to identify the main pollutants. According to the above researches, it can be concluded that these methods could be used to assess the relationships between variables and possible pattern in distribution of measured data. As a result, in the study, we mainly use CA to identify several zones with different water quality and PCA to find the most important factors that describe the natural and anthropogenic influences.

The Xiangjiang watershed is the most developed and urbanized region in the province, and the general economic quantity (GDP) of the three main cities (Changsha, Zhuzhou, Xiangtan) takes one third of the total of Hunan province. In the basin, mining and metallurgical industry is much more developed, and Zhuzhou metallurgical group corporation, Shuikoushan mining bureau and Xiangtan steel group distributed along the river. As a result, it highlighted the heavy metal pollution for the rivers. At the same time, due to uncontrolled disposal of urban and industrial waste as well as improper disposal of solid and toxic waste from industrial, agricultural and other human activities, more and more wastewater discharged into river. For instance, according to statistics, the annual use of the pesticide chemical fertilizers increased approximately 6% (Chen et al. 2004). Because of the mentioned reasons, water environment situation of the watershed is increasingly deteriorating. The increasing water pollution not only causes the deterioration of water quality but also threatens human health and the balance of aquatic ecosystems, economic development and social prosperity. Meanwhile, with an increasingly understanding of the importance of drinking water quality to public health and raw water quality to aquatic life, the government of Hunan Province implemented Regulations of Water Pollution Control in Xiangjiang watershed in 1998. Since then, many efforts have been devoted to restoring and protecting Xiangjiang watershed. Although many regulations were manipulated, plenty of pollutants in Xiangjiang river were increasing all the time (Chen et al. 2004), because of many reasons, including improper land use, local protectionism for the cities along the river, and so on. Hence, how to reduce pollution is the current major problem in the watershed.

In the present study, a large data matrix, obtained during a 7-year (1994–2000) monitoring program, is subjected to different multivariate statistical techniques to extract information about the similarities or dissimilarities between sampling sites, and the influence of possible source on the water quality parameters of Xiangjiang watershed. The specific objectives are to: (1) identify several zones with different water quality, (2) extract the parameters that are most important in assessing variations in river water quality of different zones, (3) find a good approach to assess the water quality reasonably. It can be helpful to the managers to understand the main pollutants of each cluster of sampling sites, and to take effective measures to manage the water resources, respectively.

Material and methods

Study sites

The Xiangjiang river, as the mother river of the Hunan province, originates from Haiyang Mountain of Guangxi Province, China, flows through Yongzhou City, Hengyang City, Zhuzhou City, Xiangtan City, Loudi City, Changsha City and Yueyang City and discharges to Dongting Lake and its flows complemented of several tributaries along its course, mainly including Xiaoshui River, Chunlingshui River, Zhengshui River, Leishui River, Lushui River, Lianshui River, Liuyanghe River, etc. (Fig. 1). The Xiangjiang watershed, with a basin area of 85,300 km^2 in Hunan Province, almost 40% of the whole province, was selected as the study site. The population of the watershed is almost 60% of that of Hunan province. In Xiangjiang watershed, the climate belongs to typical subtropical monsoon climate: hot and humid in summers and cold and dry in winters, with average temperature of between 16 °C and 18 °C, and a mean annual precipitation of 1,400 mm. For the features of Xiangjiang watershed, this region can be regarded as an example of a red soil hilly region.

The red soil hilly region is of low ecological stability and high fluctuation in biological system and productivity, sensitive to the activities of humans, and sudden disasters. The physical environment always tends to being more and more serious, and being against human being's survival (Cao and Zhang 1995). The distribution of red soil hilly regions is broad, including 11 provinces and 619 counties in southern China, occupying an area of 1.13 million km^2 and accounting for 11% of the total land of China (Lu and Shi 2000). Here, the subtropical monsoon climate gives the region sound bioclimatic conditions (annual rainfall of 1,400–1,700 mm, mean annual temperature of 16–19 °C) and a strong potential for producing enormous quantities of biomass. Due to the long-term excess development of natural resources, red soil hilly region has become one of the most vulnerable eco-environment regions in China (Cao and Zhang 1995), characterized by serious erosion, heavy floods and droughts, degressive land productivity, and degraded ecological stability. The integrated analysis of the vulnerable eco-environment in the red soil hilly region will prompt the development of regional economics and the improvement of human habitats.

