KSCE Journal of Civil Engineering (2011) 15(6):995-1003
DOI 10.1007/s12205-011-1079-y

Environmental Engineering
www.springer.com/12205

Surface Water Quality Assessment in the Central Part of Bangladesh Using Multivariate Analysis

Mohammad A. H. Bhuiyan*, M. A. Rakib, S. B. Dampare***, S. Ganyaglo****, and Shigeyuki Suzuki*****

Received December 15, 2009/Accepted November 4, 2010

Abstract

This study deals with the natural and anthropogenic processes that influence the surface water quality in the central Bangladesh using multivariate statistical techniques. The investigation shows that the Total Suspended Solids (TSS), Total Dissolved Solids (TDS), turbidity, Electrical Conductivity (EC), Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), NO_3^-, SO_4^{2-}, Cl^-, PO_4^{3-} and microbial loads are higher than the Bangladesh standards. R-mode CA groups all 10 sampling sites into 3 statistically significant clusters, reflecting the different physicochemical characteristics and pollution levels of the sites. R-mode CA suggests common sources (industrial, agriculture and urban sewage) for TSS, EC, turbidity, temperature, COD, PO_4^{3-}, SO_4^{2-}, and Fecal Coliform (FC). The PCA/FA identifies 5 dominant factors as responsible for the data structure, explaining 88.3% of the total variance in the dataset. The multiple anthropogenic (i.e., industrial, agricultural, urban sewage) and natural sources (soil erosion, aquatic hyacinths and weeds) of water quality parameters have been identified by PCA. This work is believed to serve as a baseline data for further studies in the Turag River system as well as inform decision-makers on the proper design of sampling and analytical protocols for effective pollution management of the surface water quality in the basin.

Keywords: *water quality, cluster analysis, principal component analysis, Dhaka City*

1. Introduction

Rivers and streams are heterogeneous at different spatial scales. This heterogeneity may be attributed to a number of factors including anthropogenic input, local environmental conditions, water discharge, water velocity and degree of temporal variability of surface water chemistry (Qadir *et al.,* 2007). Considering the important contribution of rivers to the main water resources in inland areas for drinking, irrigation and industrial purposes, it is imperative to have reliable information on water quality for effective and efficient water management.

The environment, economic growth and developments of Bangladesh are all highly influenced by surface water - its regional and seasonal availability. Spatial and seasonal availability of surface water is highly responsive to the monsoon climate and physiography of the country. In terms of quality, the surface water of the country is vulnerable to pollution from untreated industrial effluents and municipal wastewater, runoff from chemical industry and agricultural fields, and oil and lube spillage from the operation of sea and river ports. The major industrial area of Bangladesh is situated on a land bordered by Turag River, Bangshi Channel (BC), Sitalakhaya River and Dhaka-Aricha highway. Dhaka Export Processing Zone (DEPZ) is located very close to the upstream point of the study area (Savar). Although there are several stringent rules for treatment of industrial wastes, disposal of untreated wastewater into drains and subsequently into the city's major streams is very common. This poses a potential health and environmental risk to the people living in central Bangladesh. As the fluvial environment of Bangladesh is mainly controlled by seasonal fluctuations, it is important to characterize the seasonal change for evaluating the temporal variations of water pollution.

The usual practice of water quality assessment is the comparison of measured physicochemical parameters with threshold values recommended by national or international bodies. The data mining of surface water monitoring results involves several approaches of which chemometric methods have been considered among the most reliable techniques, as the environmental

*Graduate Student, Graduate School of Natural Science and Technology, Okayama University, Okayama 700-8530, Japan and Assistant Professor, Dept. of Environmental Sciences, Jahangirnagar University, Dhaka 1342, Bangladesh (Corresponding Author, E-mail: amirhb75@yahoo.com)
**Graduate Student, Dept. of Environmental Sciences, Jahangirnagar University, Dhaka 1342, Bangladesh (E-mail: rakibmamun_ju@yahoo.com)
***Post Doctoral Research Fellow, Dept. of Earth Sciences, Okayama University, Okayama 700-8530, Japan, and Scientific Officer, National Nuclear Research Institute, Ghana Atomic Energy Commission, Legon-Accra (E-mail: dampare@cc.okayama-u.ac.jp)
****Scientific Officer, National Nuclear Research Institute, Ghana Atomic Energy Commission, Legon-Accra (E-mail: sganyaglo@yahoo.co.uk)
*****Associate Professor, Dept. of Earth Sciences, Okayama University, Okayama 700-8530, Japan (E-mail: zysuzuk@cc.okayama-u.ac.jp)

