Available online at www.sciencedirect.com
ScienceDirect
Procedia Environmental Sciences 2 (2010) 1220–1234
Procedia Environmental Sciences

International Society for Environmental Information Sciences 2010 Annual Conference (ISEIS)

Assessment of river water quality in Pearl River Delta using multivariate statistical techniques

Xiaoyun Fan, Baoshan Cui*, Hui Zhao, Zhiming Zhang, Honggang Zhang

School of Environment, Beijing Normal University, State Key Joint Laboratory of Environmental Simulation and Pollution Control, No. 19 Xinjiekouwai Street, Beijing 100875, China

Abstract

The Pearl River Delta (PRD) region is one of the most industrialized areas in China, and the river water is increasingly deteriorated due to anthropogenic pollution from the rapid economic development. Principal component analysis (PCA) and cluster analysis (CA) were used to identify characteristics of water quality and to assess water quality spatial pattern in this region. The results of PCA for three regions showed that the first four components of PCA analysis showed 85.52% and 89.25% of the total variance in the data sets of North River region and West River region, respectively, the first three components showed 84.63% of variance for data set of East River region. Results of CA based on the station score of PCA were that stations of North River region, East River region and West River region were grouped into four, three and four clusters, respectively corresponding to severe pollution, moderate pollution, light pollution (except for East River region) and good water quality, which indicated the similarity and dissimilarity of the river water quality. Since, the results suggest that PCA and CA techniques are useful tools for assessment of water quality and management of water resources.

© 2010 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.

Keywords: water quality, principal component analysis, cluster analysis, Pearl River Delta

1. Introduction

Surface water pollution with chemical, physical and biological contaminants by anthropogenic activities is of great environmental attention all over the world [1, 2, 3]. Surface water systems mainly mean the waters naturally open to atmosphere, for example rivers, lakes and reservoirs water [2]. Rivers play an important role in a watershed for carrying off municipal and industrial wastewater and run-off from farm land, and are one of the most susceptible water bodies to pollutants [4, 5, 6]. The constant discharges of domestic and industrial wastewater and seasonal surface run-off due to the climate all have a strong effect on the river discharge and water quality. However, rivers

* Corresponding author. Tel./fax: +86 10 58802079.
E-mail addresses: cuibs67@yahoo.com, cuibs@bnu.edu.cn (B. S. Cui).

1878-0296 © 2010 Published by Elsevier Open access under CC BY-NC-ND license.
doi:10.1016/j.proenv.2010.10.133

are the main water sources for domestic, industrial and agricultural irrigation purposes in a region [7], river water quality is one of important factors directly concerning with health of human and living beings [8]. Therefore, it is imperative and important to have reliable information on characteristics of water quality for effective pollution control and water resource management. There is a great need to evaluate the river water quality.

For the spatial variations in hydrochemistry of rivers, the usual method of water quality evaluation is to measure multiple parameters of pollutants in different monitoring stations at periodic times in a watershed with varying topographical conditions [9]. Since, there is a complex data matrix with a large number of physico-chemical parameters to evaluate the water quality [10]. This is often difficult to providing a representative and reliable estimation conclusion.

In recent years, principal component analysis (PCA) and cluster analysis (CA) have been widely used in the interpretation of complex data sets to better evaluate the water quality and a variety of environmental issues, including inspecting the spatial and temporal patterns of water quality, chemical species associated with hydrological conditions, assessment pollution sources[11, 12, 8, 13, 14, 15, 16]. The application of PCA and CA has provided an effective tool for water resources management and pollution control, and it is useful in identification of possible factors caused by natural and anthropogenic activities that influence water systems [15].

Pearl River delta (PRD), one of the most developed regions in China and Asia [17]，is an intensity region of population and industrial, the mass of sewage discharge increase year by year in this region, especially the discharge of domestic sewage almost increase sharply with population growth. The total discharge of sewage in PRD was 82.5 $\times 10^8$t in 2004, but 51.3$\times 10^8$t in 2002, increased by about 70% in two years. Although regulations of sewage treatment are increased in Guangdong province in recent years, most sewage still directly or indirectly discharge into Pearl River estuary by various pathway due to enlarged economic activities, extensive development pattern of economy and lagging of facilities used for sewage treatment. Water quality of PRD is further deteriorated [18, 19], , and has induced water scarcity in some cities although there were relatively abundant water resources [20, 21]. Due to the value and importance of freshwater resources, surface water pollution in PRD has become a serious problem, and policy-makers and researchers are trying to look for a balance between economic development and environmental protection. Water resource management has targeted water quality improvement resulting from the understanding of deleterious impact of various pollutants on water.

The object of the present study is to analyze physico-chemical parameters of river water quality from Pearl River delta using PCA and CA multivariate technique, and to assess information about the similarity and dissimilarities among the different monitoring stations, to identify water quality variables for spatial differences, and further to make sure the impact of the pollution sources on the water quality parameters.

2. Materials and methods

All mathematical and statistical computations were performed using EXCEL 2003 (Microsoft Office®). Principal component analysis (PCA) and cluster analysis (CA) of water quality data sets were made through Matlab7.0 software.

2.1. Study area and data

The study area (PRD) is part of Guangdong Province and located in the southern of China between 112°00′~115°25′E and 22°30′~23°45′N, includes metropolitan cities such as Hong Kong (6.8million inhabitants), Guangzhou (10million), and Shenzhen (7million). This region belongs to the subtropical zone with a long summer and a short winter, and a mean annual temperature of 21–22 °C. The total rainfall is 1,600-2,000 mm year-1, with 80% of the river′s annual discharge in the wet season (from April to September). For the evaluation of water quality, considering the impacts of natural change and human activities, we investigated some regions in the Pearl River Delta and learned about the situation of water environment. The study area is covered by more than a thousand intersecting large and small channels [22], and the pattern of transport and diffusion for pollutants in the river network is extremely complicated.