Sampling and parameters

The sampling strategy was designed in such a way to cover a wide range of determinants at key sites that accurately represent the water environment quality of the river systems and account for tributary inputs that can have important impacts upon downstream water quality. Various water quality parameters from the monitoring stations were collected monthly by Center for Environmental Monitoring of Hunan Province. In this study, data from 34 monitoring stations were selected under the water environment quality-monitoring network, which covers a wide range of catchments and surface water types (rivers and tributaries) of Xiangjiang Watershed (Fig. 1). The measured parameters include field pH, dissolved oxygen (DO), nitrate (NO_3^-); nitrite (NO_2^-); ammonium (NH_4^+), chemical oxygen demand (COD), biological oxygen demand (BOD_5), acid-hydrolysable (total) phosphorus (TP), Hg, As, Cd, Cr^{6+}, Cu^{2+}. The summary basic statistics of the dataset is presented in Table 1.

Cluster analysis

Cluster analysis (CA) is used to develop meaningful aggregations, or groups, of entities based on a large number of interdependent variables. The resulting clusters of objects should exhibit high internal (within-cluster) homogeneity and high external (between clusters) heterogeneity (McGarial et al. 2000). Of all cluster analysis, hierarchical agglomerative cluster is most common approach. In the study, hierarchical agglomerative CA was performed based on the normalized data set (mean of observations over the whole period) by means of the Ward's method using squared Euclidean distances as a measure of similarity. The spatial variability of water environment quality in the whole river basin was determined from CA, which divides a large number of objects into smaller number of homogenous groups on the basis of their internal correlations (Hussain et al. 2008).

Principal component analysis/factor analysis

Factor analysis, which includes PCA is a very powerful technique applied to reduce the dimensionality of a dataset consisting of a large number of interrelated variables, while retaining as much as possible the variability presented in dataset. This reduction is achieved by transforming the dataset into a new set of variables—the principal components (PCs), which are orthogonal (non-correlated) and are arranged in decreasing order of importance. Mathematically, the PCs are computed from covariance or other