system is regarded as a multivariate one (e.g., Marengo *et al.,* 1995; Stefanov *et al.,* 1999; Wunderlin *et al.,* 2001; Lu and Lo, 2002; Simeonov *et al.,* 2002; Mendiguchýa *et al.,* 2004; Astel *et al.,* 2006; Kowalkowski *et al.,* 2006; Simeonova and Simeonov, 2007; Astel *et al.,* 2008). The application of the traditional chemometric techniques such as Cluster Analysis (CA) and Principal Component Analysis (PCA) in the interpretation of complex hydrochemical data matrices improves the understanding of the water quality and ecological status of the studied systems. It allows the identification of possible factors/sources that influence water system, and offers a valuable tool for reliable management of water resources as well as rapid solution to pollution problems (Vega *et al.,* 1998; Wunderlin *et al.,* 2001; Reghunath *et al.,* 2002; Simeonova *et al.,* 2003; Shrestha and Kazama, 2007). Several efforts have been made to compare traditional chemometric techniques like Cluster Analysis (CA) or Principal Components Analysis (PCA) with some alternative approaches like Self-Organizing Maps (SOM) for small data sets (Astel *et al.,* 2007).

In this contribution, the PCA and CA techniques have been applied to evaluate the variation of water quality parameters, spatial similarities and extract the parameters that are most important in assessing and monitoring water quality in the northwestern fluvial system of Dhaka, Bangladesh. In addition, Pearson's correlation matrix was applied to the water quality parameters for a better understanding the PCA results.

2. Study Area

The locations of the 10 selected sampling sites in the study area are shown in Fig. 1. The sites were chosen mainly due to their proximity to suspected pollution sources and their ecological, environmental and aesthetic importance. Site 1 (Tongi bridge) is the closest site to the industrial pollution sources, with interference of human activities. All communities perched along Turag River floodplain use the surface water for domestic purposes. Site 2 (Ashulia) is located downstream of the small and medium industrial area of Tongi city. Site 3 (Ashulia Model Town) is located on Jirab channel, a tributary of TR. Site 4 (New Uttara Model Town), a point of AWL, is also famous as water recreational spot during monsoon season. Site 5 (Ashulia channel) is the runoff of DEPZ. Site 6 (Ashulia wetland) is the surface water reservoir of the western margin of Dhaka City (DC). It is considered as the aesthetically and ecologically significant spot of DC, and famous for rice production. Site 7 (Berulia) is situated after the confluence of Ashulia Channel (AC) with TR and downstream of the point where the untreated sewage is discharged from Uttara Model Town (UMT) of DC. Site 8 is located beside the National Botanical and Zoological garden. Site 9 is the downstream of Bangshi River (BR). Site 10 (Amin Bazar) is located on the TR, near the urban settlement of Mirpur-Amin Bazar and downstream of the confluence of the BR. Amin Bazar and Gabtoli are the biggest commercial place and river port in Bangladesh, respectively.