In this study, a large number of monitoring stations distribute in the river channels. We mainly studied the water quality of this region in dry season. We collected the data from *Material achievements of synchronous hydrological and water quality monitor of West and East River Delta in dry season during 2005.* Mainly water quality parameters analyzed were dissolved oxygen (DO), chemical oxygen demand (COD_{Mn}), biological oxygen demand (BOD_5), total phosphorus (TP), ammonia nitrogen (NH_3-N), Hg and Oil. The reliability and homogeneity of the water quality monitoring data were strictly checked by the authority before they were released. In order to evaluate the anthropogenic and natural effects in this region, the survey stations were separated into three groups based on the geographical position and river basin as follows:

Stations 2, 3, 10-12, 14, 15, 18, 19, 20, 28, 31-42 are located in North River region; Stations 4, 13, 29, 30, 43-49 are situated in Guangzhou city and East river region (East River region); Stations 1, 5-9, 16, 17, 21-27 are presented in West River region.

Forty-nine monitoring stations altogether are shown in the map below (Figure 1).

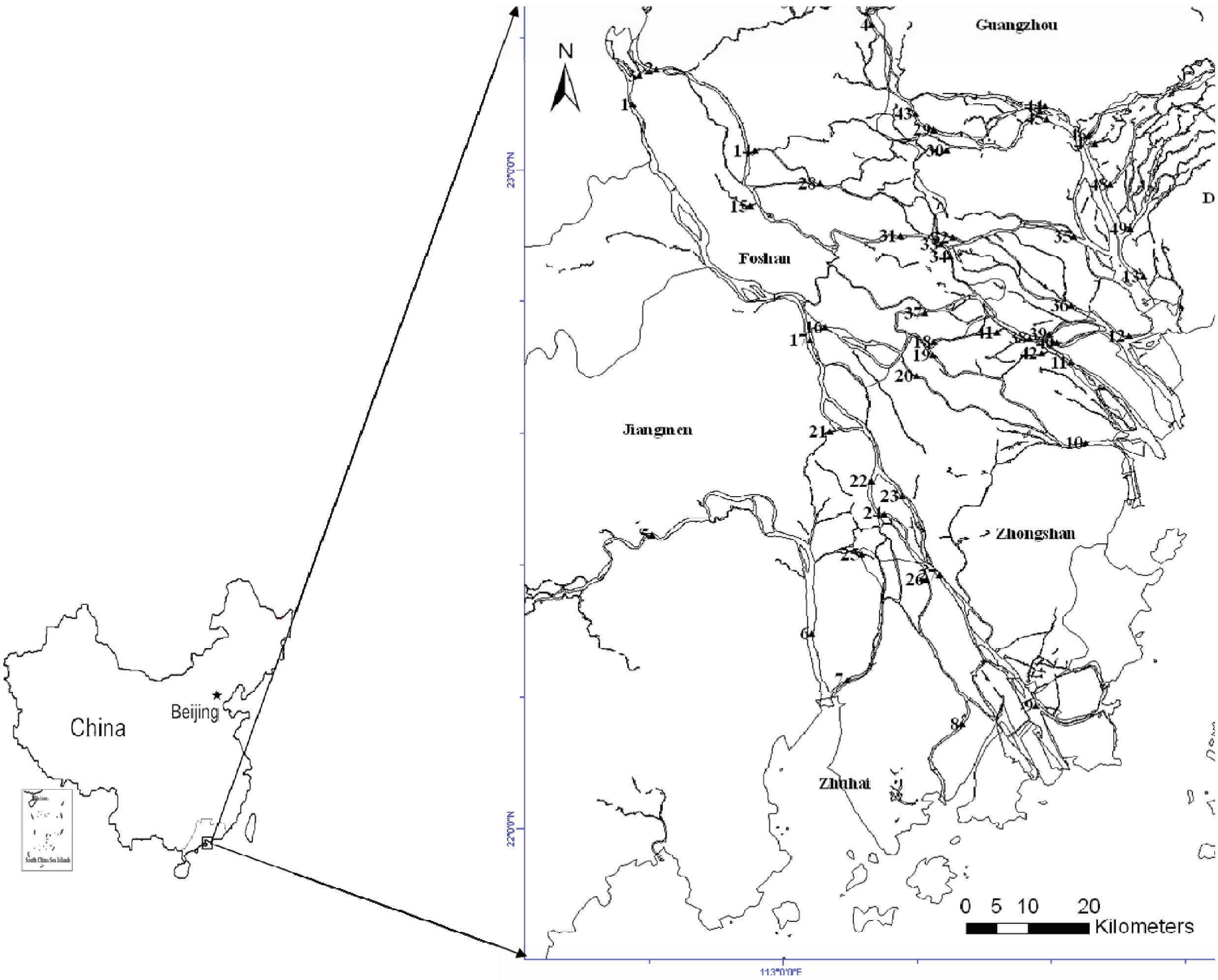

Figure1 Map of study area with the location of water quality monitoring station along with 1-Makou, 2-Sanshui, 3-Ganggen, 4-Laoyagang, 5-Shizui, 6-Guanchong, 7-Xipaotai, 8-Huangjin, 9-Guadingjiao, 10-Hengmen, 11-Fengmamiao, 12-Nansha, 13-Dahu, 14-Zidong,15-Shizaisha,16-Nanhua, 17-Tianhe,18-Haiwei, 19-Nantou, 20-Xiaolan, 21-Beijieshuizha, 22-Baiqing, 23-Daao, 24-Muzhoukou, 25-Hukeng, 26-Zhuzhou, 27-Zhuyin, 28-Lanshi, 29-Shaluowei, 30-Dashi, 31-Xiashi, 32-Sanshazuo, 33-Sanshanjiao, 34-Sanwei, 35-Sanshakou, 36-Tingjiao, 37-Rongqi, 38-Dalongjiao,39-Shangheng, 40-Xiaheng, 41-Wuzhu, 42-Huangshali, 43-Fubiaochang(2), 44-Huangpu(3), 45-Huangpuyou, 46-Dasheng, 47-Mayong, 48-Zhangpeng, 49-Sishengwei.

2.2. Principal component analysis

PCA is a multivariate statistical method which is designed to transform complexity of input variables with a large volume of information into new, uncorrelated variables, called principal components, which are linear combinations of the original variables [15]. And it is intended to have a better interpretation of original variables. In process of statistical analysis, PCA mainly involves the following six major steps: (1) start by coding the variables X_1, X_2, ⋯, X_p to have zero means and unit variance, and standard the variables to make sure they have equal weight in the further analysis; (2) calculate the covariance matrix C; (3) calculate the correlation matrix R; (4) according R to calculate the eigenvalues λ_1, λ_2,⋯, λ_p and the corresponding eigenvectors α_1, α_2, ⋯, α_P; (5) rank eigenvalues and corresponding eigenvectors by the order of numerical value and discard components interpreting a small part of total variance in dataset; and (6) develop the varible loading matrix to infer the principal parameters. The principal components (PC) can be expressed as:

$$Z_{ij}=\alpha_{i1}x_{1j}+\alpha_{i2}x_{2j}+\alpha_{i3}x_{3j}+\cdots+\alpha_{im}x_{mj} \quad (1)$$

Where z is the component score, α 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.

2.3. Cluster analysis