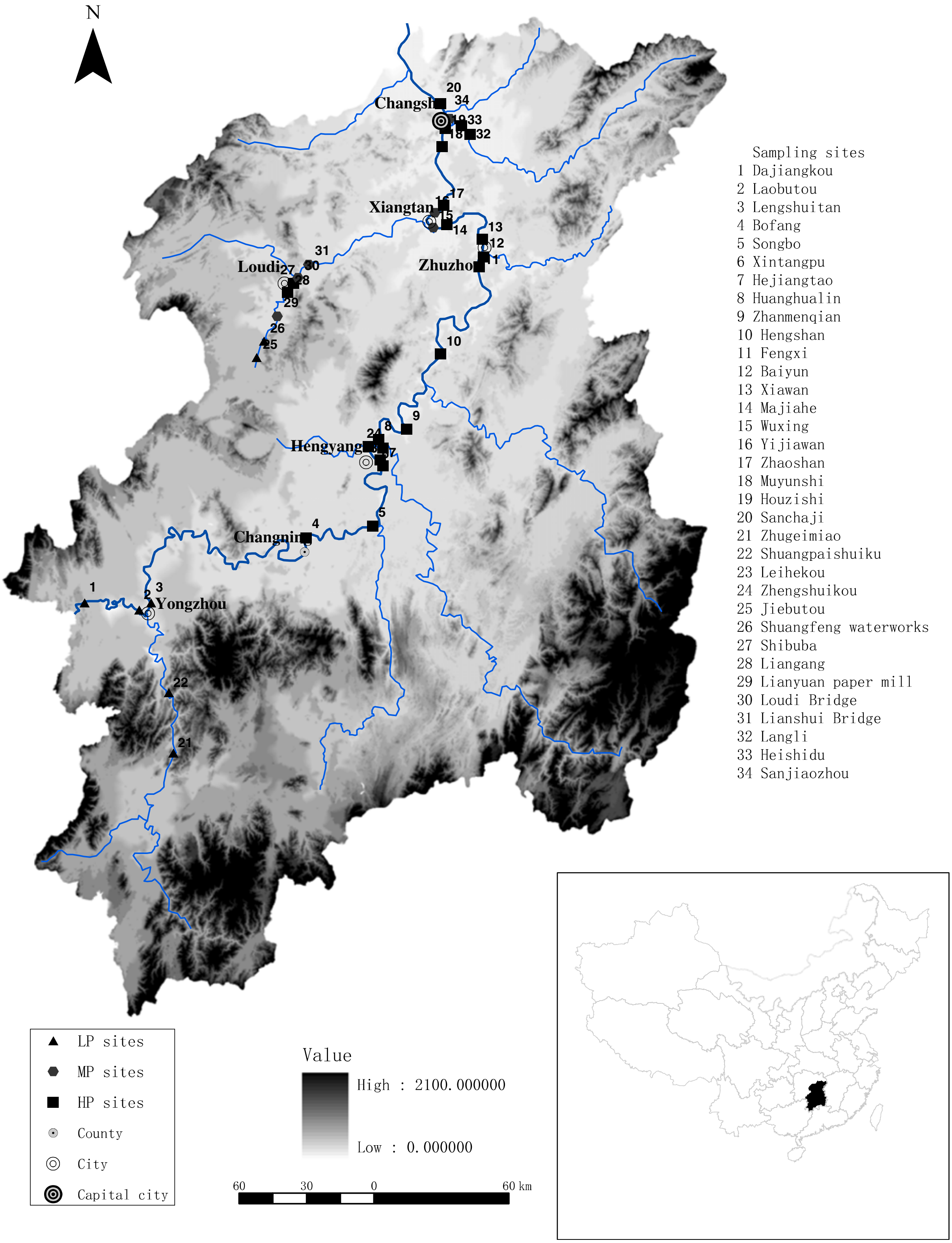

Fig. 1 Map of study area and surface water quality monitoring stations in Xiangjiang watershed

Table 1 Descriptive statistics of water quality variables

	pH (pH unit)	DO (mg/L)	COD_{Mn} (mg/L)	BOD_5 (mg/L)	NH_4^+ (mg/L)	NO_3^- (mg/L)	TP (mg/L)	As (mg/L)	Pb (mg/L)	Hg (mg/L)	Cr (mg/L)	Cd (mg/L)	Cu (mg/L)
Mean	7.69	6.23	2.98	2.76	0.28	0.83	0.15	0.016	0.012	0.00057	0.036	0.016	0.0044
Median	7.80	7.65	2.73	2.56	0.13	0.91	0.11	0.07	0.007	0.00005	0.002	0.0007	0.00325
Mode	7.80	7.80	2.89	2.56	0.12	0.35	0.05	0.004	0.002	0.00002	0.002	0.002	0.002
Standard deviation	0.38	0.89	1.30	0.84	0.42	0.41	0.04	0.06	0.019	0.004	0.005	0.05	0.004
Variance	0.15	0.79	1.69	0.71	0.17	0.17	0.37	0.003	0.001	0.00002	0.00002	0.002	0.00002
Minimum	6.62	3.4	1.12	0.51	0.06	0.04	0.14	0	0.001	0	0	0	0
Maximum	8.21	9.5	9.69	6.51	6.51	1.97	4.89	0.59	0.16	0.050	0.029	0.07	0.028

cross-product matrix, which describes the dispersion of the multiple measured parameters to obtain eigenvalues and eigenvectors. Principal components are the linear combinations of the original variables and the eigenvectors (Wunderlin et al. 2001).

