Assessment of Water Quality Using Multivariate Statistical Techniques in the Ying River Basin, China

by

Lei Lei

A thesis submitted
in partial fulfillment of the requirements
for the degree of
Master of Science
(Natural Resources and Environment)
In the University of Michigan
December 2013

Thesis committee:
Professor J. David Allan, Chair
Adjunct Associate Professor Lizhu Wang

ABSTRACT

I used multivariate statistical methods, including cluster analysis (CA), discriminant analysis (DA) and principal component analysis (PCA) to evaluate water quality in the Ying River Basin, the largest tributary of Huai River, China. A total of 12 water quality parameters were measured at each of 15 sites from 2008–2010 (540 observations), allowing investigation of temporal and spatial variation and indication of potential pollution sources. Hierarchical CA classified the 15 monitoring sites into three groups, representing heavily, moderately and least polluted sites. Three parameters (temperature, pH and TP) distinguished temporal variation with close to 67.4% correct assignment in the DA, separating summer from winter and spring-fall. In the spatial variation analysis, the DA used eight parameters (temperature, pH, DO, $COD_{\mathbf{Mn}}$, $COD_{\mathbf{Cr}}$, BOD_5, NH_4-N, and Hg) and correctly assigned about 85.7% of the sites to spatial clusters. PCA did not result in a significant data reduction in this study, but it did extract and identify significant factors/variables responsible for variation in river water quality at the three groups of sites identified by CA. Sites in Group 1 were mostly correlated with COD_{Cr}, NH_4-N and volatile phenol, suggesting that they received pollutants mainly from industrial discharge. Group 2 sites correlated most strongly with temperature, pH and DO, which may indicate that these sites were mainly affected by natural processes. Group 3 sites were dominated by COD_{Mn}, As and Hg, perhaps indicating influence by both point and non-point pollution sources.

Keywords: Ying River basin, Multivariate statistical analysis, Spatial and temporal

variation, Water Quality.

ACKNOWLEDGEMENTS

I’d first like to thank my advisor Dr. David Allan for the support and motivation he has provided to me throughout the course of this project. I’m very grateful to him for sharing his knowledge in freshwater ecosystem and always be patient with me. Thank you to Dr. Xiaohong Ruan from Nanjing University (China) for collecting and providing me with water quality data and background information about the Ying River basin. Proposal of my thesis would not have been possible without the enthusiastic help of Dr. Ruan. I would also like to thank Dr. Lizhu Wang who has given suggestive comments to me on revising work of drafts. Finally, I want to thank my parents for their endless love and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

INTRODUCTION

Rivers constitute the main inland water resource for domestic, industrial and irrigation uses in many areas, and play an important role in hydrologic and biogeochemical cycles. However, few rivers are maintained in their pristine condition due to intensive human activities, and surface water pollution is today of great environmental concern worldwide (Zhao et al., 2011). Rivers are highly vulnerable water bodies because of their role in carrying off and assimilating pollutants from both point sources (e.g., municipal wastewater and industrial discharge) and non-point sources (e.g., agricultural and urban runoff, atmospheric deposition) (Carpenter et al., 1998; Ouyang et al., 2006). Municipal and industrial wastewater discharge constitutes a constant polluting source, whereas surface runoff is a seasonal phenomenon, largely affected by climate within the basin (Singh et al., 2004). Seasonal variation in precipitation, surface runoff, interflow, groundwater flow and anthropogenic transfers have a strong effect on river discharge and, subsequently, on the concentration of pollutants in river water (Vega et al., 1998). Due to these complexities, water quality specialists and decision-makers often are confronted with significant challenges in their efforts to control water pollution (Elhatip et al., 2007). By identifying spatial and temporal patterns in river water quality, an improved understanding of the environmental conditions may help managers establish priorities for sustainable water management (Bhangu et al., 1997; Antonopoulos et al., 2001; Cooper et al., 2002). Watershed-scale analysis of water quality can illustrate the changing influence of various human activities in different sub-basins and as one proceeds from headwaters to downstream reaches.

Previous studies have demonstrated that China currently faces serious water problems; not only overexploitation and uneven spatial distribution of water resources, but also severe water pollution in China's main rivers and lakes, which both contribute to the scarcity of water of adequate quantity and quality**.** Water quality at half of the regularly monitored stations in major rivers is below the Ministry of Environmental Protection standard of Grade III (suitable for the concentrated drinking water source, swimming and aquaculture), including sites along the Yangtze River, Yellow River, Pearl River, Hai River, Huai River, Liao River, and the Songhua River (Men, 2009). Annual discharge of industrial wastes and domestic sewage into the Yangtze River is over 20 billion tons, accounting for over 42% of the waste load for the entire country (Chen et al., 2009). Since1989, some 200 serious pollution events have been recorded in the Huai River basin (Zhang et al., 2010). Growing municipal and industrial wastewater discharges due to rapid urbanization and industrialization, harmful agricultural practices, along with limited wastewater treatment facility and capacity, are the principal drivers of water pollution events. About two-thirds of the total wastewater discharged by China into rivers, lakes and the sea derives from industry, and about 80% of that is untreated. Most of the untreated discharge comes from rural industries (Wang et al. 2008).

The application of different multivariate statistical techniques, such as cluster analysis

(CA), principal component analysis (PCA), factor analysis (FA) and discriminant analysis (DA), helps in the interpretation of complex data matrices to better understand the water quality and ecological status of the studied system. Such tools facilitate the identification of possible factors that influence water quality and can aid in the reliable management of water resources as well as rapid solution to pollution problems (Lee et al., 2001; Adams et al., 2001; Reghunath et al., 2002). Multivariate statistical techniques have been applied to characterize and evaluate freshwater quality, and are useful in verifying temporal and spatial variations caused by natural and anthropogenic factors linked to seasonality (Helena et al., 2000; Singh et al., 2004, 2005). Studies investigating the spatial and seasonal variability of water quality have reported that water quality issues, such as eutrophication, are highly dependent on land use patterns and the influences of watershed runoff (Yang et al., 2010; Zhang et al., 2011). Studies undertaken in Shanghai (China) and other major cities of the world have also demonstrated a significant relationship between urbanization and surface water quality (Wang et al., 2008; Duh et al., 2008). Additionally, numerous studies have identified the pollution sources and potential influences of natural processes and anthropogenic activities on spatial-temporal variation in water quality (Fan et al., 2010; Huang et al., 2010; Wang et al., 2010).

The Ying River basin, which is the largest tributary of Huai River, was selected for a water quality assessment using multivariate statistical techniques. In this study, water quality data sets obtained during 2008-2010 in the Ying River basin were analyzed using Cluster Analysis (CA), Discriminant Analysis （DA）and Principal Component Analysis (PCA). The main objectives of this study were to: (1) examine temporal and spatial variation of selected water quality parameters; (2) identify significant parameters explaining the temporal and spatial variation of water quality; and (3) attempt to identify the main factors explaining the structure of datasets.

METHODS

The study area

The Ying River basin (34°20′ - 34°34’ N, 112°45′ - 113°15’E; 30 - 1500m elevation) is located in the east-central China between the Yellow and Yangtze River basins (Figure 1), and is the largest tributary of Huai River. It originates from the Funiu Mountain area in Henan Province, flows southeast through a region of 34 cities and counties, and finally joins the main stream of Huai River in Mohekou, Anhui Province. The Ying River is approximately 557 km long and has a drainage area of 36,728 km^2 (Gao et al. 2010). Its largest tributary is the Sha River, so the basin is also known as the Shaying River basin. Other large tributaries include the Jialu, Beiru, Li, and Quan Rivers along a north to south direction. The basin is located in a transition zone between warm-temperate and sub-tropical climates and belongs to a warm-temperate, semi-moist continental climate with cold and arid winters and warm and humid summers. Its annual mean temperature ranges from 14 ℃ to 16 ℃. Its average annual precipitation is about 769.5 mm, of which more than 65% falls during a wet season from June – September and therefore contributes to high discharge in summer (Figure

2). The average annual runoff is approximately 59.2 billion m^3 and average annual runoff depth is about 145.4mm (Wang, 2000).

The Ying River basin is highly developed in China, with a population of 24 million. It flows through several major cities, including Zhengzhou, Dengfeng, Xuchang, Luohe, Pingdingshan, Zhoukou and Jishou. The basin is one of the most densely populated regions in China, with an average population density approximately 5 times the nation's average. The river serves as an important water source for agricultural irrigation, industrial use, drinking water, domestic use, and fisheries. The upper reaches of the basin have abundant resources of coal, and heavy mining activities have led to severe pollution; whereas the middle and lower reaches of the basin are important crop production areas with a total cultivated area of 12.9 million hectares. As the Sha River sub-basin is mountainous with high precipitation, floods have occurred frequently in history causing enormous losses of local residents. Within recent decades, three large reservoirs (Zhaopingtai, Baiguishan and Gushitan Reservoirs) have been constructed in the upper reaches of the Sha River to prevent floods. In addition, numerous water control gates have been constructed throughout the basin, controlling almost all of the tributaries. Historically, these dams and floodgates have benefited the region in managing water supply, irrigation, flood control, electricity generation, etc., and thus greatly promoted social and economic development. However, as a result of intensive human activity and the many dam and floodgate constructions, hydrological regimes in the basin have changed dramatically and the pollution load discharged to rivers has risen year by year.

The river and riparian environment of the Ying River basin is in poor condition because of intensive human activities (e.g. widespread flow regulation, barriers to fish movement and excessive pollutant discharge). Water quality of the Huai River basin is the worst among the nation's seven main basins, based on reporting in the Chinese Environment Bulletin in 2005 (Zhang et al., 2010). Furthermore, the Ying River is the most polluted tributary of Huai River, contributing 43% of the total amount of discharge and pollutants to the Huai River basin. Pollution in the Ying River directly influences the water quality of the main stream of Huai River. In June 1994, a severe rainstorm caused most of the dams and floodgates in the Ying River basin to be opened simultaneously to discharge floodwaters. This flood with a high concentration of pollutants resulted in severe pollution downstream, destroying fish and shrimp and severely damaging the ecology and environment along the river (Zhang et al., 2007; Jiang et al., 2011).

Data collection and analytical methods

Water quality data collected from 15 monitoring sites along the Ying River over a three-year period (2008 – 2010) were obtained from Dr. Ruan (Nanjing University, China). Sites 1-3 were located in the upper reaches of Ying River (Figure 1) within a coal-mining area and close to the Yangcheng Industrial District in Dengfeng City. Sites 4, 10 and 11 were each located downstream of three large reservoirs. Sites 5-9 were located in middle reaches of Ying River, and Site 12-13 were located in middle reaches of Sha River. Site 14 was located at the confluence of the Ying and Sha Rivers and downstream of Zhoukou City. Site 15 was located downstream of Jishou City. Information on main human activities around each monitoring site was obtained from Wikipedia and other associated websites.

Surface water samples were collected monthly from each of the sites and analyzed using standard methods (Table 1). Twelve water quality parameters (temperature, pH, dissolved oxygen, chemical oxygen demand detected by $KMnO_4$, chemical oxygen demand detected by $K_2Cr_2O_7$, 5-day biochemical oxygen demand, ammonia–nitrogen, total phosphorous, fluorides, arsenic, mercury, and volatile phenol) were selected for statistical analysis. The sampling, preservation, transportation and analysis of water samples were performed following the standard methods: Environmental Quality Standards of Surface Water (GB3838-2002), Ministry of Environmental Protection of People's Republic of China. The specific analytical methods used are presented in Table 1.

Temperature (T) is a measure of how much heat is present in water. It influences the dissolved oxygen level as the amount of oxygen dissolved in water at saturation is higher in colder water than in warm water. Temperature is also critical for freshwater organisms because it affects the rates of biochemical reactions (i.e. photosynthesis and respiration), and directly affects survival. pH measures the acidity or alkalinity level of water. Each organism adapts to a specific range of pH, so an extreme change in pH may threaten organism survival. Chemical oxygen demand (COD) and 5-day biochemical oxygen demand (BOD_5) are both measurements of the amount of organic matter in water. The difference between them is that COD also includes reductive inorganic matter and BOD_5 mainly measures biodegradable organic matter. Excessive organic matter is decomposed by bacteria and can greatly decrease oxygen levels in water, thus threatening the survival of organisms. NH_4-N measures nitrogen in the form of ammonia and ammonium in water. Total phosphorous is a measure of all forms of phosphorous, particulate and dissolved, in a water sample. Both NH_4-N and TP are basic nutrients for plant growth and excess amounts can lead to eutrophication of a water body. Fluorides, arsenic (As), mercury (Hg) and volatile phenol are chemical parameters that represent pollutants from industrial discharge. High concentrations are toxic to freshwater organisms and a threat to human health.