Cluster analysis is an unsupervised pattern recognition technique that uncovers intrinsic structure or underlying behavior of a data set without making a priori assumption about the data, in order to classify the objects of the system into categories or clusters based on their nearness or similarity [23]. Then the results of CA show high homogeneity within cluster and high heterogeneity between clusters [8]. Hierarchical clustering is the most common approach in which clusters are grouped sequentially, by starting with the most similar pair of objects and forming higher clusters step by step and is typically illustrated by a dendrogram. The dendrogram provides a visual summary of the clustering processes, presenting the map of groups with a dramatic reduction in dimensionality of the original data. Hierarchical agglomerative CA was performed on the normalized data set by means of the Ward method, using Euclidean distances as a measure of similarity. The Euclidean distance usually gives the similarity between two samples and a distance can be represented by the difference between analytical values from both the samples. This method uses the analysis of variance approach to evaluate the distances between clusters, attempting to minimize the sum of squares of any two clusters that can be formed at each step [12, 24, 4, 16].

3. Results

3.1. Raw data and water quality trends

Results of physicochemical analysis of the water samples collected from 49 monitoring station are given as follows. Mainly concentrations of chemical pollutants such as DO, COD_{Mn}, BOD_5, TP, NH_3-N, Hg and Oil were statistical analyzed. The results showed that the DO concentration ranged from 0.85-9 mg L^{-1}, the COD_{Mn} concentrations varied from 1.0 to 7.7mg L^{-1}, the BOD_5 values ranged from 0.25 to 22.8 mg L^{-1} , the ammonia nitrogen concentrations varied from 0.25-8.14 mg L^{-1}, the concentrations of total phosphorus ranged from 0.035-0.665 mg L^{-1}, concentration of Oil varied from 0.025-0.225 mg L^{-1}, and the concentration of Hg ranged from 0.00002-0.00009 mg L^{-1}. Analysis of variance for all the parameters except TP, Oil and Hg showed that ρ values were <0.05. It means all the characteristics about pollutants differed significantly at 49 stations except TP, Oil and Hg. To summarize, river water in Pearl River Delta are suffered from different degrees of pollution and showed a significant spatial variation.

3.2. Principal component analysis

The results of the PCA for three regions based on the correlation matrix (Table 1) of chemical parameters are presented in Table 2, and the scree plot and cumulative variance proportion of principal components are shown in Figure 2. The first four principal components of PCA analysis showed 85.52% and 89.25% of the total variance for the data set in North and West River region, respectively; the first three principal components showed 84.21% of the total variance for the data set in East River region (Table 2). Score of each station of PCA in three regions are presented in Table 3, 4 and 5.

3.3. Cluster analysis

Cluster analysis was applied on score of principal components of three regions respectively, to detect spatial similarity and dissimilarity for monitoring stations distributed in the river watershed. The results of CA based on station scores of PCA for the three regions are shown in Figure 3 (a, b, c). The monitoring stations of North River region and West River region were grouped into four statistically significant clusters, respectively; and a total of three clusters of monitoring stations were obtained from East River region (Figure 3 b)

Table 1 Correlation matrices of water quality parameters for different regions.