PCA can be used to reduce the variable numbers and explain the same amount of variance with fewer variables (principal components) (Wu and Wang 2007). Factor analysis attempts to explain the correlations between the observations in terms of the underlying factors, which are not directly observable (Yu et al. 2003).

This study comprises application of multivariate statistical techniques to analyze water quality dataset obtained from Xiangjiang River and its tributaries in Hunan province. Statistical calculations were performed using the "Statistical Package for the Social Sciences Software—SPSS 11.5 for Windows".

Results and discussion

Spatial similarity and site grouping

In this study, sampling sites classification was performed by the use of cluster analysis (z-transformation of the input data, squared Euclidean distance as similarity measure and Ward's method of linkage) and dendrogram was generated, which grouped all 34 sampling sites of the basin into three statistically significant clusters (Fig. 2). Since we used hierarchical agglomerative cluster analysis, the number of clusters was also decided by water environment quality, which is mainly effected by land use and industrial structure. Grouped stations under each cluster can be seen in Fig. 2. Based on the results of cluster analysis, results can be concluded.

Cluster I (Stations 1–2–3–21–22–25–26)

Sites mainly located at middle reach of the watershed (Station 1–2–3–21–22) were grouped under Cluster I, which were basically at Yongzhou City. In addition, Station 25, 26 located upstream of Lian Creek, showed the similar water environment quality characteristics with these stations. Yongzhou City is an underdeveloped area. According to statistical yearbook of Hunan 2001 (see Table 2), the per capita GDP of Yongzhou City was 4,358 RMB, much lower than that of other cities in Xiangjiang Watershed; the primary industry accounts for 31.7%, much higher than that of the average level of Hunan Province and the urbanization and industrialization level is relatively low. Hence, the impact of human being's activities on the riverine ecosystem is relatively low. Although the mining and the direct discharged domestic wastewater contaminated the water, the cluster I corresponds to relatively Less polluted (LP), because the inclusion of the sampling location suggests the self purification and assimilative capacity of the river are strong.

Cluster II (15–16–29–30–31–34)

This cluster sites mainly located Xiangtan City and Loudi City, which were between Zhuzhou City and Changsha City. Sites 29, 30, 31 located middle stream of Lianshui River (a branch of Xiangjiang River), and the others located at the main stream. In addition, Station 34 corresponded to Liuyanghe River. Cluster II correspond to moderately pollution.