Data pretreatment and statistical analysis

The original data set was pretreated before conducting multivariate statistical analysis. Temperature data for site 1-3 in March and May, 2010 were missing, and were estimated using average values from data in 2008 and 2009. Observations below the limit of detection were set to zero. In order to avoid the influence of occasional extreme pollution events during the period of study, outliers were screened by making box plots and 25 data points (mainly from COD_{Mn} and NH_4-N recordings) subsequently were eliminated from the data set. Normality of the data was examined using Shapiro-Wilk's test and Q-Q plots, and natural logarithmic transformation was carried out for COD_{Mn}, COD_{Cr}, NH_4-N, As, Hg, and volatile phenol. River water quality data sets were subjected to multivariate statistical techniques: cluster analysis (CA), discriminant analysis (DA), and principal component analysis (PCA). DA was applied to raw data, whereas CA and PCA were applied to data that was standardized through z-scale transformation to avoid misclassifications arising from the different orders of magnitude of both numerical values and variance of the parameters analyzed. Mean differences among seasonal and spatial groups were examined using one-way ANOVA at a significant level of 0.05. All mathematical and statistical computations were made using SPSS Statistics (version 21) and Microsoft Office Excel 2007.

Cluster analysis (CA) is one of a large family of statistical techniques whose main purpose is to categorize entities (e.g., sampling sites) into distinct groups or clusters according to some criteria, such that the within-group similarity is maximized and among-group similarity is minimized. Hierarchical agglomerative clustering is the most common approach, which provides intuitive similarity relationships between any one sample and the entire data set (McKenna, 2003). The Euclidean distance is a commonly used distance coefficient, which usually gives the similarity between two samples and a "distance" that can be represented by the "difference" between analytical values from both the samples (Otto, 1998). The result of hierarchical clustering is typically illustrated by a dendrogram (a tree-like plot), which provides a visual summary of the agglomeration processes, depicting a picture of the clusters and their similarity, with a dramatic reduction in dimensionality of the original data set (Shrestha et al., 2007). In this study, hierarchical cluster analysis was used to classify the 15 sampling sites into groups based on characteristics of water quality, to examine the spatial pattern of water quality. The analysis was performed on normally standardized data set by means of Ward's method using squared Euclidean distance as a measure of similarity. The Ward's method uses an analysis of variance approach to evaluate the distances between clusters in an attempt to minimize the sum of squares of any two clusters that can be formed at each step. The spatial variability of water quality in the whole river basin was determined from CA, using the linkage distance, reported as Dlink/Dmax, which represents the quotient between the linkage distances for a particular case divided by the maximal linkage distance. The quotient is then multiplied by 100 as a way to standardize the linkage distance represented on the y-axis (Wunderlin et al., 2001; Simeonov et al., 2004; Singh et al., 2004).

Discriminant analysis (DA) seeks to describe the relationships among two or more pre-specified groups of sampling entities based on a set of two or more discriminating variables. DA involves deriving the linear combinations (i.e., canonical functions) of the discriminating variables that will best discriminate among groups. The canonical functions are defined as weighted linear combinations of the original variables, where each variable is weighted according to its ability to discriminate among groups. The first canonical function defines the specific linear combination of variables that maximizes the ratio of among group to within group variance in any single dimension.

It constructs a discriminant function for each group, as follows:

$$f\,(G_i) = k_i + \sum_{j=1}^{n} w_{ij} \times p_{ij}$$

where i is the number of groups (G), k_i is a constant inherent to each group, n is the number of parameters used to classify a set of data into a given group, and w_{ij} is the weight coefficient, assigned by DA to a given parameters (p_{ij}) (Johnson and Wichern 1992; Wunderlin et al. 2001; Lattin et al. 2003; Singh et al. 2004).

PCA is designed to transform the original variables into new, uncorrelated variables (axes), called principal components, which are linear combinations of the original variables. The new axes lie along the directions of maximum variance. PCA provides an objective way of finding indices of this type so that the variation in the data can be accounted for as concisely as possible (Brumelis et al., 2000). PCA provides information on the most meaningful parameters that describe the majority of the data set, affording data reduction with minimum loss of original information (Helena et al., 2000). The principal component (PC) can be expressed as:

$$z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}x_{3j} + \ldots + a_{im}x_{mj}$$

where z is the component score, a is the component loading, x the measured value of variable, i is the component number, j the sample number and m the total number of variables.

Factor analysis (FA) follows PCA. The main purpose of FA is to reduce the contribution of less significant variables to simplify even more of the data structure coming from the PCA. This purpose can be achieved by rotating the axis defined by PCA according to well established rules to construct new variables, also called varifactors (VF). A Principal Component (PC) is a linear combination of observed water quality variables, whereas a VF can include unobservable, hypothetical, latent variables (Vega et al., 1998; Helena et al., 2000). PCA analysis used normalized variables to extract significant PCs to further reduce the contribution of variables with minor significance; these PCs were subjected to varimax rotation (raw) generating VFs (Simeonova et al., 2003; Bu et al., 2010; Zhang et al., 2009). As a result, a small number of variables would usually account for approximately the same amount of information as do the much larger set of original variables. The FA can be expressed as:

$$z_{ji} = a_{f1}f_{1i} + a_{f2}f_{2i} + a_{f3}f_{3i} + \ldots + a_{fm}f_{mi} + e_{fi}$$

where z is the measured variable, a is the factor loading, f is the factor score, e the residual term accounting for errors or other source of variation, i the sample number and m the total number of factors.

RESULTS

Cluster analysis

Cluster analysis (CA) was employed to identify groups of similar monitoring sites and explore spatial heterogeneity of water quality. It generated a dendrogram, grouping the 15 sites into three distinct clusters at $(D_{link}/D_{max}) \times 100 < 40$ (Figure 3). Group 1 included sites 1- 3, located along the Ying River's upper reaches (Figure 1). Group

2 included sites 4-8 along the middle reaches of the Ying River, and sites 10-13 along the Sha River to its confluence with the Ying. Within Group 2, the three sites below reservoirs (Sites 4, 10 and 11) were clustered. Group 3 included the three lower-most sites along the Ying River, of which site 14 was just below the confluence of the Sha, Jialu and Ying Rivers. The classifications were statistically significant because the sites in these groups had similar features and human influences.

Seasonal and spatial variations of water quality

Seasonal averages computed for each of the 12 water quality variables showed distinct seasonal variation in some but not all of the measures (Figure 4). There exists significant difference ($p < 0.05$) in average temperature and dissolved oxygen among the four seasons. Temperature tends to be highest in summer and lowest in winter, and a clear inverse relationship between temperature and dissolved oxygen is observed. The average pH value is slightly higher in spring and summer than in fall and winter. The average concentrations of BOD_5, $COD_{\mathbf{Mn}}$, and TP all showed peaks in summer and then a decrease in autumn, although these differences were not significant. However, the three site groups exhibited different seasonal variation in COD_{Mn} (Figure 5). NH_4-N exhibited lower average concentrations in summer and fall, and higher average concentrations in winter and spring. In addition, strong seasonal variations were also observed in As and Hg.

Group averages for each of the 12 water quality variables were also computed and significant ($p < 0.05$) spatial variation was observed in many of the variables (Figure 6). Average temperature of Group 1 sites is significantly lower than the other two groups, presumably because these sites are located in the headwater with high elevation and relatively low air temperature. These sites may also receive discharge of groundwater, which has lower temperature than surface water. BOD_5, NH_4-N and TP have similar trends of spatial variation such that Group 1 has the highest average values, followed by Group 3, and the lowest average values appear in Group 2. DO exhibits absolutely inverse trend that Group 1 has the lowest oxygen level, which indicates that high loads of organic pollution in Group 1 sites may be depleting oxygen level below saturation.

Discriminant analysis

Temporal DA

Temporal variation in water quality was further evaluated through discriminant analysis (DA). Temporal DA was performed on the raw data after dividing the whole data set into seasonal groups (spring, summer, autumn and winter). Both standard and stepwise modes of DA were applied. In the stepwise mode, one variable that minimized the overall Wilk's Lambda statistic was entered or removed at each step. Season was the dependent variable while all monitored water quality parameters were independent variables.

As shown in Table 3, the values of Wilk's lambda and chi-square statistic for each discriminant function (DF) varied from 0.232 to 0.992 and from 13.887 to 457.916

respectively, ($p < 0.01$), indicating that the temporal DA was credible and effective. For the standard DA, the first function explained almost all ($R = 93.4\%$) of the total variance in dependent variables. A small Wilk's Lambda and a large chi-square also support this interpretation, with a p-value less than 0.01. The stepwise DA had similar results, which indicated that 98.2% of the total group differences in the data set were explained by its first DF. Therefore, the first DF alone was sufficient to explain the difference of water quality among four seasons, separating summer and winter from spring and fall (Figure 7). The stepwise DA identified three variables (temperature, pH and TP) as the most important discriminating variables and its first function was mostly correlated with temperature (coefficient = 0.949) (Table 4). Classification functions (CFs) and the classification matrices (CMs) obtained from standard and stepwise modes of DA are shown in Tables 5 and 6. In the standard mode, all variables were included to construct CFs which correctly classified 68.4% of the original grouped cases using 12 variables. In stepwise mode, the DA correctly assigned 67.4% of the cases using only three discriminating variables.

Spatial DA

Spatial variation in water quality also was evaluated using DA with groups identified by CA. The main objectives were to test the significance of discriminant functions obtained and to determine the most significant variables associated with differences among the spatial groups. The groups were the dependent variables, while all the measured water quality parameters constituted the independent variables. Both standard and stepwise modes of DA were applied.

As shown in Table 7, the values of Wilk's lambda and the chi-square for each discriminant function varied from 0.225 to 0.561 and from 148.138 to 379.892, with p-value less than 0.01, indicating that the spatial DA was credible and effective. In stepwise DA, eight variables (temperature, pH, DO, $COD_{\mathbf{Mn}}$, $COD_{\mathbf{Cr}}$, BOD_5, NH_4^+-N, and Hg) were selected as the most important discriminating variables. The two DFs explained 62.1% and 37.9% of the group differences, respectively. The first DF separated Group 1 from Groups 2 and 3 (Figure 8), and was significantly (coefficients > 0.3) correlated with pH, DO and temperature (Table 8). The second DF established some separation between Group 2 and Group 3, and was significantly correlated with COD_{Mn}, BOD_5, and NH_4-N. The CFs and CMs obtained from two modes were shown in Tables 9 and 10. In the standard mode, when all 12 variables were included, the constructed CFs produced 88.5% accuracy in assigning cases. However, in stepwise mode, DA produced 85.7% correct assignment using only eight discriminating variables.

Principle component analysis

Principal component analysis (PCA) was performed on normalized data sets (12 parameters × 15 monitoring sites) to reduce the dimensions of the original data sets and to identify latent factors affecting water quality. The number of significant principal components (PCs) was determined based on both scree plot and eigenvalue–one criterion. The eigenvalue-one criterion indicates that PCs with eigenvalues greater than one are regarded as significant when the correlation matrix is used in the analysis. In this study, PCA extracted two significant PCs with eigenvalues > 1, explaining about 76% of the total variance in corresponding water quality data sets. Varimax rotation was performed on extracted PC axes to improve the interpretation of PCA, as it increased the absolute values of larger loadings and reduced the absolute values of smaller loadings within each component. Liu et al. (2003) classified the factor loadings as "strong," "moderate," and "weak," corresponding to absolute loading values of >0.75, 0.75– 0.50, and 0.50–0.30, respectively. VF1, accounting for 46% of the total variance, had strong positive loadings on NH_4-N, TP and volatile phenol, and strong negative loadings on temperature, pH and DO. VF2, accounting for 32% of the total variance, has strong positive loadings on COD_{Mn}, As and Hg (Table 11).

Principal component loadings and scores for the first two PCs were both displayed in a scatter plot (Figure 9). The PCA demonstrated a similar clustering result for monitoring sites as CA. Three clusters of monitoring sites occupied different subspaces in the two dimensional ordination space composed by PC1 and PC2. Water quality of Sites 1, 2 and 3 (Group 1) was mostly correlated with COD_{Cr}, NH_4-N and volatile phenol. Water quality of Sites 9, 14 and 15（Group 3）were dominated by COD_{Mn}, As and Hg. Lastly, Sites 4, 5, 6, 7, 8, 10, 11 and 12 (Group 2) are mostly correlated with temperature, pH, DO.