	DO	COD	BOD	TP	NH3-N	Oil	Hg
North River region							
DO	1						
COD	0.3042	1					
BOD	0.0628	0.4140	1				
TP	-0.4613	-0.3858	-0.0446	1			
NH3-N	-0.7388	-0.1303	-0.0499	0.0587	1		
Oil	0.0362	0.2734	-0.0070	-0.1185	-0.0528	1	
Hg	-0.6318	-0.3038	-0.1203	0.4095	0.4312	0.0205	1
East River region							
DO	1						
COD	-0.4311	1					
BOD	-0.7371	0.5914	1				
TP	-0.2239	0.0914	0.4367	1			
NH3-N	-0.4090	0.0124	0.4940	0.8734	1		
Oil	-0.3410	0.3559	-0.1238	-0.0879	-0.2276	1	
Hg	0.1843	0.4413	0.2208	0.0709	0.0057	-0.2758	1
West River region							
DO	1						
COD	-0.3201	1					
BOD	-0.0458	0.3872	1				
TP	-0.5926	0.4944	0.0838	1			
NH3-N	-0.6082	0.5150	0.3319	0.5652	1		
Oil	-0.0757	0.6737	0.1617	0.3402	0.2251	1	
Hg	-0.0920	-0.2517	-0.3871	-0.1150	0.2033	-0.0228	1

4. Discussion

4.1. Regional correlation of water quality parameters

Before applying PCA, correlation analysis was carried out. And the analytical results are listed in Table 1. According to the results of correlation matrices, for North River region, significant negative correlations existed between DO and NH_3-N, DO and Hg; there was significant positive correlation between COD_{Mn} and BOD_5; and TP and Oil variables were independent. As to East river region, significant negative correlation existed between COD_{Mn} and DO; significant positive correlations appeared between COD_{Mn} and BOD_5, TP and NH_3-N. Similar to the above two regions, there were significant negative correlations between DO and TP, DO and NH_3-N; and positive correlations existed between COD_{Mn} and NH_3-N, COD_{Mn} and Oil, TP and NH_3-N. The analysis results indicated that significant positive and negative correlations existed between parameters of water quality. Based on these relationships, it can be concluded that a few factors can explain the main variance of all the parameters, and be used to find an internal structure and assist in determination of pollutant sources not directly accessible to [14, 5].

4.2. Spatial variability and source identification of water quality using principal component analysis

Due to the complexity of relationships between parameters of water quality, it is difficult to draw a clear conclusion directly. However, principal component analysis can explain the structure of the data in detail by extracting the latent information. The scree plot with eigenvalues was used to determine the number of principal

Table 2 Loadings and eigenvalues of water quality parameters on significant principal components for river water samples from North River region, East River region and West river region

	PC1	PC2	PC3	PC4
North River region				
DO	-0.883	-0.324	-0.027	0.054
COD_{Mn}	-0.591	0.643	0.047	0.021
BOD_5	-0.271	0.576	0.669	0.208
TP	0.614	-0.212	0.281	0.618
NH_3-N	0.673	0.470	-0.060	-0.491
Oil	-0.170	0.457	-0.703	0.464
Hg	0.782	0.182	-0.084	0.188
Eigenvalue	2.673	1.359	1.035	0.920
Variability (%)	38.185	19.409	14.784	13.142
Cumulative (%)	38.185	57.594	72.378	85.520
East River region				
DO	-0.747	-0.398	0.353	
COD_{Mn}	0.567	0.636	0.414	
BOD_5	0.894	0.113	0.154	
TP	0.712	-0.505	-0.153	
NH_3-N	0.760	-0.552	-0.222	
Oil	0.031	0.750	-0.430	
Hg	0.198	-0.019	0.919	
Eigenvalue	2.804	1.698	1.422	
Variability (%)	40.054	24.258	20.319	
Cumulative (%)	40.054	64.312	84.631	
West River region				
DO	-0.658	0.501	0.333	0.173
COD_{Mn}	0.835	0.303	0.262	0.005
BOD_5	0.432	0.592	-0.381	0.508
TP	0.782	-0.215	-0.080	-0.410
NH_3-N	0.792	-0.335	-0.185	0.348
Oil	0.588	0.254	0.720	-0.003
Hg	-0.135	-0.775	0.364	0.464
Eigenvalue	2.920	1.517	1.017	0.793
Variability (%)	41.718	21.674	14.53	11.331
Cumulative (%)	41.718	63.392	77.922	89.253

components (PCs) that could be kept for further research and comprehending the underlying data set [3, 23]. In this study the scree plot (Figure 2) showed that the first four PCs were be used to further analysis for North and West River regions, and the first three PCs for East River region, respectively explained 85.52%, 84.63% and 89.25% of the original dataset.

Projection of the original variables on the subspace of the PCs are called loading and coincides with the correlation coefficients between PCs and variables [9, 23]. And the component loadings can be used to determine the relative importance of a variable as compared to other variables in a PC, and do not reflect the importance of the component itself [3]. Therefore, the positive values on each component are related to important inputs, and the negative values correspond to low inputs [14].

As to water quality of North River region, loading of four retained PCs are presented in Table 2. PC1 explained 38.19% of the total variance and was highly contributed to by variables organic related parameters (DO and COD_{Mn}), inorganic nutrients (NH_3-N and TP) and metal Hg. Since the component seems to mainly measure the synthesized information of all parameters of water quality, and may be related to anthropogenic pollution of industrial, domestic and agricultural source. This component also reveals that the parameter Oil with low loadings were less important in accounting for river water quality in the North River region. PC2 explained 19.41% of the total variance and was positively contributed to by BOD_5 and COD_{Mn}. It indicated that the component was related to organic pollutants from domestic and industrial wastewater. PC3 explained 14.78% of the total variance in the dataset and consist of a strong loading of BOD_5, and the negative variable loading of this component was Oil. This component was mainly