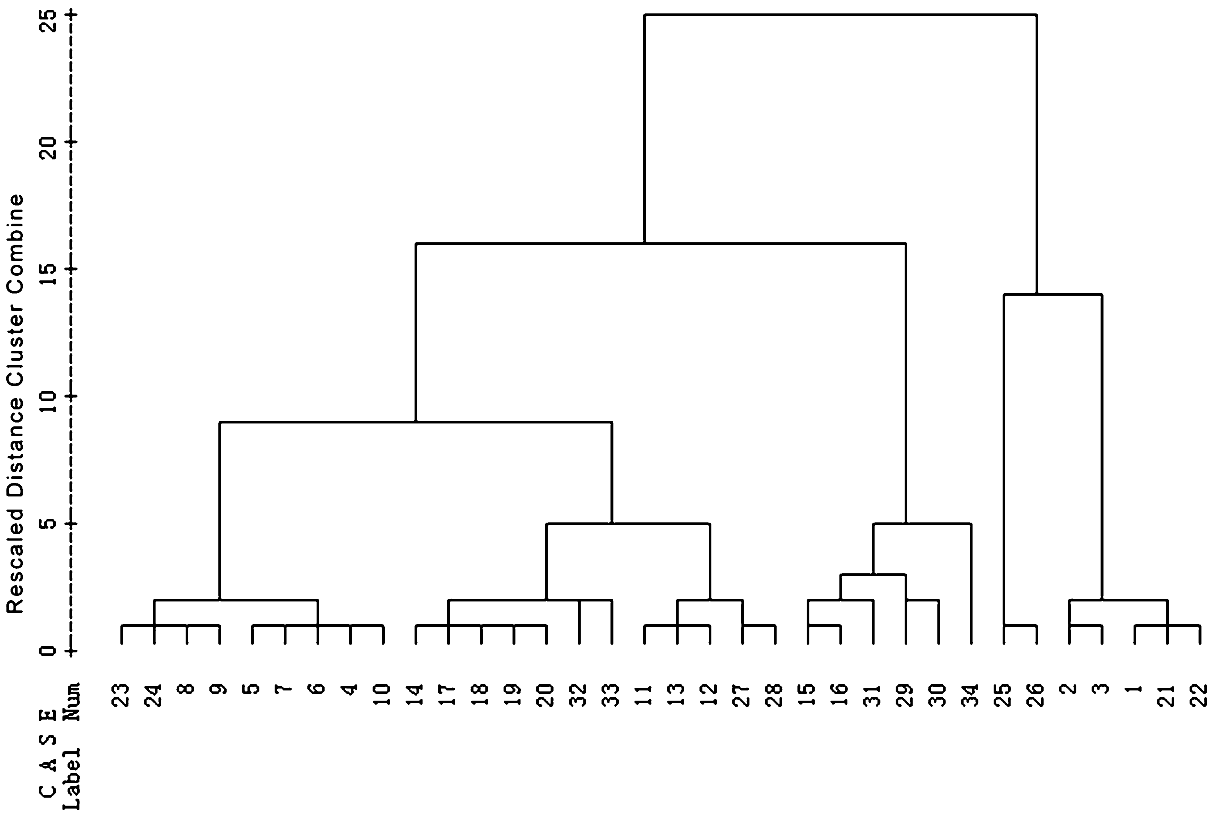

Fig. 2 Ward's minimum variance dendogram

Cluster III (the rest stations)

The rest stations mainly located Hengyang City, Zhuzhou City and Changsha City, which were the most important industrial cities of Hunan Province. These cities were the origin of nonferrous metals and the base of smelting, agriculture and heavy industry. Therefore, these sampling stations received pollutants mostly from domestic wastewater, wastewater treatment plants and industrial effluents. They were corresponded to highly polluted sites.

The results indicate that the CA technique is useful in offering reliable classification of surface water in the whole region and make it possible to design a future spatial sampling strategy in an optimal method, which can reduce the number of sampling sites. In recent years, some studies have been performed on the water environment quality from the overall watershed (Chen 2005; Sun et al. 2006), while the study allowed the identification of three different zones of the watershed with different water quality. As a result, we can take different measures to control water pollution in the zones after the PCA below, respectively.

Data structure determination and source identification

Principal component analysis/factor analysis was performed on the normalized data sets separately for the three different regions, viz., LP, MP and HP, as delineated by CA techniques, to compare the composition structure between analyzed water samples and

Table 2 The per GDP of the main cities in the watershed

	City						Whole province
	Changsha	Xiangtan	Zhuzhou	Hengyang	Yongzhou	Loudi	
Per GDP (RMB)	11,262	7,643	7,949	5,011	4,358	4,239	5,639

identify the factors influencing each one (Table 3). PCA of the three data sets yield four PCs for each sites with Eigenvalues >1, explaining 71.62%, 71.77% and 72.01% of the total variance in each water quality data sets. An Eigenvalue gives a measure of the significance for the factor, which with highest Eigenvalue is the most significant. Eigenvalues of 1.0 or greater are considered significant (Kim and Mueller 1978). And the factor loadings were classified as ‘strong’, ‘moderate’ and ‘weak’, corresponding to absolute loading values of >0.75, 0.75–0.50 and 0.50–0.30, respectively (Liu et al. 2003).

For the data set pertaining to LP sites, among four PCs, PC1, explaining 25.71% of the total variance, has moderate loadings on DO, TP, As and strong loading on Pb. PC2, explaining 17.48% of the total variance, has a moderate negative loading on nitrate nitrogen and strong loadings on Cu and ammonia nitrogen. PC3, explaining 15.63% of the total variance, has strong positive loadings on COD and moderate positive loadings on BOD_5. PC2 and PC3 represent organic pollution from domestic waste and nonpoint source pollution. PC4, explaining the lowest variance (12.80%), has moderate loadings on Cd and pH and strong negative loading on Cr. PC1 and PC4 represent the pollution from industrial pollution and mining activities.