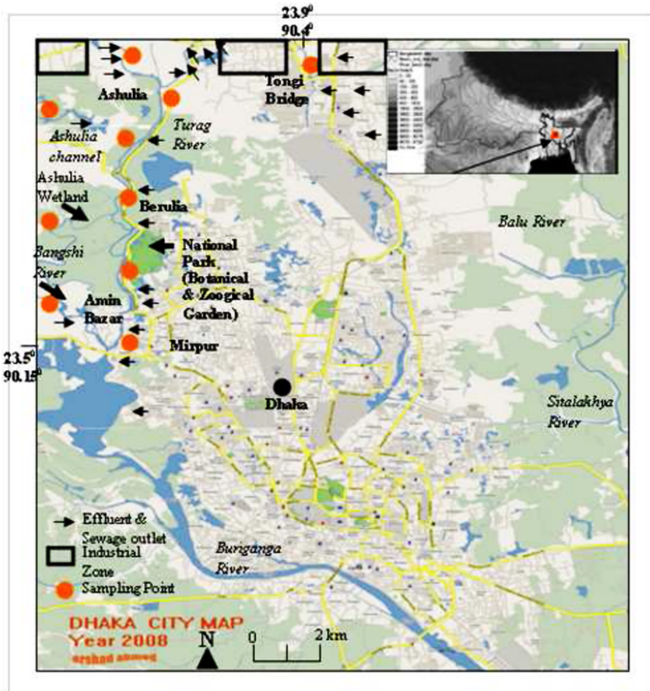

Fig. 1. Location Map of the Study Area

3. Materials and Methods

3.1 Methods for Chemical and Physical Analyses

Fifty samples from the first 25 cm of the water column were collected at each site using a pre-sterilized bottom-weighted polyethylene flask, repeatedly washed with water from those sites. Sampling was done at each site on monthly basis from May, 2007 to May, 2008. The geographic positions of the sample sites were recorded by GPS (Explorist, model: 200). Temperature, pH and conductivity measurements were performed *in situ* with a portable meter equipped with a temperature sensor and a membrane electrode (HANNA Instruments, model: HI 9143). Clark cell probes method was used for DO and BOD measurements (Johnston and Williams, 2006). TSS measurement was conducted by a simple gravimetric method. Alkalinity was measured by titrimetric method. The membrane filtration method was used for the determination of *F. coliforms* (FC) and *F. streptococci* (FS) (USEPA, 1985; Clesceri *et al.,* 1998). Chemical oxygen demand was measured by the dichromate reflux method (Michael 1975; Clair *et al.,* 2003). Anions (SO_4^{2-}, PO_4^{3-}, Cl^-, NO_3^-) were analyzed by colorimetric method (APHA, 1995).

Rainfall data were collected from the Weather Forecasting Center of Bangladesh Meteorological Department (BMD) and discharged water data obtained from Bangladesh Water Development Board (BWDB). All units of measurements of the basic water quality parameters are shown in Table 1.

3.2 Statistical Analysis

Water quality data were subjected to univariate analysis: range, mean and standard deviation and multivariate analysis: Cluster

Table 1. The Water Quality Parameters with Their Abbreviations and Units used in the Study

Parameter	Abbreviation	Unit
Water temperature	Temp	°C
Total suspended solids	TSS	mg/l
Total dissolved solids	TDS	mg/l
Turbidity	Turb	NTU
Total alkalinity ($CaCO_3$)	Alkal	mg/l
Acidity/alkalinity	pH	pH unit
Electrical conductivity	Mmho cm^{-1}	µS/cm
Dissolved oxygen	DO	mg/l
Biochemical oxygen demand	BOD_5	mg/l
Chemical oxygen demand	COD	mg/l
NO_3^{-1}		mg/l
SO_4^{-2}		mg/l
PO_4^{-3}		mg/l
Cl^{-1}		mg/l
Fecal coliform	FC	cfu/100ml
Total coliform	TC	cfu/100ml

Analysis (CA) and Principal Component Analysis (PCA)/Factor Analysis (FA) using SPSS 16.0 for windows. Prior to such analyses, the raw data were commonly normalized to avoid misclassifications due to the different order of magnitude and range of variation of the analytical parameters (Aruga *et al.,* 1995).