Table 3 Score of PCA for each station in North River region

	PC1	PC2	PC3	PC4	Total score
2	-1.1857	-0.5909	0.1817	-0.4504	-0.5997
3	-1.3450	-1.4452	0.0438	-0.0022	-0.7880
10	0.5197	-1.6193	-0.3509	-0.5673	-0.2419
11	0.7569	-0.7489	-0.3814	-1.0112	-0.0451
12	0.9260	0.4792	0.1534	-0.7319	0.3735
14	-0.9164	-0.2226	-0.1608	-0.3090	-0.4575
15	-0.9475	-0.6766	0.6108	-0.2426	-0.4346
18	-0.7217	-1.0442	0.0869	-0.0580	-0.4730
19	-0.5245	-0.7033	0.4701	0.4419	-0.2094
20	-0.1937	-0.1433	0.4964	0.6232	0.0533
28	-0.5783	0.1908	-0.6510	-0.6215	-0.3616
31	-0.5096	0.3252	0.9461	-0.1628	-0.0129
32	-0.5634	2.4744	2.0324	0.6806	0.6548
33	-0.4641	0.7820	0.8931	0.8493	0.2179
34	-0.2449	0.3035	0.6199	-0.4774	-0.0055
35	1.8276	1.0956	0.0079	-1.6405	0.6969
36	1.0625	-0.2577	-0.2693	-0.5698	0.2414
37	-0.8524	-0.1345	-1.1487	0.4373	-0.4644
38	0.1164	1.2392	0.3171	-0.4428	0.2738
39	0.0110	1.1274	-2.1270	1.2615	0.0734
40	-0.0634	0.9990	-2.5790	1.4060	-0.0279
41	1.4225	-0.0232	-0.3615	-1.3416	0.3096
42	2.4678	-1.4065	1.1698	2.9291	1.2267

Table 4 Score of PCA for each station in East River region.

	PC1	PC2	PC3	Total score
4	0.5964	-1.3522	-0.6512	-0.2215
13	-1.2260	-1.0903	-0.1335	-0.7826
29	1.5670	-0.0389	0.1820	0.6551
30	1.5433	0.5150	0.7953	0.9046
43	0.8259	-1.2015	-0.7330	-0.1096
44	-0.9498	-0.5151	0.1910	-0.4666
45	-0.8898	-0.0245	-0.0264	-0.3677
46	-0.4696	1.1692	-1.9968	-0.3100
47	0.1665	1.6085	-0.0285	0.4511
48	-0.5465	1.0216	0.3499	0.1000
49	-0.6174	-0.0919	2.0512	0.1471

Table 5 Score of PCA for each station in West River region

	PC1	PC2	PC3	PC4	Total score
1	-1.1896	2.4871	0.0159	0.4572	0.0968
5	0.3710	0.2201	-0.6890	-1.4394	-0.0608
6	-0.0313	0.1415	-0.6306	0.2216	-0.0489
7	0.2500	0.6957	-1.3514	-0.0599	0.0519
8	1.8733	-1.0621	-1.0676	1.5972	0.5773
9	-0.3449	-1.2717	-0.1743	-0.6458	-0.5180
16	-1.3118	-0.6503	0.9386	0.2809	-0.5200
17	-1.2807	-0.7134	0.9065	1.7991	-0.3534
21	2.1261	0.9181	2.6041	-0.0107	1.4632
22	-0.3147	0.5181	-0.2056	0.7402	0.0350
23	-0.1158	-0.8625	0.4723	-1.7547	-0.3654
24	-0.1170	-0.7126	0.5959	-0.2763	-0.1479
25	0.4570	-0.2038	-0.8223	0.6640	0.1022
26	-0.6705	-0.4793	-0.0005	-0.7678	-0.4707
27	0.2988	0.9750	-0.5920	-0.8056	0.1587

associated with organic related parameters and Oil parameters. PC4 accounted for 13.14% of the total variance and was highly contributed to by variable TP, and was bound up with inorganic nutrients parameters.

For East River region, analysis results of loading and cumulative (%) on three retained PCs are also given in Table2. The total explained variance of PC1 was 40.05%, and had a strong loading of organic related parameters (BOD_5), inorganic nutrients (NH_3-N and TP), and was negatively influenced by DO. This component seemed to be composite influence of organic pollutants and inorganic nutrients. The pollution in this region was ascribed to concentrated urban and industrial activities which induced a large number of wastewater discharge into river nearby. Other researchers also provided the similar issue [25, 26, 20]. Among these literatures, Ho and Hui [25] utilized photogrammetic to survey the sources of contamination in East river and found that the pollution were related to human activities, such as domestic and industrial discharges, agricultural chemical applications and soil erosion due to deforestation. PC2 represented 24.26% of the total variance and consists of a strong loading of parameter Oil.