For the data set representing the MP sites, among total four significant PCs, PC1, explaining about 25.51% of the total variance, has strong positive loading on Cd and moderate positive loadings on Cu, As and Pb. These factors represent the contribution of the industrial from heavy industry. PC2, explaining about 17.29% of the total variance, has strong positive loading on, COD and moderate loadings on DO and ammonia nitrogen, which represent pollution from domestic wastewater of Xiangtan City and Loudi City. PC3, explaining about 15.80% of the total variance, has strong positive loadings on nitrate nitrogen and total phosphorus, moderate loading on BOD_5. In these areas, farmers use the fertilizer, which represents point and non-point source pollution from orchard and agriculture areas. PC4 (13.15%) has moderate positive loadings on PH and Cr. The factor represents the pollution of electroplating wastewater.

Lastly, for the data set pertaining to water quality in HP site. PC1, explaining 33.49% of total variance, has strong loadings on Pb, TP, Cd, Cr and pH, and moderate loadings on Cu and Hg. This factor can be

Table 3 Loadings of experimental variables (12) on principal components for LP sites, MP sites and HP sites data sets

Variables	PC1	PC2	PC3	PC4
LP sites				
pH	−0.130	0.321	−0.552	−0.300
DO	**−0.716**	0.117	0.003	0.231
COD_{Mn}	−0.018	0.102	**0.829**	0.103
BOD_5	0.286	0.051	**0.714**	−0.209
NH_4–N	0.180	**−0.757**	0.334	0.225
NO_3–N	−0.022	**−0.702**	−0.126	−0.244
TP	**0.710**	0.170	0.173	0.054
As	**0.705**	−0.179	0.234	0.263
Pb	**0.805**	0.279	0.128	0.021
Hg	0.147	0.290	−0.133	−0.320
Cd	0.406	0.318	−0.176	**0.722**
Cr	0.278	0.166	−0.163	**−0.835**
Cu	−0.141	**0.786**	0.181	−0.044
Eigenvalue	2.571	1.748	1.563	1.280
% Total variance	25.706	17.479	15.631	12.803
Cumulative % variance	25.706	43.185	58.816	71.619
MP sites				
pH	0.179	−0.051	−0.186	**0.746**
DO	−0.31	**0.706**	0.032	0.181
COD_{Mn}	−0.006	**0.909**	0.029	0.114
BOD_5	−0.092	0.359	**0.705**	0.179
NH_4–N	0.008	**0.713**	0.255	−0.156
NO_3–N	−0.258	−0.057	**0.757**	0.074
TP	−0.026	0.095	**0.754**	−0.358
As	**0.743**	0.084	−0.113	−0.227
Pb	**0.734**	0.072	−0.076	0.100
Hg	−0.135	0.009	−0.146	0.080
Cd	**0.817**	−0.056	−0.099	−0.104
Cr	−0.188	−0.121	0.071	**0.708**
Cu	**0.704**	−0.090	−0.008	0.366
Eigenvalue	2.051	1.729	1.581	1.315
% Total variance	25.514	17.292	15.812	13.154
Cumulative % variance	25.514	42.806	58.618	71.772
HP sites				
pH	**−0.892**	−0.106	0.111	−0.129
DO	0.310	−0.371	0.028	**−0.708**
COD_{Mn}	0.035	**0.768**	0.106	−0.029
BOD_5	0.282	**0.703**	0.219	−0.084
NH_4–N	0.181	0.424	−0.584	0.199
NO_3–N	0.077	0.159	**0.735**	0.229
TP	**0.797**	−0.092	−0.049	0.118
As	0.363	**0.738**	−0.157	0.170
Pb	**0.900**	0.124	−0.010	−0.255
Hg	**−0.746**	0.338	0.166	0.380
Cd	**0.863**	−0.218	−0.138	0.009
Cr	**0.792**	−0.396	−0.016	0.214
Cu	**0.739**	0.178	−0.157	−0.176
Eigenvalue	3.349	1.617	1.165	1.067
% Total variance	33.493	16.169	11.646	10.699
Cumulative % variance	33.493	49.662	61.308	72.008