3.2.1 Principal Component Analysis/Factor Analysis

PCA is designed to transform the original variables into new, uncorrelated variables (axes), called the principal components. This process reduces the dimensionality of data by a linear combination of original data to generate new latent variables which are orthogonal and uncorrelated to each other (Nkansah *et al.*, 2010). 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 (Sarbu and Pop, 2005). The Principal Component (PC) provides information on the most meaningful parameters, which describe a whole data set affording data reduction with minimum loss of original information (Helena *et al.*, 2000). The principal component can be expressed as:

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

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

Factor Analysis (FA) is similar to PCA except for the preparation of the observed correlation matrix for extraction and the underlying theory (Tabachnick and Fidell, 2007). The major objective of FA is to reduce the contribution of less significant variables to simplify even more of the data structure coming from PCA. This goal can be achieved by rotating the axis defined by PCA, according to well established rules, and constructing new variables, also Called Varifactors (VF) (Shrestha and Kazama, 2007). In fact, PC is a linear combination of observable variables, whereas VF can include unobservable, hypothetical, latent variables (Vega *et al*., 1998; Helena *et al.,* 2000). PCA of the normalized variables was presented to extract significant PCs and to further decrease the contribution of variables with less significance; these PCs were subjected to varimax rotation (raw) generating VFs (Howitt and Cramer, 2005; Shrestha and Kazama, 2007 and references therein). Therefore, a small number of factors will generally explain about the same amount of information as do the much larger set of original observations (Shrestha and Kazama, 2007).

The FA can be expressed as:

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

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

The correlation coefficient matrix measures how well the variance of each constituent can be explained by relationships with each other (Liu *et al.,* 2003). The terms “strong”, “moderate”, and “weak” were applied to factor loadings and refer to absolute loading values as >0.75, 0.75-0.50 and 0.50-0.30, respectively, following the approach of Liu *et al.* (2003).

3.2.2 Cluster Analysis (CA)

Cluster analysis is described as a group of multivariate techniques whose primary purpose is to accumulate objects based on the characteristics they possess. Cluster analysis classifies objects, so that each object is analogous to the others in the cluster with regard to a predetermined selection criterion. The resulting clusters of objects are supposed to show high internal homogeneity and high external (between clusters) heterogeneity. Each cluster thus describes, in terms of the data collected, the class to which its members belong; and this description may be abstracted through use from the particular to the general class or type (Einax *et al.,* 1997). Hierarchical agglomerative clustering (HCA) is the most general approach, which presents perceptive similarity between any one sample and the entire data set, and is usually demonstrated by a dendrogram (tree diagram) (McKenna, 2003). The dendrogram presents a visual outline of the clustering processes, presenting a picture of the groups and their proximity, with a dramatic reduction in dimensionality of the original data. The Euclidean distance usually provides the similarity between two samples and a distance can be symbolized by the difference between analytical values from the samples (Otto, 1998).

In this study, HCA was presented on the normalized data set using the Ward's method as agglomeration technique and squared Euclidean distance as a measure of similarity. The Ward's method employs an analysis of variance approach to assess the distances between clusters in an attempt to reduce the Sum of Squares (SS) of any two clusters that can be formed at each step. The spatial inconsistency of water quality in the whole basin was determined from CA, using the linkage distance.