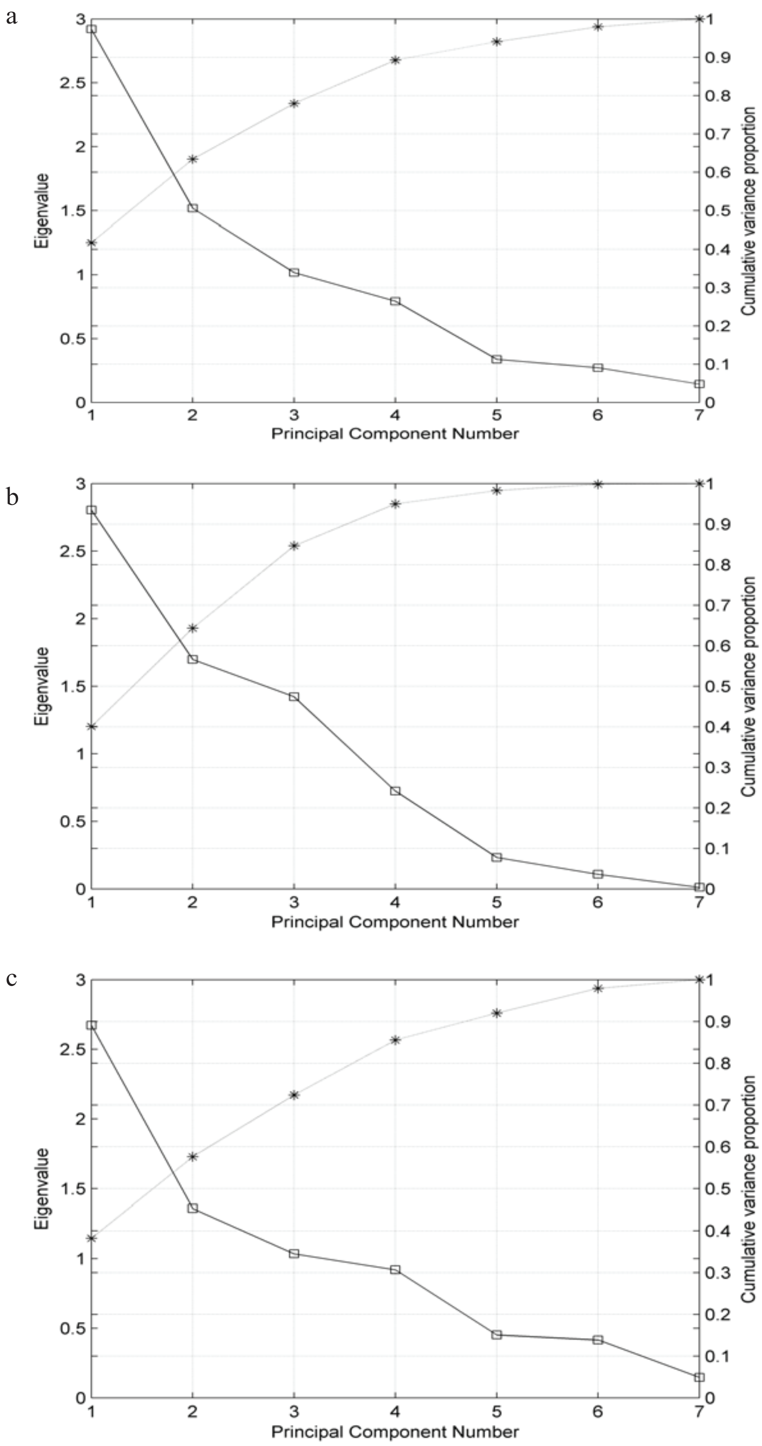

Figure2 Scree plot and cu n, b: East River region, c: West River region).

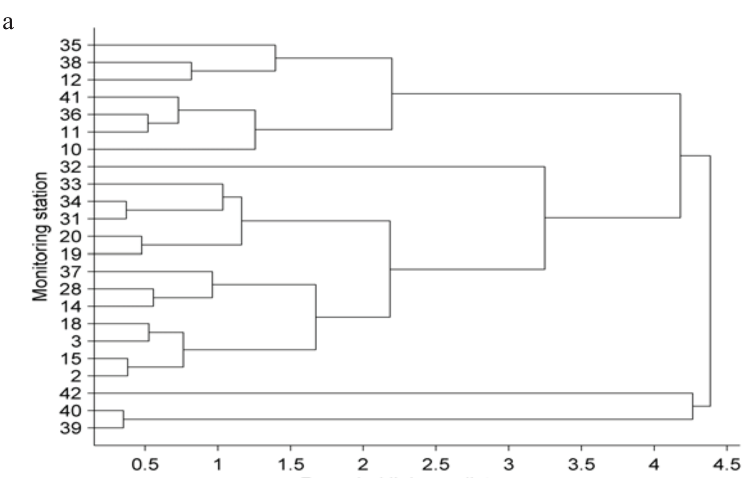

Figure3. Result of cluster analysis based on the score of PCA (a: North River region, b: East River region, c: West River region). The dissimilarity defined by Euclidean distance and the combination of cluster is based on Ward method.

This component was associated with influences from Oil pollutants. PC3 explained 20.32% of the total variance and contributed to by variable Hg, so this component linked with impacts from metal.

As shown in Table2, the result of PCA indicated four components underlying the characteristics of river water quality in West River region. Similar to the results of North and East River region, PC1 (41.72% of total variance in the dataset) in this region was also related to common influences of almost all the parameters of water quality. And the non-point source pollution for the use of chemical fertilizers and pesticides is becoming increasingly serious associated with the development of agriculture. PC2 accounted for 21.67% of the total variance and consisted of a negative loading of Hg. It indicated that this component was related to factor of metal. PC3 explained 14.53% of the total variance was an Oil-related component with a high loading of Oil. PC4 represented 11.33% of the total variance in the dataset and had a strong contribution from BOD_5, so this component was associated with organic pollutants parameters.

For the three regions, PC1 was related to synthesis influence of almost all the parameters of river water quality. But the information of PC2, PC3 and PC4 (except East River region) related to was different. Similar researches applied PCA for evaluation of spatial variations in river water quality has earlier been used [8, 12, 24, 23, 27].

4.3. Analysis of stations with parameters variance in the three regions

Table 3, table 4 and table 5 show the PC score of every station in North River region, East River region and North River region, respectively.