interpreted as representing influences from industrial discharge and agriculture activities. PC2, explaining 16.17% of the total variance, has moderate positive loadings on BOD_5 and As, and strong positive loading on COD_{Mn}. PC3, explaining 16.17% of the total variance, has moderate positive loading on nitrogen. These factors represent pollution from domestic wastewater and non-point sources. PC4, explaining 10.67% of the total variance, has moderate negative loading on DO, which is due to anaerobic conditions in river from the loading of high dissolved organic matter.

Through the PCA, the sources of the pollutants were identified in the three zones. As mentioned above, it can be helpful to the government and managers, who can lay down different regulations and policies in the three zones respectively.

Heavy metal pollution in Hunan Province

Based on the results of PCA analysis, we can know that heavy metals are the main pollutants in the basin. One hundred forty-six industrial enterprises, which are the key source of pollution, distributed in the basin, and discharged wastewater as high as 8.39×10^8 m^3/a, in which the metal is as high as 5,013.13 t/a (Wang et al. 2004), mainly composed of mercury, cadmium, lead, zinc, copper, chromium and other metals. The present paper critically introduced high pollution level of dissolved heavy metals at two ecologically distinct zones along the river: Changning in Hengyang City, Zhuzhou City. Because of its mining and smelting activities, about 28 high-pollution factories distributed along the Changning section, such as Shuikoushan mining bureau, and discharged wastewater as high as $2{,}752\times10^4$ m^3/a, whose toxicant content far excess the water quality standards. Zhuzhou City is the base of heavy industry in Hunan Province. The chemical and smelting industries have been recognized basic industries of the city. The wastewater, discharged from these industries, achieved $8{,}671\times10^4$ m^3/a, with a high concentration of heavy metal. For example, the emissions of harmful waste water of Zhuzhou Smelter is $1{,}650\times10^4$ m^3/a, and the content of heavy metals in waste water were 36.69 t/a for Zinc, 35.80 t/a for lead, 35.49 t/a for cadmium, 12.28 t/a for arsenic, which causing serious pollution in Zhuzhou City section.

Conclusions

In this case study, different multivariate statistical techniques were used to evaluate spatial variations in surface water quality of Xiang river basin. Hierarchical cluster analysis grouped 34 sampling sites into three clusters of similar water quality characteristics. Based on obtained information, it is possible to design an optimal sampling strategy, which could reduce the number of sampling stations and associate costs. Also this analysis allowed the identification of four different zones in the river, with different water quality. Although the factor/principle component analysis did not result in a significant data reduction, it helped extract and identify the factors/sources responsible for variations in river water quality at three zones. Varifactors obtained from factor analysis indicate that parameters responsible for water quality variations are mainly related to industrial and domestic wastewater.

With serious situation of water pollution in the Xiangjiang watershed, the management of water quality of the different zones is becoming more and more important as well as the planning of the whole watershed. According to the sources of pollution, different measures should be adopted, in order to control the total quantity of the pollutants and achieve the goal of sustainable use of the water resources. It could be helpful to managers and government agencies in water quality management.

As a result, multivariate statistical methods including factor, principal component, and cluster analysis can be used to understand complex nature of water quality issues and determine priorities to improve water quality.

Acknowledgement The study was funded by the National 863 High Technologies Research Program of China (Grant No. 2007AA10Z222) and the Natural Science Foundation of China for Distinguished Young Scholars (Grant No. 50225926, Grant No. 50425927).

References

Boyacioglu, H., & Boyacioglu, H. (2007). Water pollution sources assessment by multivariate statistical methods in the Tahtali Basin, Turkey. *Environmental Geology*, *54*, 275–282.