DISCUSSION

Temporal variation of water quality

Temporal trends were observed in some water quality parameters. Notably, temperature was highest in summer and dissolved oxygen was inversely related to temperature due to its saturation relationship. Averaged across all sites, the concentrations of BOD_5 and $COD_{\mathbf{Mn}}$ also showed peaks in summer and then a decrease in autumn, and these variables may be primarily determined by temperature. Xia et al. (2002) noted that pollutants that have a high concentration during dry season and a low concentration during wet season tend to come from point sources whose supply is constant, whereas the inverse pattern can be attributed to non-point sources that are mobilized by high run-off during wet periods.

Interestingly, the three site groups exhibited quite different seasonal variation in COD_{Mn} (Figure 5). In group 1, COD_{Mn} are lowest in summer when precipitation and runoff are greatest, indicating that point source pollution of organic matters dominates in these sites. Groups 2 and 3 exhibited the reverse pattern, suggesting that COD_{Mn} is influenced mainly by non-point sources at these sites.

A pattern of low average concentrations of NH_4-N in summer and fall, and higher average concentrations in winter and spring, strongly indicates point source pollution for this parameter, which is associated with municipal discharge and animal waste from livestock farms. During spring and winter, both decreased precipitation and increased agricultural withdraws for irrigation contribute to lower flows and thus the higher concentrations of NH_4-N. Gao et al. (2010) also observed a higher NH_4-N concentration during spring and winter in the Ying River basin.

Clustering of monitoring sites and pollution source identification

Cluster analysis was successfully employed in identifying three groups of similar monitoring sites, and the results of principal component analysis additionally verified the reliability of the clustering result. Although the principle component analysis did not result in significant variable reduction in this study, it helped extract and identify significant variables responsible for variation in river water quality among the three different site groups.

As indicated by PCA, Group 1 water quality correlated most strongly with COD_{Cr}, NH_4-N and volatile phenol. Although the three sites that form Group 1 (Sites 1, 2, and 3) are located in the upper reaches of Ying River with high forest coverage, nonetheless they represent the most heavily polluted area of the watershed. Site 1 (Dajindian) located in the headwater of Ying River is an important mining area with abundant resources of coal and metals. Sites 2 (Gaocheng) is located within the Yangcheng Industry District of Dengfeng City, and Site 3 (Jiangzhuang) is just downstream of this district. Thus, high values for COD_{Cr}, NH_4-N, and volatile phenol are presumably due to industrial discharges (point sources) from the Yangcheng Industry District in Dengfeng city, where heavy industries are concentrated. The main industrial activities of this region include coal-fired power generation, aluminum fabrication, cement producing, and beneficiation (a variety of processes whereby extracted ore from mining is separated into mineral and gangue; the former is suitable for further processing or direct use). All of these generate quantities of pollutants into the environment. NH_4^+- N from industrial activities may enter water bodies through two pathways. The coal-fired power plants and cement factories emit great quantities of gases and dusts containing NH_4-N into the atmosphere, which enter waterways by atmospheric deposition. On the other hand, wastewater from coking plants contains high concentrations of NH_4-N and organic matters which are discharged directly into the river. Untreated domestic wastewater (non-point sources) also contains high loads of organic matter from human and kitchen wastes, adding to the high values for COD and NH_4-N at these sites. Volatile phenols may come from coal gas cleaning and coking process.

Group 2 includes nine sites (Figure 3) that are relatively less polluted as evidenced by the lowest mean concentration of pollutants. Sites 4, 10, and 11 in this group are located downstream of large reservoirs and exhibit the best water quality, illustrating the self-purification and assimilating function of these water bodies. The remaining Group 2 sites are located in the middle reaches of the Ying River (Sites 5-8) and Sha River (Sites 12-13), where agriculture dominates. Thus, these sites likely receive pollution mainly from non-point sources (i.e. agricultural and orchard plantation activities, and unsewered domestic wastewater). Group 2 sites are less influenced by industrial discharge, and water quality variation cannot be clearly associated with specific human activities. These sites show variation mainly in temperature, pH, and DO.

Group 3 (Sites 9, 14, and 15) corresponds to moderately polluted sites and water quality was dominated by high values for COD_{Mn}, As, and Hg. Sites 14 and 15 are situated downstream of Zhoukou and Jieshou Cities, respectively. Organic matter inputs from livestock farms, unsewered domestic wastewater, municipal sewage treatment plants, and industry discharges influence these sites to varying degrees. Animal waste and fodder from numerous livestock farms contribute organic pollutants at Sites 9 (Zhifang) and 15 (Jieshou). Sites 14 (Zhoukou) and 15 (Jieshou) have similar industrial activities and more diverse sources of organic matter, including pollutants from leather processing (mostly animal proteins and fats), food and liquor processing (starch, protein, oil, alcohol), fabrication (fats, cellulose), and printing and dyeing (lignin, cellulose and starch). The ratio of BOD_5 and COD usually serves as a measure of biodegradation of organic matter in water. Although both Group 1 and Group 3 sites have high concentrations of organic matter, Group 3 sites have higher BOD_5/COD values than Group 1, implying the sites in these two groups have different organic pollution sources. Wastewater from leather processing and dyeing industries contain high loads of arsenic (As). Plastic, pharmaceutical and chemical industries in these two cities can produce wastewater containing mercury (Hg).

Other factors influencing seasonal and spatial variations in water quality

In addition to seasonal variation and point and non-point pollution from anthropogenic activities, the water quality in the Ying River basin is also affected by other factors. As the natural watercourse of the Ying River has been interrupted by numerous dams and floodgates，the control of floods by water gates is of great significance. During the dry season when floodgates are closed to reserve water, pollutants discharged into the river are concentrated in a reduced volume of water potentially leading to a considerable increase in pollutant concentrations at some sampling sites. In the wet season, floodgates are opened when heavy storms occur in the river’s upper reaches, and water with accumulated pollutants will flow downstream, causing severe pollution incidents in lower reaches. As pollutants are exported, water quality within the Ying River basin may subsequently improve. In recent years, three severe water pollution incidents (1989, 1994, and 2004, respectively) in the Huai River basin were all caused by concentrated pollutants flowing down through the Ying River system (Zhang et al. 2007). As a result, researchers and managers are now developing strategies on how to operate multiple dams and floodgates in a coordinated manner within the entire Huai River basin.

Jialu River is one of the most polluted tributaries of the Ying River, although it was not included in this study. Gao et al. (2010) reported that the values for NH_4-N, TN, TP, and COD_{Mn} in Jialu River are higher than in Sha River and upper Ying River. Site 14 (Zhoukou) located downstream of the confluence where the Jialu River joins the

Ying River, and Ying River may have been strongly influenced by the pollutants from the Jialu River.

CONCLUSION

Multivariate statistical methods were successfully applied in this study to evaluate temporal and spatial variation in river water quality and to identify possible anthropogenic sources of water quality patterns at monitoring sites in the Ying River basin. The results are useful for river water quality management. Hierarchical CA grouped 15 monitoring sites into three groups based on their similarity of water quality characteristics, thus providing a useful classification of the surface watercourses that can be used for optimizing a future spatial monitoring network in the basin with lower costs. For example, the number of monitoring sites could be reduced by selecting only one site from each of the three groups. Furthermore, the pollution of Group 1 and Group 3 sites is relatively serious and should be controlled.

Pollution in the Ying River basin likely derives from three sources: (1) excess industrial discharge of different types (paper making, food processing, cement producing, metallurgy, leather processing, fabrication, coking etc.); (2) increased pollution from large-scale livestock farms, and likely pesticides and chemical fertilizers used in farmlands; (3) municipal and domestic sewage from a dense population and limited wastewater treatment facilities in less developed areas.

TABLES

Table 1. Units, analytical methods, and detection limit of water quality parameters monitored in the Ying River basin from 2008 – 2010.

Parameter	Abbreviation	Unit	Method	Detection Limit (mg/L)
Temperature	T	℃	Thermometer	
pH	pH		Glass electrode method	
Dissolved oxygen	DO	mg/L	Iodometric method	0.2
Chemical oxygen demand detected by $KMnO_4$	COD_{Mn}	mg/L	Titration method	0.5
Chemical oxygen demand detected by $K_2Cr_2O_7$	COD_{Cr}	mg/L	Dichromate method	10
5-day biochemical oxygen demand	BOD_5	mg/L	Dilution and seeding test	2
Ammonia - nitrogen	NH_4 - N	mg/L	Nessler's reagent spectrophotometry	0.05
Total phosphorous	TP	mg/L	Ammonium molybdate spectrophotometric method	0.01
Fluorides	Fluorides	mg/L	Fluorine reagent spectrophotometry	0.05
Arsenic	As	mg/L	Cold atomic fluorescent spectrophotometry	0.00006
Mercury	Hg	mg/L	Cold atomic absoption spectrophotometry	0.00005
Volatile phenol	Volatile phenol	mg/L	4-AAP spectrophotometric method	0.002

Table 2. The means and standard deviations for twelve water quality parameters measured monthly at 15 sites from 2008-2010. S.D= 1 standard deviation. See Table 1 for parameter abbreviations.

Parameters		Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8
T	Mean	10.83	10.70	10.70	16.20	17.46	17.55	17.49	17.76
	S.D	7.23	6.96	7.14	9.15	8.73	8.83	8.52	8.88
pH	Mean	7.43	7.43	7.45	7.78	7.86	7.87	7.75	7.96
	S.D	0.35	0.35	0.36	0.31	0.24	0.23	0.22	0.30
DO	Mean	4.93	6.21	5.62	8.23	7.60	6.83	7.57	8.00
	S.D	2.92	1.41	1.72	1.84	1.31	1.57	1.65	1.97
COD_{Mn}	Mean	5.10	4.48	3.86	2.46	2.63	3.79	3.46	3.16
	S.D	2.05	1.06	1.53	0.58	0.63	0.92	0.56	0.94
COD_{Cr}	Mean	41.58	29.59	31.25	18.65	23.42	29.38	26.07	26.32
	S.D	27.43	21.05	25.08	4.33	4.09	12.59	6.87	10.42
BOD_5	Mean	11.57	11.56	7.83	2.61	2.97	3.12	2.65	2.87
	S.D	7.10	8.33	5.80	1.00	1.24	1.78	0.72	1.38
NH_4-N	Mean	3.21	4.42	1.73	0.12	0.11	0.20	0.12	0.14
	S.D	3.43	4.43	2.06	0.060	0.052	0.068	0.067	0.054
TP	Mean	0.46	0.39	0.26	0.048	0.077	0.098	0.070	0.068
	S.D	0.40	0.48	0.24	0.024	0.085	0.052	0.033	0.059
Fluorides	Mean	0.75	0.89	0.89	0.84	0.56	0.60	0.62	0.56
	S.D	0.29	0.17	0.19	0.20	0.15	0.13	0.15	0.14
As	Mean	0.0021	0.0015	0.0015	0.00035	0.00037	0.0013	0.00047	0.0018
	S.D	0.0040	0.0014	0.00083	0.00070	0.00065	0.0037	0.00098	0.0034
Hg	Mean	0.00047	0.000058	0.000042	0.000024	0.000026	0.000030	0.000030	0.000027
	S.D	0.00010	0.00014	0.000084	0.0000015	0.0000045	0.000012	0,000022	0.000013
Volatile	Mean	0.0097	0.065	0.0098	0.00096	0.00098	0.00097	0.0010	0.00098
Phenol	S.D	0.0088	0.31	0.014	0.00017	0.00013	0.00014	0.00022	0.00013

(continued Table 2)

Parameters		Site 9	Site 10	Site 11	Site 12	Site 13	Site 14	Site 15
T	Mean	17.34	17.22	17.32	16.89	17.04	16.80	17.43
	S.D	8.97	9.08	8.90	8.74	8.81	9.04	9.08
pH	Mean	7.69	7.86	7.86	7.84	7.83	7.70	7.61
	S.D	2.27	0.35	0.32	0.23	0.17	0.23	0.28
DO	Mean	6.33	7.29	7.25	7.33	7.57	8.65	7.22
	S.D	2.83	0.90	0.82	1.85	2.02	1.88	0.96
COD_{Mn}	Mean	7.84	2.46	2.65	3.32	3.15	4.41	5.66
	S.D	3.29	0.76	0.78	0.84	0.81	1.78	1.38
COD_{Cr}	Mean	25.31	7.05	9.67	34.62	30.10	12.08	23.62
	S.D	13.57	3.61	4.98	23.41	17.03	7.04	7.66
BOD_5	Mean	12.91	2.44	2.16	2.57	2.81	6.03	1.13
	S.D	8.95	1.22	0.91	0.91	1.09	3.43	0.25
NH_4-N	Mean	1.58	0.091	0.078	0.76	0.42	0.44	2.09
	S.D	1.58	0.037	0.051	0.67	0.39	0.37	1.81
TP	Mean	0.33	0.056	0.050	0.14	0.10	0.11	0.26
	S.D	0.26	0.042	0.028	0.13	0.057	0.13	0.11
Fluorides	Mean	0.91	0.59	0.54	0.61	0.58	0.72	0.82
	S.D	0.26	0.15	0.12	0.13	0.14	0.32	0.14
As	Mean	0.0039	0.00028	0.00030	0.00056	0.00051	0.0023	0.0033
	S.D	0.0043	0.00049	0.00056	0.0012	0.0013	0.0033	0.0032
Hg	Mean	0.00017	0.000028	0.000030	0.000035	0.000032	0.00016	0.000024
	S.D	0.00035	0.000011	0.000016	0.000023	0.000022	0.00038	0.0000043
Volatile	Mean	0.0046	0.00098	0.00098	0.00097	0.00098	0.0010	0.00096
Phenol	S.D	0.0079	0.00015	0.00012	0.00014	0.00010	0.00000	0.00025

Table 3. Wilk's lamda and chi-square test for the discriminant analysis of temporal variation in water quality across four seasons.