4. Results and Discussion

4.1 Water Quality

The univariate overview of water quality parameters is presented in Table 2. The mean concentration of temperature (temp), TSS, TDS, turbidity, alkalinity, EC, DO, BOD, PO_4^{3-}, COD, SO_4^{2-}, NO_3^-, Cl^- and FC are 28.75±2.47ºC, 26.82±5.91 mg/l, 757.08±118.44 mg/l, 24.19±4.16 NTU, 725 mg/l ($CaCO_3$), 948.50±5 μScm^{-1}, 5.08±0.51 mg/l, 0.93±0.29 mg/l, 48.05 mg/l, 85.05±21.09 mg/l, 470±1.01mg/l, 7.14±0.92 mg/l, 82.25±18.978 mg/l and 58429 x10^7±131182 cfu/100 ml, respectively. The results show that most of the parameters (TSS, TDS, turbidity, alkalinity, EC, PO_4^{3-}, DO, BOD, COD, SO_4^{2-}, NO_3^-, Cl^- and FC) express significant changes, whereas temp and pH show minimum changes in all cases. TSS, TDS, turbidity, EC, BOD, COD and FC show the highest concentration in the range of 8.50-50.00 mg/l, 225-1364 mg/l, 8.97-38.00 NTU, 450-1325 μScm^{-1}, 0.02 to 2.77 mg/l, 8.50-105 mg/l and 5006×10^7 to 430105×10^7 cfu/100 ml, respectively, which are 2-3 folds of national (DoE, 1997) and international (WHO, 2004) standards (Table 2). These parameters increase in the low flow period due to evaporative effects, whereas lower values are observed in the high flow period as the surface water is diluted by rain water. Some input of salt and ions are also contributed by human activities. Noticeable depletion of DO and alarming level of COD concentrations are recorded at all points, indicative of potential ecological and environmental risk. The sharp decline in DO may have resulted from introduction of organic matter in the water which consumed oxygen during decomposition (Masamba and Mazvimavi, 2008). The decrease in DO is generally accompanied by a corresponding decrease in pH. FC in low flow period is always found to be higher than in the high flow season. Similar results have been found elsewhere (Palamuleni *et al.,* 2004; Elmanama *et al.,* 2006). However, the values obtained in this work are still quite high compared to other values for open waters (Masamba and Mazvimavi, 2008). The values of FC in Gangetic River system were found to be 2-1600 cfu/100 ml (Baghel *et al.,* 2005). The range of 0-642 cfu/100 ml FC was evaluated in Oregon (Bracken *et al.,* 2006), whereas the value of 345-2558 cfu/ml were found in Kitwe Stream, Zambia (Ntengwe 2006) and 154 1000 CFU/100 ml were recorded in Marimba River, Zimbabwe (Mvungi *et al.,* 2003). Fatoki *et al.* (2001) determined 200-1000 CFU/100 ml on the Umtata River in South Africa. In Malawi, FC was found at a range of 400-18,500 CFU/100 ml for Lunzu Stream (Palamuleni *et al.,* 2004). The high values of FC and TC in these study areas are attributed to informal settlements with improper sanitation.

4.2 Cluster Analysis

4.2.1 Spatial Similarity and Site Grouping

Dendrogram derived from R-mode CA (Fig. 2a) shows that all sampling sites on the stream and wetlands are grouped into three statistically significant clusters. Group 1 consists of stations TB, AS and AB. Group 2 consists of stations AMT, NUMT, AC, AW, NP and BR. Group 3 consists of only BE. The group classifications change with significant level because the sites in these groups have similar characteristic features and natural backgrounds that are affected by similar sources.

Group 1, clustered by TB, AS and AB, corresponds to relatively High Polluted (HP) sites. The stations are situated at extreme upstream and downstream sites of the stream. These stations receive pollution from point sources, mostly from industrial and municipal activities. Group 2 corresponds to relatively Moderate Pollution (MP) sites. In group 2, stations are situated in between upstream and downstream sites of the study area. These stations are also contaminated with industrial effluents. Group 3 (consisting of only BE) also corresponds to relatively high polluted sites (HP). This station is highly influenced by urban sewage system of Dhaka City Corporation (DCC). The station receives pollution from point and nonpoint sources, i.e., agricultural and livestock farms, domestic wastewater and surface runoff from villages.

The cluster analysis has provided a useful classification of the surface watercourses in the study area, which could be used to design an optimal future spatial monitoring network with lower cost (Simeonov *et al.,* 2003; Singh *et al.,* 2004). On the basis of our results, the number of monitoring sites could be reduced and chosen only from groups 1, 2 and 3.

4.2.2 Grouping/affinity Among Different Variables

R-mode cluster analysis performed on the measured basic water quality parameters reveals two distinct groups or clusters for the annual average data (Fig. 2b). For annual average (Fig. 2b), cluster 1 includes TDS, turbidity, DO, Cl^-, temperature, COD and SO_4^{2-} whereas cluster 2 includes TSS, EC, alkalinity, pH, PO_4^{3-}, NO_3^-, BOD, FC and TC. However, the presence of nutrients under usual conditions supports the growth of bacteria as well as other microorganisms, leading to higher BOD levels as well as a corresponding increase in the fecal coliform count (Elmanama *et al.,* 2006). This phenomenon has also been observed by different researchers (e.g., Jannasch 1968; Rozen and Belkin, 2001). From these clusters, it is difficult to identify the individual groups which come from a single source; rather, they reflect a complex assimilation of industrial (COD, EC, Cl^-, temperature), agriculture (PO_4^{3-}, SO_4^{2-}, TSS, turbidity) and urban sewage (FC) sources. Among them, turbidity materials often come from soil erosion in the Pleistocene Terrace and agriculture fields at the upstream areas. These parameters come