Cao, X. Z., & Zhang, G. S. (1995). Formation and countermeasures of the vulnerable eco-environment of red soil hilly region. *Rural Eco-environment*, *11*(4), 45–48 (In Chinese).

Carpenter, S. R., Caraco, N. E., Correll, D. L., Howarth, R. W., & Smith, V. H. (1998). Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications*, *8*(3), 559–568.

Chapman, D. (1992). *Water Quality Assessment. In: Chapman D. On behalf of UNESCO, WHO and UNEP*. London: Chapman & Hall, 585.

Chen, G. H. (2005). Problems and solutions of the protection and utilization of water resources in Hunan. *Journal of Hunan Economic Management College*, *16*(6), 3–5 (in Chinese).

Chen, Y. S., Wu, F. C., Lu, H. Z., & Yao, C. S. (2004). Analysis on the water quality changes in the Xiangjiang River from 1981 to 2000. *Resources and Environment in the Yangtze Basin*, *13*(5), 508–512 (in Chinese).

Hussain, M., Ahmed, S. M., & Abderrahman, W. (2008). Cluster analysis and quality assessment of logged water at an irrigation project, eastern Saudi Arabia. *Journal of Environmental Management*, *86*(1), 297–307.

Jarvie, H. P., Whitton, B. A., & Neal, C. (1998). Nitrogen and phosphorus in east coast British rivers: speciation, sources and biological significance. *Science of the Total Environment*, *210/211*, 79–109.

Kim, J. O., & Mueller, C. W. (1978). *Introduction to factor analysis: what it is and how to do it. Quantitative applications in the social sciences series*. Newbury Park, CA: Sage.

Liu, C. W., Lin, K. H., & Kuo, Y. M. (2003). Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. *Science of the Total Environment*, *313*, 77–89.

Lu, R. K., & Shi, Z. Y. (2000). Features and recover of degraded red soil. *Soil*, *4*, 198–209 (In Chinese).

McGarial, K., Cushman, S., & Stafford, S. (2000). *Multivariate statistics for wildlife and ecology research*. New York: Springer.

Mendiguchía, C., Moreno, C., Galindo-Riaño, M. D., & García-Vargas, M. (2004). Using chemometric tools to assess anthropogenic effects in river water: A case study: Guadalquivir River (Spain). *Analytica Chimica Acta*, *515* (1,5), 143–149.

Praus, P. (2005). Water quality assessment using SVD-based principal component analysis of hydrological data. *Water SA*, *31*(4), 417–422.

Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling & Software*, *22*(4), 464–475.

Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M., & Kouimtzis, Th. (2003). Assessment of the surface water quality in northern Greece. *Water Research*, *37*(17), 4119–4124.

Sun, S. Q., Hu, G. H., Wang, Y. Z., & Li, C. (2006). Water environmental health risk assessment of Xiangjiang River. *Journal of Safety and Environment*, *6*(2), 12–15 (in Chinese).

Wang, Q. H., Wang, S. Y., & Liu, M. Y. (2004). Safety evaluation on pollution of Xiang River Valley in Hunan Province. *China Water and Wastewater*, *20*(8), 104–106 (in Chinese).

Wu, M. L., & Wang, Y. S. (2007). Using chemometrics to evaluate anthropogenic effects in Daya Bay, China. *Estuarine, Coastal and Shelf Science*, *72*(4), 732–742.

Wunderlin, D. A., Días, M. P., AméMaría, V., Pesce, S. F., Hued, A. C., & Bistoni, M. Á. (2001). Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia river basin (Cordoba–Argentina). *Water Research*, *35*(12), 2881–2894.

Yu, S. X., Shang, J. C., Zhao, J. S., & Guo, H. C. (2003). Factor analysis and dynamics of water quality of the Songhua River Northeast China. *Water, Air and Soil Pollution*, *144*(1–4), 159–169.

Zhou, F., Guo, H. C., Liu, Y., & Jiang, Y. M. (2007). Chemometrics data analysis of marine water quality and source identification in Southern Hong Kong. *Marine Pollution Bulletin*, *54*(6), 745–756.