Mode	Function	R	Eigenvalue	Wilk's lambda	Chi-square	p-level
Standard mode	1	93.4	2.618	.232	457.916	0.000
	2	4.4	0.125	.838	55.442	0.000
Stepwise mode	1	98.2	2.475	.275	409.351	0.000
	2	1.4	0.036	.957	13.887	0.00

Table 4. Structure matrix for the discriminant analysis of Table 3.

Standard Mode		**Stepwise Mode**	
Parameters	**Function 1**	**Parameters**	**Function 1**
As	.140	As	.208
BOD_5	.053	BOD_5	.061
COD_{Cr}	.005	COD_{Cr}	.038
COD_{Mn}	.052	COD_{Mn}	.127
DO	-.236	DO	-.125
Fluorides	.012	Fluorides	.051
Hg	.032	Hg	-.011
NH_4-N	-.088	NH_4-N	-.028
pH	-.041	pH	-.043
Temperature	**.923**	Temperature	**.949**
TP	.044	TP	.045
Volatile phenol	-.065	Volatile phenol	-.105

Table 5. Classification function coefficients for the discriminant analysis (DA) of Table 3.

Parameters	Standard mode DA				Stepwise mode DA			
	Spring	Summer	Fall	Winter	Spring	Summer	Fall	Winter
Temperature	.228	.735	.371	-.200	-.035	.484	.129	-.438
pH	82.387	80.521	80.910	82.388	71.274	69.078	69.328	71.112
DO	1.045	.827	.964	1.170				
COD_{Mn}	-.243	-.271	-.278	-.201				
COD_{Cr}	.182	.191	.197	.193				
BOD_5	.019	.036	.034	-.002				
NH_4-N	.161	-.150	-.123	.029				
TP	11.171	13.937	10.852	10.050	21.859	24.260	21.609	20.089
Fluorides	30.805	31.860	31.422	33.055				
As	-203.255	-269.951	-36.939	-144.733				
Hg	-10161.995	-10674.717	-11977.221	-12683.404				
Volatile phenol	909.062	903.641	942.536	891.810				
(Constant)	-339.734	-334.801	-330.678	-337.947	(Constant)	-279.253	-273.090	-266.876

Table 6. Classification matrix for the discriminant analysis (DA) of Table 3.

Monitoring seasons	Percent correct (%)	Seasons assigned by DA			
		Spring	Summer	Fall	Winter
Standard mode					
Spring	47.2	68	20	24	32
Summer	88.2	3	127	14	0
Fall	45.1	30	47	65	2
Winter	93.1	5	0	5	134
Total	68.4	106	194	108	168
Stepwise mode					
Spring	46.5	68	20	24	32
Summer	84.7	3	127	14	0
Fall	46.5	30	47	65	2
Winter	91.7	5	0	5	134
Total	67.4	110	176	125	165

Table 7. Wilk's lamda and chi-square test for a discriminant analysis of spatial variation in water quality across three groups of sites.

Mode	Function	R	Eigenvalue	Wilk's lambda	Chi-square	p-level
Standard mode	1	60.3	1.353	.225	379.892	0.000
	2	39.7	0.890	.529	162.069	0.000
Stepwise mode	1	62.1	1.281	.246	359.687	0.000
	2	37.9	0.782	.561	148.138	0.000

Table 8. Structure matrix for a discriminant analysis of Table 7.

	Standard Mode			Stepwise Mode	
Parameters	**Function 1**	**Function 2**	**Parameters**	**Function 1**	**Function 2**
As	.097	.269	As	-.006	.135
BOD_5	-.301	**.469**	BOD_5	-.325	**.484**
COD_{Cr}	-.235	.064	COD_{Cr}	-.243	.056
COD_{Mn}	.123	**.694**	COD_{Mn}	.103	**.747**
DO	**.360**	-.062	DO	**.372**	-.047
Fluorides	-.092	**.332**	Fluorides	-.179	.132
Hg	.154	.211	Hg	.151	.234
NH_4-N	-.340	**.390**	NH_4-N	-.363	**.398**
pH	**.397**	-.280	pH	**.417**	-.277
Temperature	**.354**	-.135	Temperature	**.368**	-.125
TP	-.219	**.443**	TP	-.202	.255
Volatile phenol	-.110	.042	Volatile phenol	.071	.010

Table 9. Classification function coefficients for a discriminant analysis of Table 7.

	Standard Mode			Stepwise mode		
	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
Temperature	.872	1.047	1.089	.781	.958	.992
pH	113.780	119.451	117.207	107.982	113.566	110.956
DO	3.472	4.007	4.571	3.016	3.601	4.110
COD_{Mn}	-.341	-.278	.821	-.008	.009	1.147
COD_{Cr}	.463	.483	.400	.557	.576	.500
BOD_5	-.380	-.552	-.569	-.073	-.256	-.239
NH_4-N	-.904	-1.209	-1.059	-.753	-1.137	-.929
TP	23.084	20.628	23.242			
Fluorides	25.878	25.009	27.538			
As	297.451	302.659	455.172			
Hg	-13394.829	-12488.898	-9698.953	-16641.976	-15603.749	-12315.586
Volatile phenol	-38.387	-42.422	-42.729			
(Constant)	-454.190	-501.135	-494.665	-420.971	-468.414	-456.505

Table 10. Classification matrix for a discriminant analysis (DA) of Table 7.

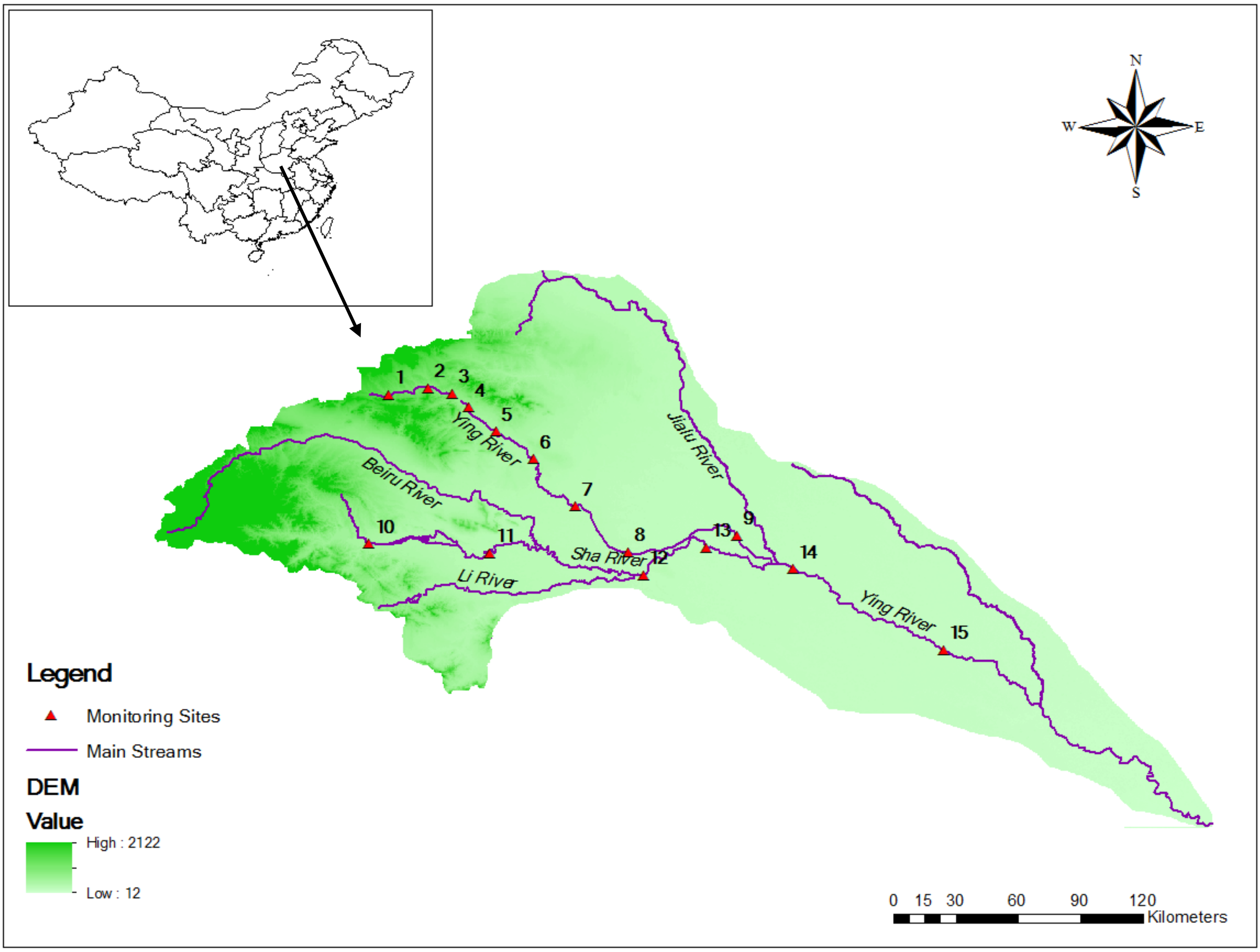
Monitoring sites	Percent correct (%)	Regions assigned by DA		
		Group 1	Group 2	Group 3
Standard mode				
Group 1	78.7	85	11	12
Group 2	95.1	9	308	7
Group 3	78.7	9	14	85
Total	88.5	103	333	104
Stepwise mode				
Group 1	76.9	83	15	10
Group 2	92.0	9	298	17
Group 3	76.0	8	18	82
Total	85.7	100	331	109

Table 11. Loadings of water quality variables on significant principal components.

Water quality variables	Rotated Components	
	VF1	VF2
Temperature	**-0.95**	0.00
pH	**-0.86**	-0.34
DO	**-0.81**	-0.21
COD_{Mn}	0.26	**0.93**
COD_{Cr}	0.63	0.03
BOD_5	0.66	0.62
NH_4-N	**0.89**	0.33
TP	**0.82**	0.53
Fluorides	0.54	0.63
As	0.15	**0.91**
Hg	-0.07	**0.91**
Volatile phenol	**0.76**	0.05
Eigenvalue	5.57	3.85
% of Total variance	46%	32%
Cumulative % variance	46%	78%

FIGURES

Figure 1. Location of monitoring sites in the Ying River basin, China.



Monitoring Sites:

1. **Dajindian**
2. **Gaocheng**
3. **Jiangzhuang**
4. **Baisha Reservoir**
5. **Yuzhou**
6. **Yingyang**
7. **Huaxing**
8. **Wuliu**
9. **Zhifang**
10. **Zhaopingtai Reservoir**
11. **Baiguishan Reservoir**
12. **Yancheng**
13. **Luohe**
14. **Zhoukou**
15. **Jieshou**

Figure 2. Monthly mean runoff of Jieshou section (Site 15), 2008 – 2010.