Table 2. Description of Water Quality Parameters (for 12 months period) of Respective Sampling Sites

Name of Station		Temp	TSS	TDS	Turbidity	pH	Alkal	EC	DO	BOD	COD	PO_4^{3-}	SO_4^{2-}	NO_3^-	Cl^-	FC	TC
		°C	mg/l	mg/l	NTU	pH unit	mg/l	µS/cm	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	cfu/100ml	cfu/100ml
Tongi Bridge (TB)	Min	25.00	12.60	252.00	9.70	6.90	200.00	780.00	2.50	0.06	21.00	3.29	47.00	2.58	39.00	68500	5048
	Max	32.00	45.00	1350.00	36.50	8.20	1233.00	1320.00	6.10	2.15	105.00	7.25	151.00	12.70	198.00	676760	1344000
	Mean	28.40	31.05	784.33	23.37	7.48	763.17	1043.33	4.83	0.90	72.50	5.22	100.00	7.41	84.92	430105	946782
	SD	1.90	9.98	397.30	9.42	0.31	424.71	183.20	1.12	0.63	28.22	1.44	37.55	3.64	41.19	372630	674524
Ashulia (As)	Min	24.00	8.50	293.00	17.25	7.20	200.00	552.00	2.75	0.02	21.36	3.11	38.00	2.75	49.00	6513	4720
	Max	31.00	39.00	1367.00	38.00	7.90	1250.00	1275.00	6.00	2.52	96.67	6.89	145.00	12.51	110.00	57500	1150000
	Mean	27.28	26.54	838.33	27.14	7.47	757.50	983.25	4.72	0.89	56.57	4.62	94.42	7.42	81.67	36053	809835
	SD	1.80	7.84	374.47	6.28	0.18	373.27	210.82	1.13	0.86	27.13	1.24	33.71	3.35	21.38	32007	577360
Ashulia M Town (AMT)	Min	25.00	12.74	225.00	12.57	7.20	200.00	500.00	3.00	0.01	18.64	2.75	14.00	2.23	38.00	1520	1350
	Max	31.00	35.00	1350.00	35.31	7.80	1175.00	1200.00	6.00	2.12	83.01	6.10	144.00	12.10	214.00	34500	2563000
	Mean	27.95	26.00	745.83	24.65	7.43	703.75	891.25	4.83	0.75	46.47	4.51	77.42	6.61	92.08	23320	1811360
	SD	2.00	7.62	415.67	7.00	0.16	296.98	241.28	1.02	0.69	23.46	1.21	34.02	3.81	42.81	18010	1282175
New Uttara Model Town (NUMT)	Min	25.00	8.50	314.00	11.76	7.10	250.00	550.00	3.21	0.06	15.47	2.69	38.00	2.98	36.00	857	2500
	Max	32.00	39.25	1307.00	32.00	7.80	1200.00	1250.00	6.50	2.10	95.14	5.97	134.00	10.78	179.00	11500	3200000
	Mean	27.70	23.40	762.83	23.24	7.43	687.08	907.67	5.17	0.80	46.60	4.12	77.75	6.49	88.75	7526	2260974
	SD	2.20	9.73	387.62	6.10	0.18	295.77	232.60	1.05	0.71	29.27	0.97	30.47	2.74	42.10	6179	1601250
Ashulia channel (AC)	Min	25.00	9.90	316.00	12.47	7.20	200.00	450.00	2.75	0.07	14.97	2.15	35.00	2.79	33.00	553	2170
	Max	31.00	36.25	1289.00	31.25	7.80	1150.00	1100.00	6.30	2.15	85.44	6.50	103.00	12.68	187.00	10400	2860000
	Mean	27.90	22.58	802.58	23.12	7.51	679.17	780.83	5.02	0.89	38.38	4.63	70.17	7.12	90.58	6963	2020791
	SD	1.90	7.88	357.89	5.93	0.17	289.56	212.02	1.16	0.79	25.03	1.43	22.09	3.57	45.64	5477	1431085