For North River region, basing on PCs score station 42 had got the highest scores among all the stations, and station 35 and 32 came second; score of them were much higher than others, it indicated that the rivers which the three stations located at were suffered from serious pollution. Among these stations, station42 is located in Zhongshan city and both station 35 and 32 are located in Panyu district of Guangzhou city, where many manufacturing factories were built and a large number of wastewater from domestic and industrial plants were discharged. The station 2 and 3 had the lower scores than others, which suggested both of them had better river water quality; these stations are located in the upstream of North River and far from cities, so they had received little pollution.

As to East River region, scores of station 30 and 29 were higher than other stations according to the result of PCA, and score of station 13 was the lowest which significantly lower than others. It can be concluded that in this region rivers with station 30 and 29 were seriously polluted by all kinds of pollutants, because there were a mass of factories nearby the two stations in Guangzhou city. But water quality was a little better in river flowing through station 13, which was because the station was located at the Humen where water quality was influence by tide and had better purification capacity.

Among of all the stations in West River region, station 21 had got the highest score based on the result of PCA, and the second was station 8; the lower scores appeared in station 9 and 16. It seemed to that in this region water quality of station21 were deteriorated seriously and the next was station 8 also with bad water quality compared with other stations. However, river where station 8 located at had better water quality in this region.

4.4. Spatial similarity and site grouping in different regions

CA has been proved useful in solving classification problems where the object is to sort cases or variables into groups, such that the degree of association is strong between members of the same cluster and weak between members of different clusters [28]. Since, in a watershed CA technique is a useful tool that offers reliable classification of surface water and makes possible to adequately serve for spatial assessment in an optimal manner [8]. Sites in the same cluster often possess similar characteristics and natural background source types [4]. To survey anthropogenic and natural influences on characteristics of water quality in monitoring stations of three regions, CA was used to classify these stations based on results of PCA.

4.4.1. North River region

CA generated a dendrogram grouping the monitoring stations into four groups at $(Dlink/Dmax)\times 100<15$, and there were significant differences between clusters (Figure 3A). Among all the stations, station 42 was clarified into cluster 1, where river water quality was polluted seriously, and mainly influenced by synthesis pollutants from domestic and industrial waste water and agriculture runoff. Cluster 2 consisted of seven stations (10, 11, 12, 35, 36, 38 and 41) with moderate pollution of river water. And the dendrogram clarify the abnormality of station 39 and station 40 which make up one group as cluster3, water quality of which was light pollution and influenced by organic and inorganic nutrients pollutants. The remaining stations (2, 3, 14, 15, 18-20, 28, 31, 32, 33, 34, 37) were grouped into cluster4. Water quality of this group was better compared with the above three clusters.

4.4.2. East River region

The dendrogram of Figure 3B showed that a total three clusters of monitoring stations were obtained at (Dlink/Dmax)×100<15 by CA based on the score of PCA in East River region. The monitoring stations 4, 29, 30 and 43 were grouped into cluster1 with water quality of severe pollution. Water quality in these stations was comprehensively affected by all the parameters in this study. Cluster 2 consisted of station 46, 47 and 48, where water quality was received moderate pollution, and was mainly caused by manufacturing industry wastewater and shipping. Cluster 3 was grouped by stations 13, 44, 45 and 49 which received light pollution, and have better water quality.

4.4.3. West River region

According to result of CA, the monitoring stations in this region were grouped into four clusters concluded from the dendrogram (Figure 3C) at (Dlink/Dmax)×100<15. Cluster 1 only consists of one station (21), had received severe pollution from domestic and industrial wastewater and agricultural runoff. Similar to cluster 1, cluster 2 was grouped by station 8, water quality of which was better than cluster 1, but were moderate polluted. Cluster 3 consisted of six stations (9, 16, 17, 23, 24 and 26), was low polluted region. Seven stations (1, 5, 6, 7, 22, 25 and 27) were classified into cluster 4. Water quality in these stations was better than others, and was suffered from less pollution compared with other stations.

This suggests that the CA technique was reasonably able to distinguish the similarity and dissimilarity of spatial patterns for different types of water quality in the study area, and will be possible for rapid management of water resources in an optimal manner. The same characteristics are also reported by other researchers [8, 29, 30, 4].

5. Conclusion

In this study, PCA and CA multivariate techniques have been used for water quality spatial assessment and pollution sources identification in PRD. The results of the PCA suggested parameters responsible for water quality variations in North River region was mainly related to organic related parameters (DO and COD_{Mn}), inorganic nutrients (NH_3-N and TP) and metal Hg; but in East River region, mainly related to organic related parameters (BOD_5) and inorganic nutrients (NH_3-N and TP), and in West River Region, mainly related to organic related parameters (COD_{Mn}) and inorganic nutrients (NH_3-N and TP). Cluster analysis was applied on score of principal components of three regions respectively. The monitoring stations of North River region, East River region and West River region were grouped into four, three and four statistically significant clusters, respectively. Stations within a same cluster had similar water quality characteristics. Therefore, the method used here can offer a useful tool for assessment of water quality and management of water resources in some regions with a large number complex water quality datasets involved.

Acknowledgments

This study was funded by the National Natural Science Foundation (U0833002; 50939001) and National Key Basic Research Project of People's Republic of China (2003CB415104).

References

[1] May AM, Mutasem E, Mark DS, John, NL. Factors influencing development of management strategies for the Abou Ali River in Lebanon. *Sci Total Environ* 2006; 362: 15-30.

[2] Noori R, Sabahi MS, Karbassi AR, Baghvand, A, Taati Zadeh, H. Multivariate statistical analysis of surface water quality based on correlations and variations in the data set. *Desalination*, doi:10.1016/j.desal.2010.04.053.