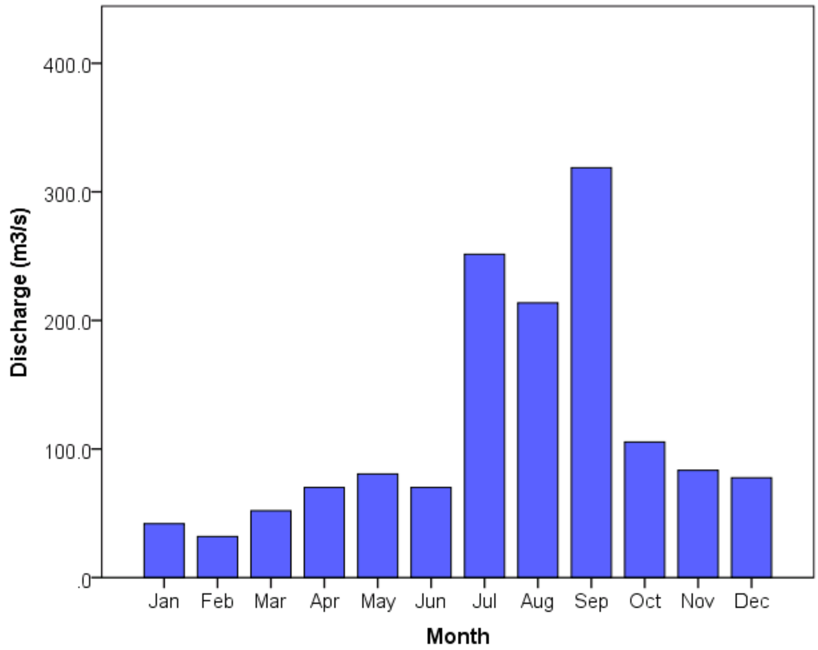


Figure 3. Dendrogram showing spatial clustering of monitoring sites.

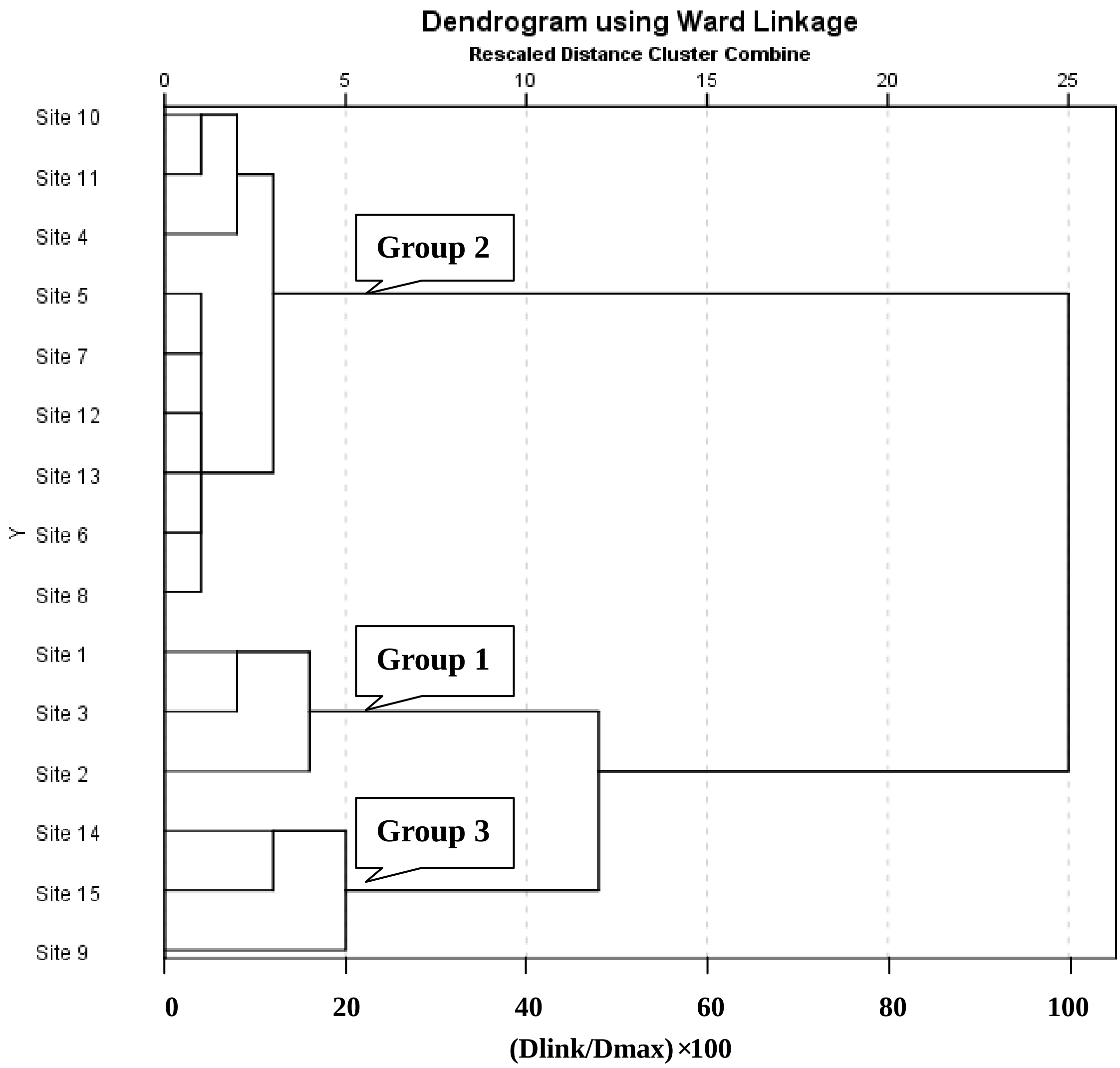
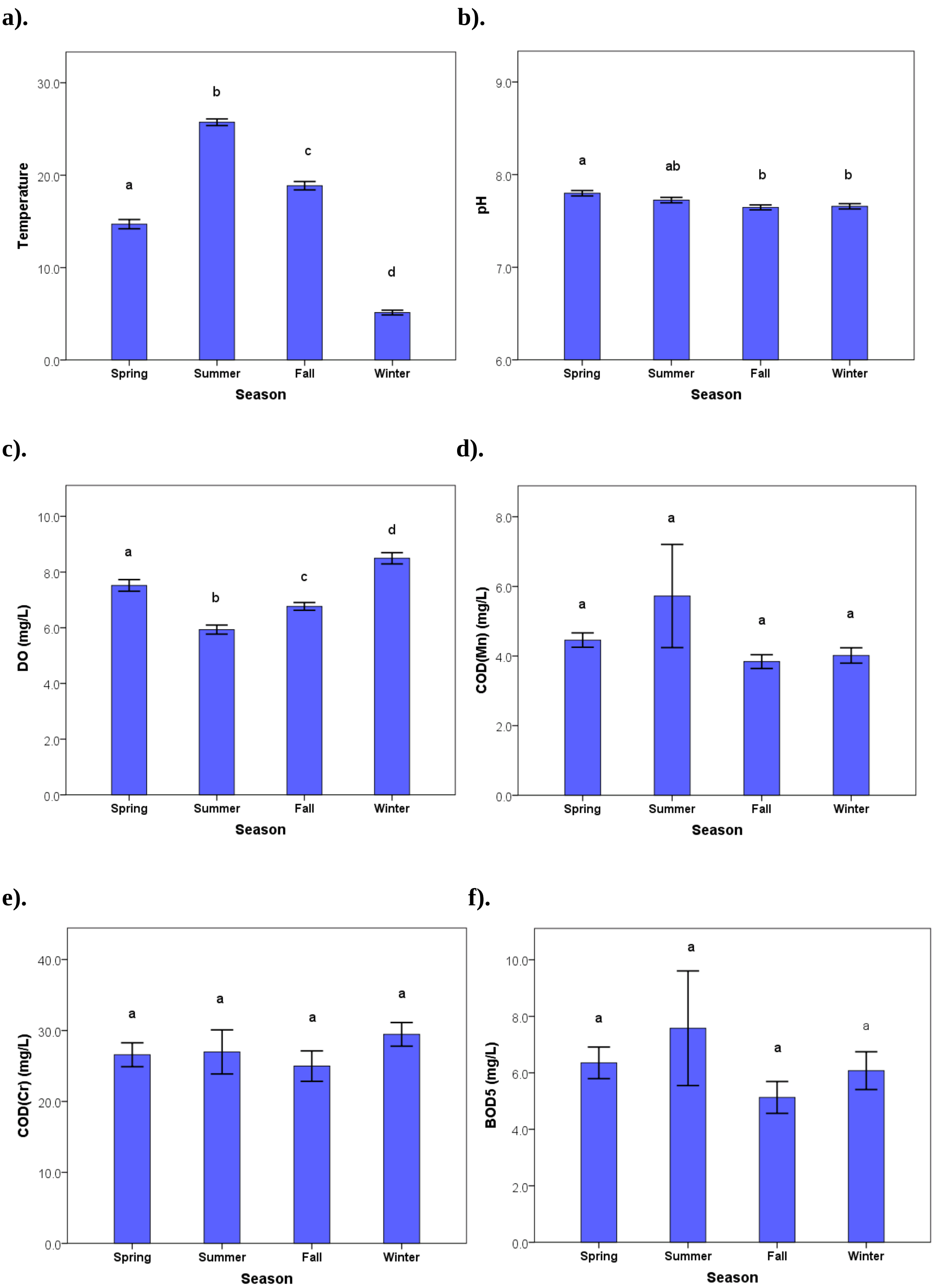
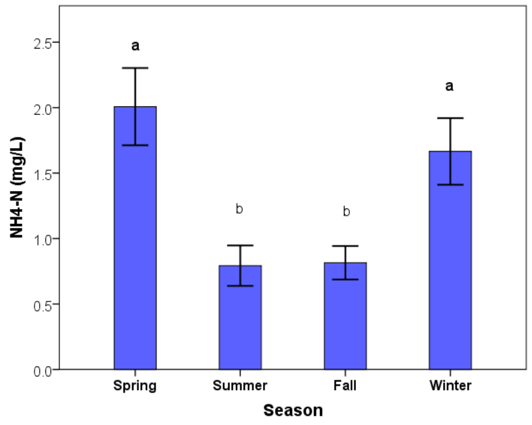


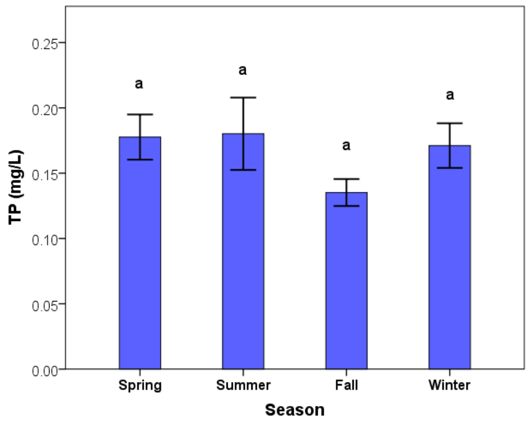
Figure 4. Bar plots with means and standard errors for all parameters, showing seasonal variation at a significant level of 0.05.



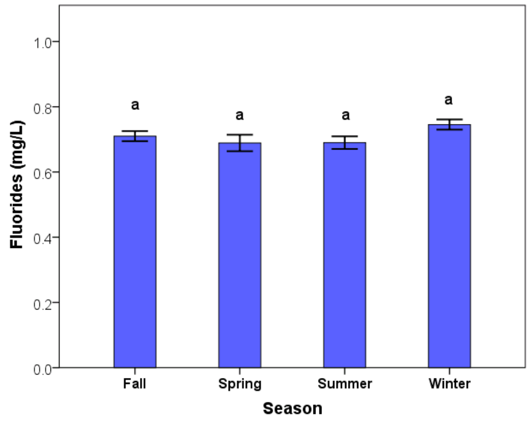
g).


NH4-N (mg/L)
2.5
2.0
1.5
1.0
0.5
0.0
a
b
b
a
Spring
Summer
Fall
Winter
Season

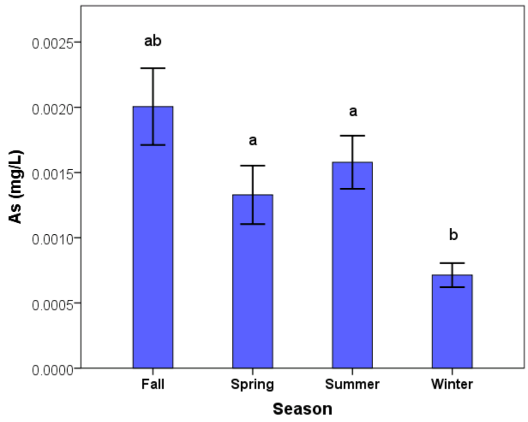
h).