Ashulia wetland (AW)	Min	24.00	11.25	367.00	8.97	7.00	200.00	550.00	3.65	0.01	11.61	1.57	39.00	2.12	38.00	225	3500
	Max	32.00	41.32	1308.00	30.50	7.80	1100.00	1250.00	6.40	2.13	65.00	6.21	145.00	10.64	168.00	8500	8700000
	Mean	27.65	26.17	840.83	22.98	7.48	666.67	897.92	5.44	0.89	30.87	3.96	89.92	6.62	81.83	5851	6149354
	SD	2.30	9.55	351.32	6.83	0.22	250.76	238.24	0.92	0.72	14.19	1.63	35.94	3.05	38.76	4363	4351750
Berulia (BE)	Min	24.00	17.00	380.00	11.90	7.30	300.00	760.00	3.00	0.04	8.50	2.00	35.00	2.11	47.00	4650	1100
	Max	31.00	50.00	1258.00	34.50	7.80	1250.00	1250.00	6.60	2.55	53.00	6.63	141.00	12.25	149.00	53400	4500000
	Mean	27.29	34.92	761.67	21.25	7.54	854.58	1011.33	4.88	1.13	28.97	4.16	80.33	7.60	85.67	34471	3181203
	SD	2.40	10.06	315.00	8.22	0.15	329.73	146.13	1.27	0.78	15.72	1.62	29.69	3.49	34.14	29025	2250550
National Park (NP)	Min	23.00	14.41	354.00	11.56	7.10	150.00	625.00	3.12	0.04	8.50	2.17	32.00	2.75	25.00	1420	1100
	Max	30.50	39.48	1276.00	30.11	7.80	1100.00	1200.00	6.20	2.77	41.50	6.56	139.00	12.27	111.00	8500	4500000
	Mean	27.16	24.10	760.50	22.37	7.43	710.00	925.75	5.24	0.91	23.59	4.57	76.67	6.86	68.83	5006	3181203
	SD	2.40	7.90	353.40	6.02	0.20	276.70	200.76	1.14	0.87	11.68	1.33	32.80	3.22	29.88	4960	2250550
Bangshi River (BR)	Min	24.00	8.70	388.00	15.70	7.30	200.00	675.00	3.20	0.02	15.23	2.59	42.00	2.10	42.00	2560	3200
	Max	32.00	37.18	1107.00	30.23	7.80	1100.00	1275.00	6.30	2.10	96.12	7.15	141.00	12.41	158.00	12500	2300000
	Mean	28.00	23.34	723.67	23.08	7.51	684.17	937.08	5.31	0.91	43.49	4.87	95.08	7.79	95.67	7029	1624083
	SD	2.60	8.43	260.56	4.94	0.16	283.40	214.76	1.01	0.78	27.39	1.38	31.54	3.57	35.42	7530	1151600
Amin Bazar (AB)	Min	25.00	10.10	362.00	10.83	7.20	400.00	908.00	3.00	0.02	17.00	3.67	31.00	2.87	40.00	4250	4800
	Max	32.00	44.00	1011.00	32.00	7.90	1245.00	1325.00	6.70	1.78	102.00	7.58	137.00	13.50	156.00	43800	1250000
	Mean	28.00	28.57	673.33	23.67	7.46	890.42	1100.33	4.95	0.72	50.01	5.38	78.75	7.61	93.00	27966	880489
	SD	2.10	10.32	195.88	6.82	0.19	255.96	131.42	1.28	0.65	33.14	1.28	30.82	3.64	38.01	24025	627400
Mean																	
Bangladesh Standard		28-30	10.00	1000.00	10.00	6.5-8.5	200-500	700	6.00	0.20	8.00	6.00	400.00	10.00	150-600.00	1000.00	
WHO (2004) standard		25.00	-	1000.00	10.00	6.5-8.5	200.00	150.00	-	6.00	10.00	-	250.00	50.00	250.00	0	0