[3] Ouyang Y, Nkedi-Kizza P, Wu QT, Shinde D, Huang, CH. Assessment of seasonal variations in surface water quality. *Water Res* 2006; 40: 3800-3810.

[4] Singh KP, Malik A, Mohan D, Sinba, S. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomi River (India)- a case study. *Water Res* 2004; 38: 3980-3992.

[5] Singh KP, Malik A, Sinha S. Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques- a case study. *Anal Chim Acta* 2005; 538: 355-374.

[6] Wang XL, Lu YL, Han JY, He, GZ, Wang, TY. Identification of anthropogenic influences on water quality of rivers in Taihu watershed. *J Environ Sci-China* 2007; 19: 475-481.

[7] Yu S, Shang J. Factor analysis and dynamics of water quality of the Songhua River, Northeast China. *Water Air Soil Poll* 2003; 144: 159-169.

[8] Kazi TG, Arain MB, Jamali MK, Jalbani, N, Afridi, HI, Sarfraz, RA., et al. Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. *Ecotox Environ Safe* 2009; 72: 301-309.

[9] Razmkhah H, Abrishamchi A, Torkian A. Evaluation of spatial and temporal variation in water quality by pattern recognition techniques: A case study on Jajrood River (Tehran, Iran). *J Environ Manage* 2010; 91:852-860.

[10] Chapman D. Water quality assessment, ed. D. Chapman on behalf of UNESCO, WHO and UNEP. Chapman and Hall, London; 1992, 585.

[11] Bouza-Deaño R, Ternero-Rodríguez M, Fernández-Espinosa AJ. Trend study and assessment of surface water quality in the Ebro River (Spain). *J Hydrol* 2008; 361: 227-239.

[12] Omo-Irabor OO, Olobaniyi SB, Oduyemi K, Akunna, J. Surface and ground water quality assessment using mutivariate analytical methods: a case study of the Western Niger Delta, Nigeria. *Phys Chem Earth* 2008; 33: 663-673.

[13] Ouyang Y. Evaluation of river water quality monitoring stations by principal component analysis. *Water Res* 2005; 39: 2621-2635.

[14] Parinet B, Lhote A, Legube B. Principal component analysis: an appropriate tool for water quality evaluation and management—application to a tropical lake system. *Ecol Model* 2004; 178: 295-311.

[15] Shrestha S, Kazama F. Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. *Environ Modell Softw* 2007; 22: 464-475.

[16] Wu ML, Wang YS, Sun CC, Wang, HL, Dong, JD, Yin, JP, et al. Identification of coastal water quality by statistical analysis mthods in Daya Bay, South China Sea. *Mar Pollut Bull* 2010; 60: 852-860.

[17] Chau KW, Jiang YW. Three-dimensional pollutant transport model for the Pearl River Estuary. *Water Res* 2002; 36: 2029–2039.

[18] He GF, Yuan GM, Li FQ. Effect of economic development on water quality in Zhujiang Estuary. Marine environmental science 2004; 23(4): 50-52, 70. (in Chinese)

[19] Li B, Huang GQ. Ecology effect and counter measure of urbanization in Pearl River Estuary. Marine environmental science 2008; 27(5): 543-546. (in Chinese)

[20] Zhu ZY, Deng QL, Zhou HY, Ouyang, TP, Kuang, YQ, Huang, NS, et al. Water Pollution and Degradation in Pearl River Delta, South China. *Ambio* 2002; 31(3): 226-230.

[21] Ouyang TP, Zhu ZY, Kuang YQ. Assessing impact of urbanization on river water quality in the Pearl River Delta Economic Zone, China. *Environ Monit Assess* 2006; 120: 313-325.

[22] Leung Y, Lee Y, Lam KC, Lin, K., Zeng, FT. An environmental decision-support system for the management of water pollution in a tidal river network. *Int J Geogr Inf Sci* 2005; 19(4): 483–500.

[23] Vega M, Parda R, Barrada E, Deban, L. Assessment of seasona and polluting effects on the quality of river water by exploratory data analysis. *Water Res* 1998; 32: 3581-3592.

[24] Simeonov V, Statis J A, Samara C, Zachariadis, G., Voutsa, D., Anthemidis, A., et al. Assessment of the surface water quality in Northern Greece. *Water Res* 2003; 37: 4119-4124.

[25] Ho KC, Hui KCC. Chemical contamination of the East River (Dongjiang) and its implication on sustainable development in the Pearl River Delta. *Environ Int* 2001; 26: 303-308.

[26] Ouyang TP, Zhu ZY, Kuang YQ. River water quality and pollution sources in the Pearl River Delta, China. *J Environ Monitor* 2005; 7: 664-669.

[27] Wunderlin DA, Diaz MP, Ame MV, Pesce, SF, Hued, AC, Bistoni, MA. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality, a case study: Suquia river basin (CordobaArgentina). *Water res* 2001; 35: 2881-2894.

[28] Maria JB, Graca C. Identification of similar environmental areas in Tagus Estuary by using multivariate analysis. *Ecol Indic* 2006; 6: 508-515.

[29] Kim JH, Kim RH, Lee L, Cheong, TJ, Yum, BW, Chang, HW. Multivariate statistical analysis to identify the major factors governing groundwater quality in the coastal area of Kimje, South Korea. *Hydrol Process* 2005; 19: 1261-1276.

[30] Panda UC, Sundaray SK, Rath P, Nayak, BB, Bhatta, D. Application of factor and cluster analysis for characterization of river and estuarine water systems-a case study: Mahanadi River (India). *J Hydrol* 2006; 331(3-4): 434-445.