TP (mg/L)
0.25
0.20
0.15
0.10
0.05
0.00
a
a
a
a
Spring
Summer
Fall
Winter
Season

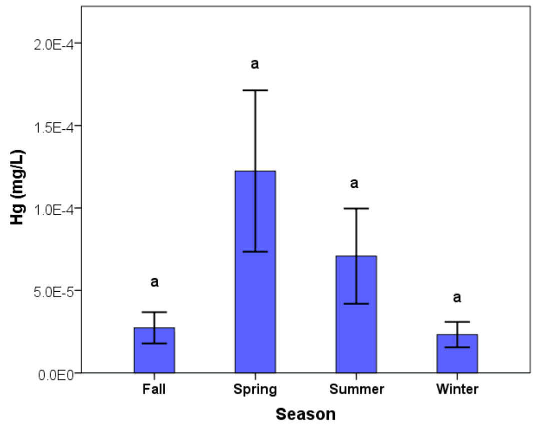
i).


Fluorides (mg/L)
1.0
0.8
0.6
0.4
0.2
0.0
a
a
a
a
Fall
Spring
Summer
Winter
Season

j).


As (mg/L)
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
ab
a
a
b
Fall
Spring
Summer
Winter
Season

k).


Hg (mg/L)
2.0E-4
1.5E-4
1.0E-4
5.0E-5
0.0E0
a
a
a
a
Fall
Spring
Summer
Winter
Season

l).

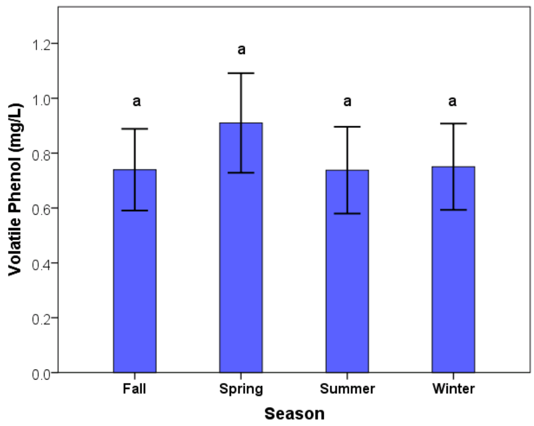
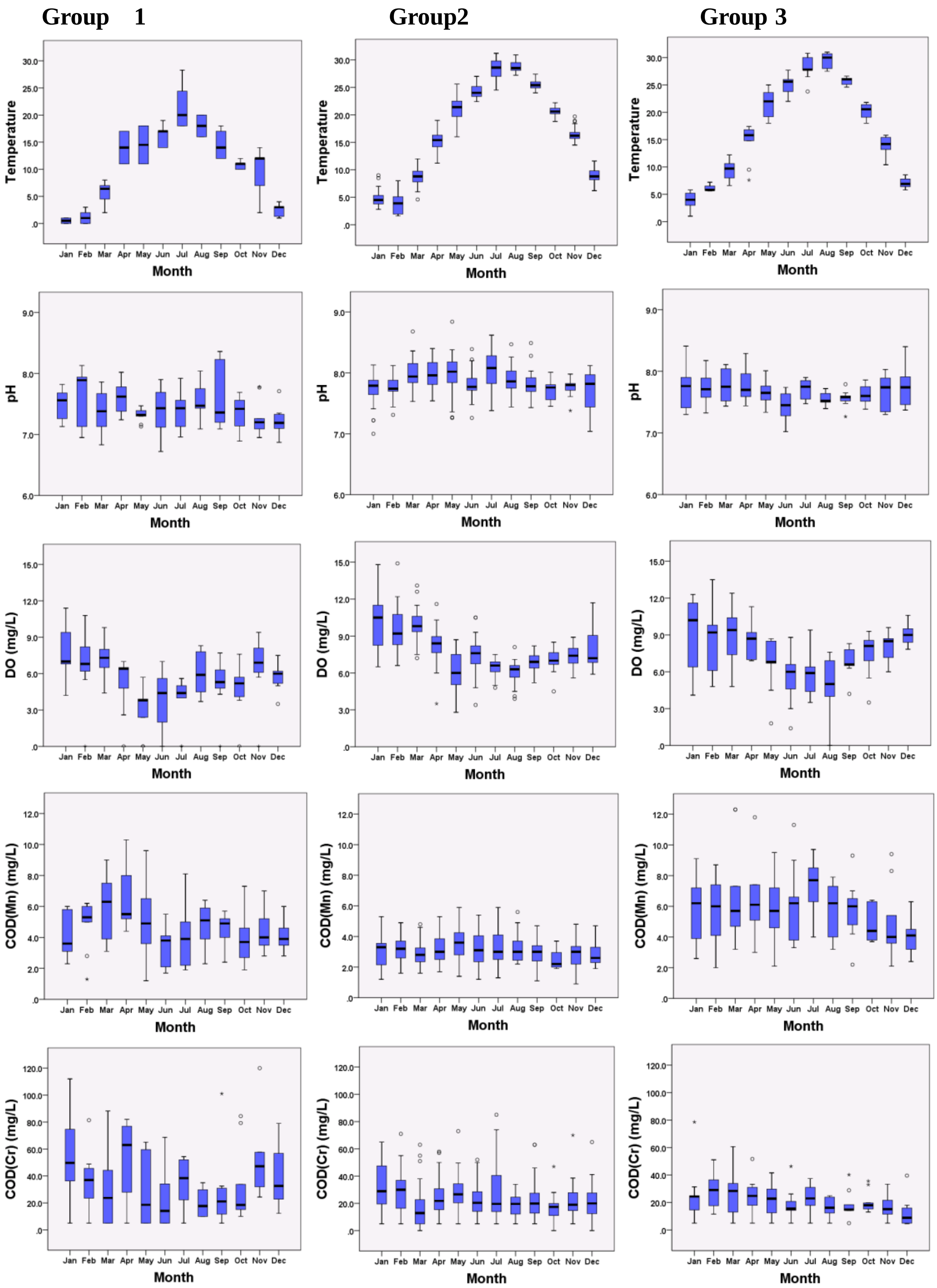
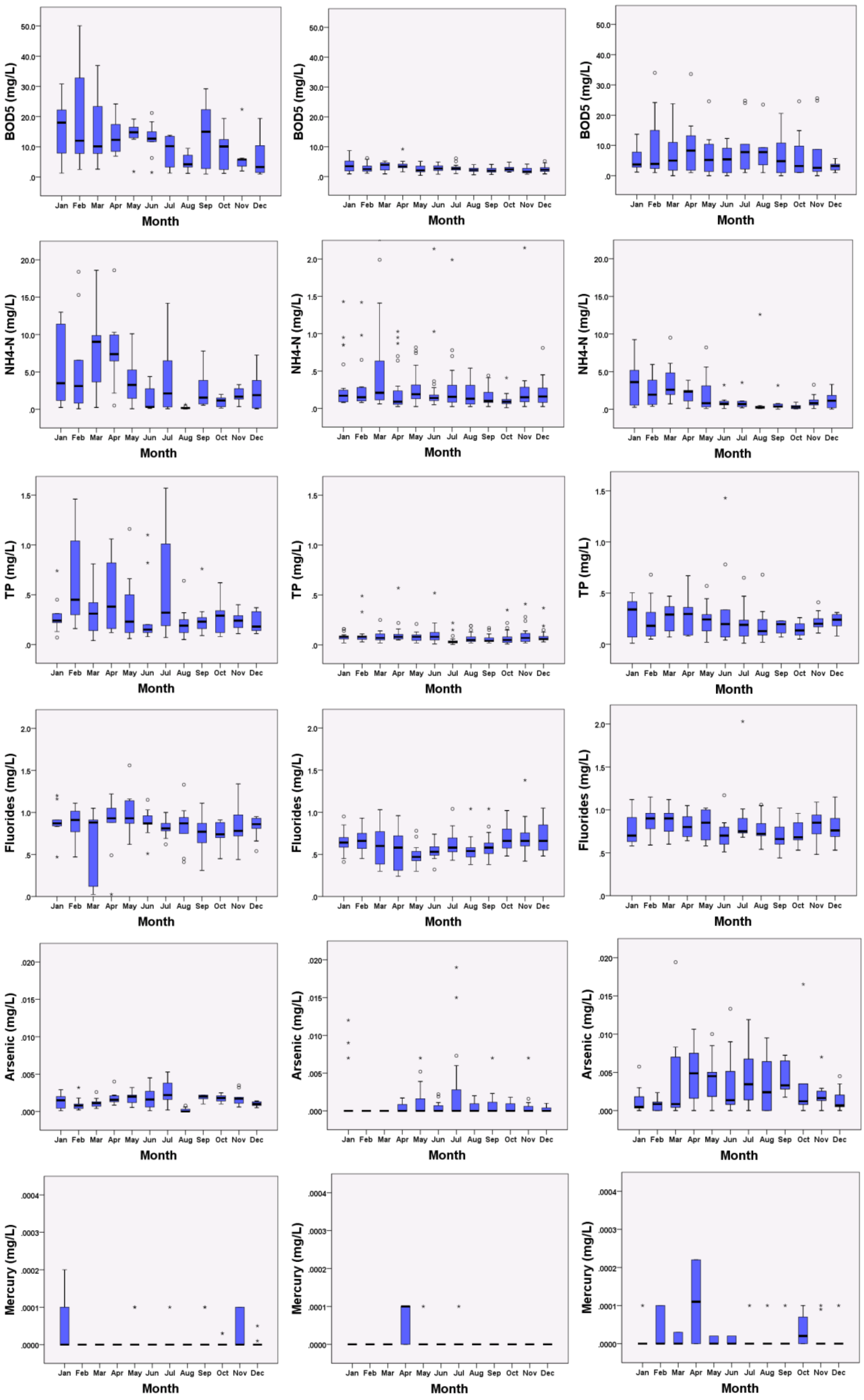

Volatile Phenol (mg/L)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
a
a
a
a
Fall
Spring
Summer
Winter
Season

Figure 5. Seasonal variation in water quality for the three sites groups.




BOD5 (mg/L)
NH4-N (mg/L)
TP (mg/L)
Fluorides (mg/L)
Arsenic (mg/L)
Mercury (mg/L)
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Month

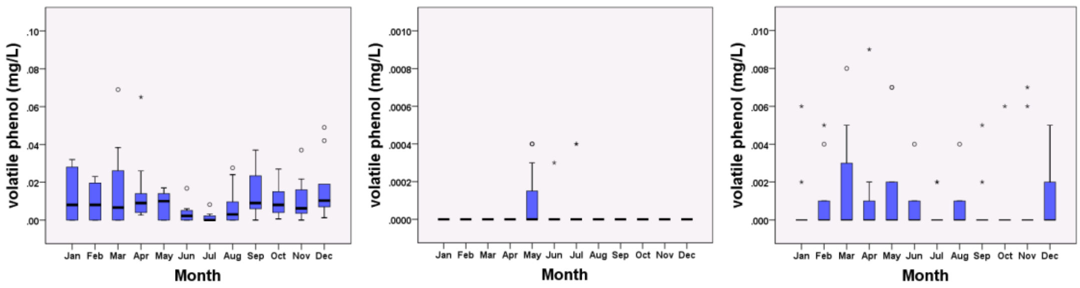
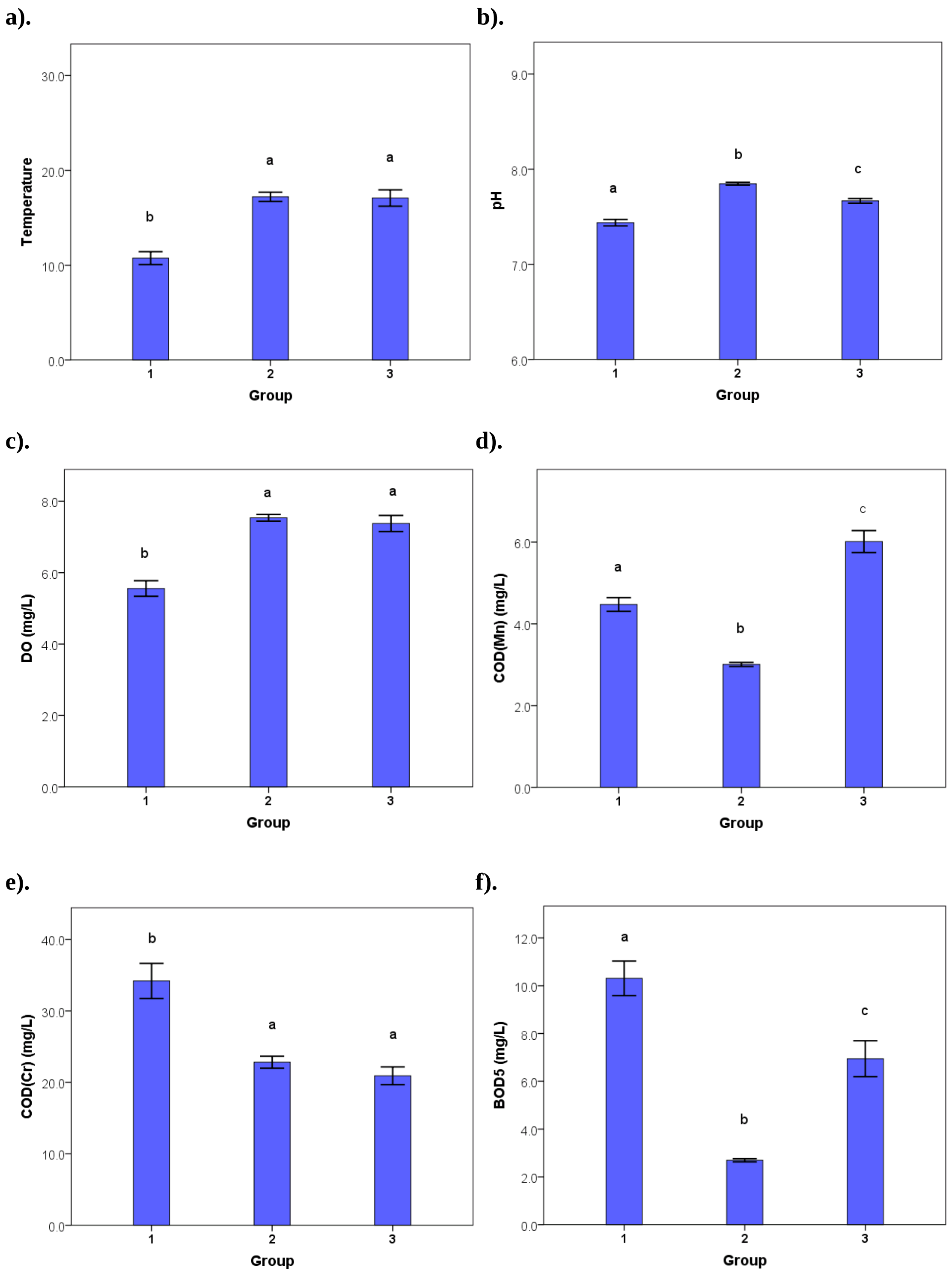
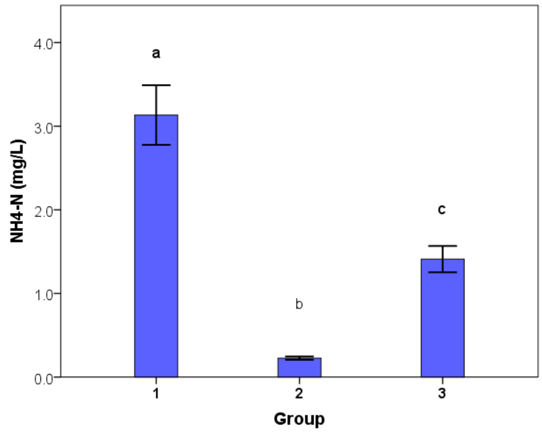

volatile phenol (mg/L)
Month
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

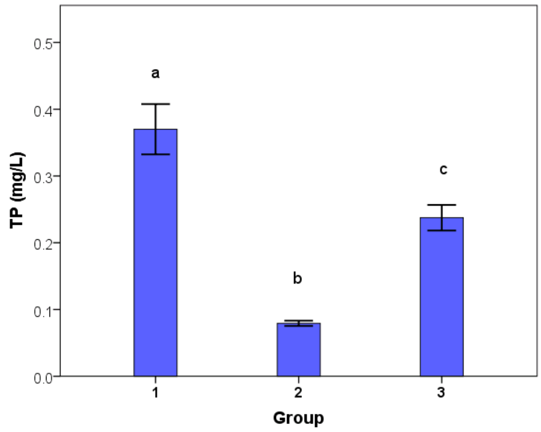
Figure 6. Bar plots with mean values and standard errors for all parameters, showing spatial variation at a significant level of 0.05.



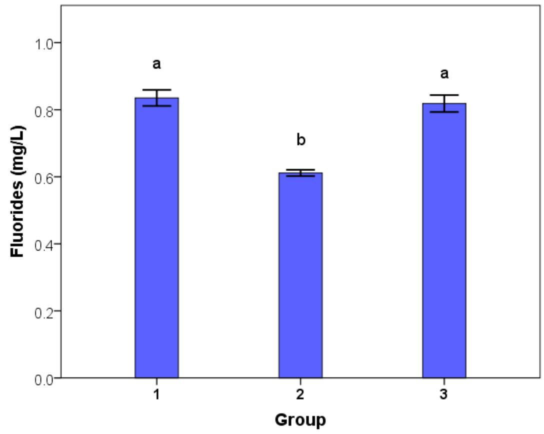
g).


NH4-N (mg/L)
4.0
3.0
2.0
1.0
0.0
a
b
c
1
2
3
Group

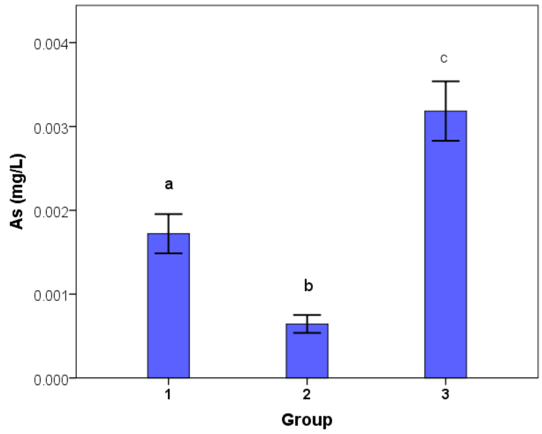
h).


TP (mg/L)
0.5
0.4
0.3
0.2
0.1
0.0
a
b
c
1
2
3
Group

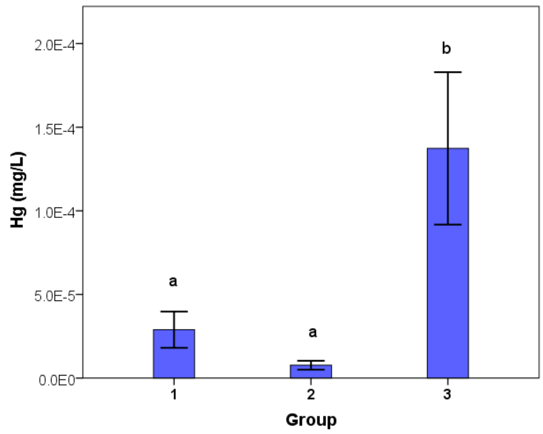
i).


Fluorides (mg/L)
1.0
0.8
0.6
0.4
0.2
0.0
a
b
a
1
2
3
Group

j).


As (mg/L)
0.004
0.003
0.002
0.001
0.000
a
b
c
1
2
3
Group

k).


Hg (mg/L)
2.0E-4
1.5E-4
1.0E-4
5.0E-5
0.0E0
a
a
b
1
2
3
Group

l).

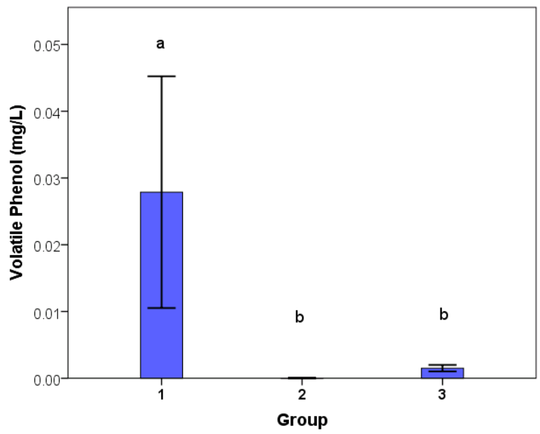

Volatile Phenol (mg/L)
0.05
0.04
0.03
0.02
0.01
0.00
a
b
b
1
2
3
Group

Figure 7. Scatter plot for the discriminant analysis of temporal variation in water quality across four seasons (stepwise mode). In the plot: 1 – spring, 2 – summer, 3 – fall, 4 – winter.

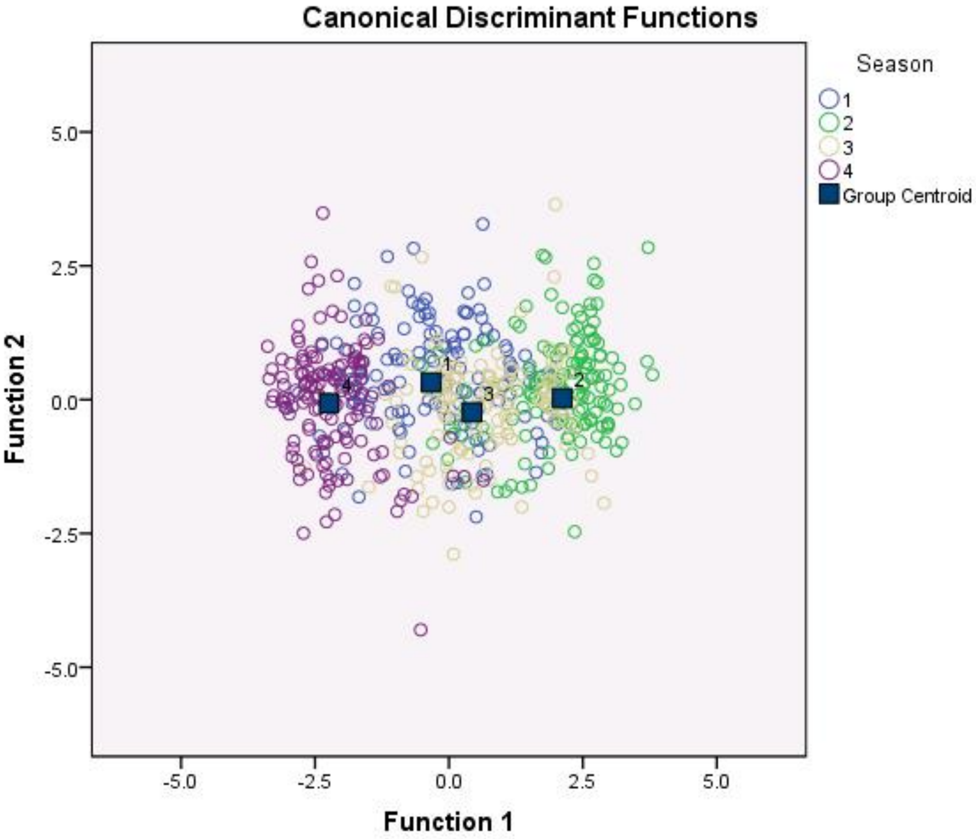


Figure 8. Scatter plot for the discriminant analysis of spatial variation in water quality across 3 sites groups (stepwise mode).

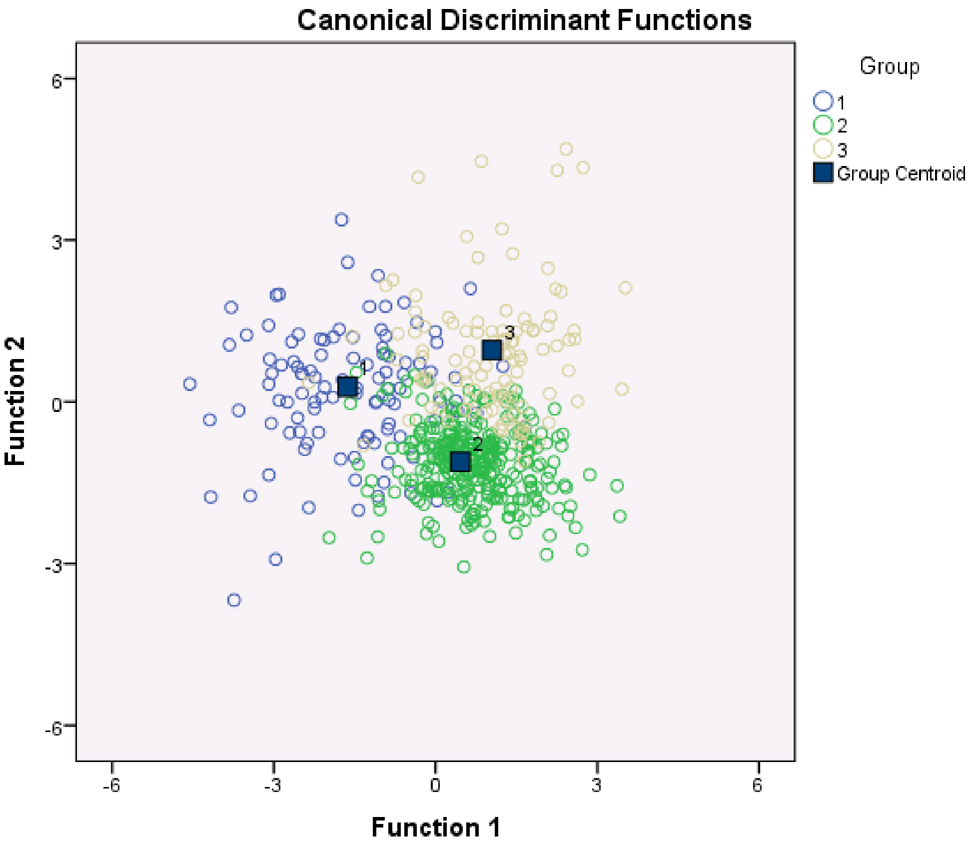
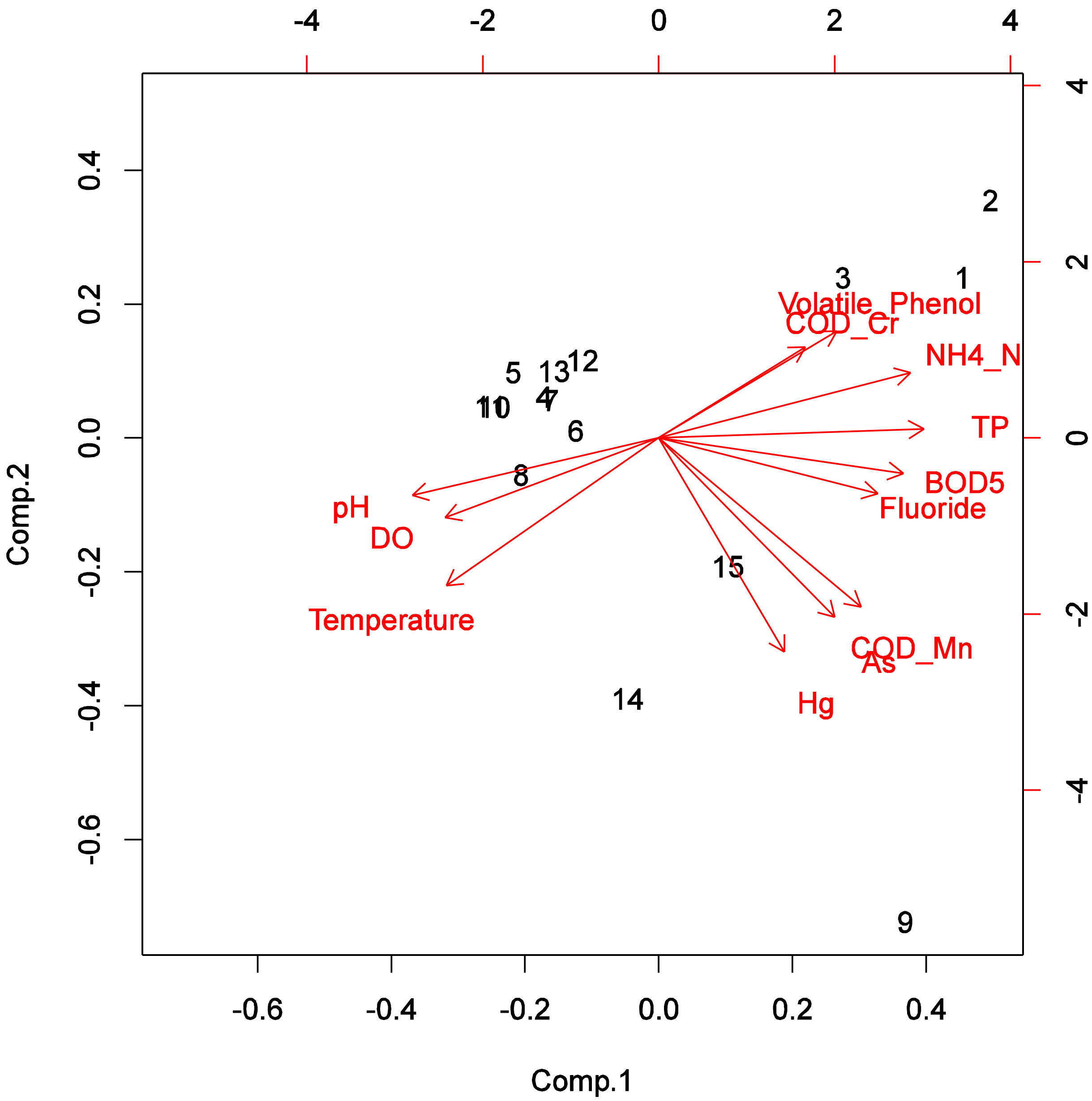


Figure 9. Scatter plot of loadings and scores of PCA. In the plot, numbers 1 – 15 correspond to monitoring Sites 1 – 15.



LITERATURE CITED

Adams, S., Titus, R., Pietesen, K., Tredoux, G., Harris, C., 2001. Hydrochemical characteristic of aquifers near Sutherland in the Western Karoo, South Africa. Journal of Hydrology 241: 91-103.

Antonopoulos, V. Z., Papamichail, D. M., and Mitsiou, K. A., 2001. Statistical and trend analysis of water quality and quantity data for the Strymon River in Greece, Hydrology and Earth System Sciences 5: 679–692.

Bhangu, I., and Whitfield, P. H.,1997. Seasonal and long-term variations in water quality of the Skeena River at Usk, British Columbia, Water research 31: 2187–2194.

Brumelis, G., Lapina, L., Nikodemus, O., Tabors, G., 2000. Use of an artificial model of monitoring data to aid interpretation of principal component analysis. Environmental Modelling & Software 15 (8): 755-763.

Bu, H. M., Tan, X., Li, S. Y., Zhang Q. F., 2010. Temporal and spatial variations of water quality in the Jinshui River of the South Qinling Mts., China. Ecotoxicology and Environmental Safety 73: 907-913.

Carpenter, S.R., Caraco, N.F., Correll, D.L., Howarth, R.W., Sharpley, A.N., Smith, V.H., 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecological Applications 8: 559–568.

Chen, D. Q., F. Xiong, K. Wang & Y. H. Wang, 2009. Status of research on Yangtze fish biology and fisheries. Environmental Biology of Fishes 85: 337–357.

Cooper, D. M., House, W. A., May, A., and Gannon, B., 2002. The phosphorus budget of the Thame catchment, Oxfordshire, UK: 1. Mass balance, The Science of the Total Environment 282: 233–251.

Duh, J., Shandas, V., Chang, H., George, L., 2008. Rates of urbanization and the resilience of air and water quality. The Science of the Total Enviroment 400: 238–256.

Elhatip, H., Hinis, M. A., and Gulgahar, N., 2007. Evaluation of the water quality at Tahtali dam watershed in Izmir, Turkey by means of statistical methodology, Stochastic Environmental Research and Risk Assessment 22: 391–400.

Fan, X.Y., Cui, B.S., Zhao, H., Zhang, Z.M., Zhang, H.G., 2010. Assessment of river water quality in Pearl River Delta using multivariate statistical techniques. Procedia Environmental Sciences 2: 1220-1234.

Gao, H. L., Li, H. T., Zhao, F. L., 2010. Temporal and spatial variation of water pollution in Shaying River (Henan section). Water Resources Protection 26(3): 23-26.

Helena, B., Pardo, R., Vega, M., Barrado, E., Ferna ńdez, J.M., Ferna ńdez, L., 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis. Water Research 34: 807-816.

Huang, F., Wang, X.Q., Luo, L.P., Lou, L.P., Zhou, Z.Q., Wu, J.P., 2010. Spatial variation and source apportionment of water pollution in Qiantang River (China) using statistical techniques. Water Research 44:1562-1572.

Jiang, Y., Peng, Q. D., Luo, H. H., Ma, W., 2011. Analysis of spatial and temporal variation of water quality in Huaihe River Basin. ShuiLi XueBao 42(11): 1283-1288.

Johnson, R.A., Wichern, D.W., 1992. Applied Multivariate Statistical Analysis. Prentice-Hall, Englewood Cliffs, NJ.

Lattin, J., Carroll, D., & Green, P. (2003). Analyzing Multivariate Data. New York: Duxbury.

Lee, J.Y., Cheon, J.Y., Lee, K.K., Lee, S.Y., Lee, M.H., 2001. Statistical evaluation of geochemical parameter distribution in a ground water system contaminated with petroleum hydrocarbons. Journal of Environmental Quality 30: 1548-1563.

McKenna Jr., J.E., 2003. An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. Environmental Modelling & Software 18 (3): 205-220

Men, W., 2009. System engineering for water pollution control at the watershed level in China. Frontiers of Environmental Science and Engineering in China 3(4): 443–452.

Otto, M., 1998. Multivariate methods. In: Kellner, R., Mermet, J.M., Otto, M., Widmer, H.M. (Eds.), Analytical Chemistry. WileyeVCH, Weinheim.

Ouyang Y., Nkedi-Kizza, P., Wu, Q. T., Shinde, D., and Huang. C. H., 2006. Assessment of seasonal variations in surface water quality," Water Research 40: 3800-2610.

Reghunath, R., Murthy, T.R.S., Raghavan, B.R., 2002. The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. Water Research 36: 2437-2442.

Shrestha, S., and Kazama, F., 2006. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan.

Environmental Modelling & Software 22: 464-475.

Simeonova, P., Simeonov, V., Andreev, G., 2003. Environmetric analysis of the Struma River water quality. Central European Journal of Chemistry 2: 121-126.

Simeonov, V., Simeonova, P., Tsitouridou, R., 2004. Chemometric quality assessment of surface waters: two case studies. Chemical and Engineering Ecology 11(6): 449-469.

Singh, K.P., Malik, A., Mohan, D., Sinha, S., 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India): a case study. Water Research 38: 3980-3992.

Singh, K.P., Malik, A., Sinha, S., 2005. Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques: a case study. Analytica Chimica Acta 538, 355-374.

Vega, M., Pardo, R., Barrado, E., Deban, L., 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Research 32, 3581-3592.

Wang, M., Webber, M., Finlayson, B., Barnett, J., 2008. Rural industries and water pollution in China. Journal of Environmental Management 86: 648-659.

Wang, X. L., Han, J. Y., Xu, L. G., Zhang, Q., 2010. Spatial and seasonal variations of the contamination within water body of the Grand Canal, China. Environmental Pollution 158: 1513-1520.

Wang, X. P., 2000. Analysis of precipitation characteristics in upstream area of Shaying River. Hydrology 20(1): 35-42.

Wang, J., Da, L., Song, K., Li, B., 2008. Temporal variations of surface water quality in urban, suburban and rural areas during rapid urbanization in Shanghai, China. Environmental Pollution 152, 387–393.

Wunderlin, D.A., Diaz, M.P., Ame, M.V., Pesce, S.F., Hued, A.C., Bistoni, M.A., 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, 2881-2894.

Xia, X. H., Zhou, J. S., Yang, Z. F., 2002. Nitrogen contamination in the Yellow River Basin of China. Journal of Environmental Quality 31(3): 917-925.

Yang, Y. H., Zhou, F., Guo, H. C., Sheng, H., Liu, H., Dao, X., He, C. J., 2010.

Analysis of spatial and temporal water pollution pattern in Lake Dianchi using multivariate statistical methods. Environemtal Monitoring and Assessment 170: 407-416.

Zhao, J., Fu, G., Lei, K., Li, Y. W., Li, X. M., Li, J., 2011. Assessment of spatial-temporal patterns of surface water quality in the Min River (China) and implications for management. 2011 International Conference on Computer Distributed Control and Intelligent Environmental Monitoring 1983-1989.

Zhang, Y. Y., J. Xia, T. Liang & Q. Shao, 2010. Impact of water projects on river flow regimes and water quality in Huai River basin. Water Resource Management 24: 889–908.

Zhang, Y., Guo, F., Meng, W., Wang, X., Q.,2009. Water quality assessment and source identification of Daliao river basin using multivariate statistical methords. Environemtal Monitoring and Assessment 152: 105-121.

Zhang, X., Wang, Q. S., Liu, Y. F., Wu, J., Yu, M., 2011. Application of multivariate statistical techniques in the assessment of water quality in the Southwest New Territories and Kowloon, Hong Kong. Environemtal Monitoring and Assessment 173: 17-27.

Zhang, Y. Y., Xia, J., Wang, G. S., Jiang, Y., Zhao, C. S., 2007. Research on influence of dam's union dispatch on water quality in Huaihe River Basin. Engineering Journal of Wuhan University 40(4): 31-35.