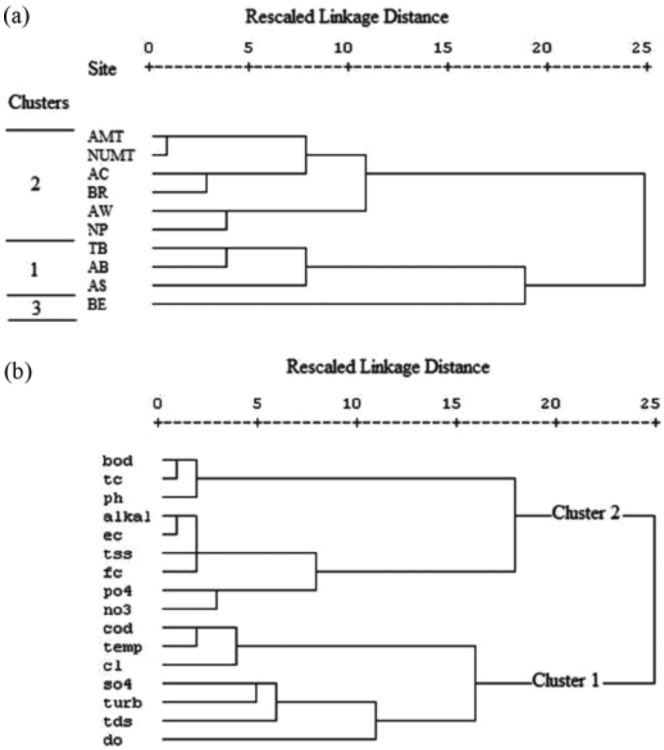

Fig. 2. Dendrograms Showing the Clustering of (a) Sampling Sites and (b) Water Quality Parameters

from the organic matter and micronutrients within the basin area. It can be concluded that the water quality parameters of the study area are dominated by anthropogenic sources.

4.3 Principal Component Analysis (PCA) and Factor Analysis (FA)

The rotation of the principal components was executed by the varimax method with Kaiser Normalization. Five VFs are obtained for water quality parameters through FA performed on the PCs, which indicates that five main controlling factors influenced the quality of surface water in the study area. Corresponding VFs, variable loadings, and the variance explained are presented in Table 3.

Varifactor 1 (VF1) explains 26.0% of total variance and is positively loaded with inorganic, organic and mineral related parameters (i.e., PO_4^{3-}, NO_3^{1-}, TSS, alkalinity, EC and FC), and negatively and weakly participated by DO (Table 3). The negative factor loading of DO on this factor suggests utilization of dissolved oxygen to decompose the organic matter by bacterial (FC) function (Singh *et al.,* 2004). These factors are the input from anthropogenic (urban sewage and agricultural) and natural sources. VF2, which explains 22.4% of total variance, is loaded with physical and biological parameters (i.e., BOD, pH and TC) and negatively loaded with turbidity. These factors indicate dominance of domestic and urban sewage sources in the river system. VF3 explains 17.3% of total variance and loaded with COD, PO_4^{3-}, Cl^- and temperature. These factors represent the contribution of point and nonpoint pollution from industrial and agricultural areas. VF4 explains 11.8% and is loaded with SO_4^{2-}, TDS and turbidity. These factors represent the contribution of nonpoint pollution from agricultural and soil erosion processes. In this areas, farmers use ammonium sulfate fertilizers, and the surface runoff receives sulphate via surface runoff and irrigation waters. TDS and turbidities result from soil erosion due to cultivation and monsoon rainfall from upland areas in the river basin. VF5, explaining 10.7% of total variance, is weakly loaded by COD, turbidity and FC and are less significant factors that

Table 3. Rotated Component Matrix of Five-factor Model

R-mode	VF1	VF2	VF3	VF4	VF5
Parameters					
COD	0.222	-0.282	**0.671**	0.395	0.453
PO4	**0.589**	-0.325	**0.621**	0.055	-0.162
NO3	**0.628**	0.450	0.331	0.157	-0.104
SO4	0.289	0.034	0.163	**0.876**	-0.096
CL	-0.116	0.247	**0.829**	-0.105	0.323
TSS	**0.728**	0.399	-0.155	0.109	0.258
TDS	-0.477	0.186	-0.374	0.673	0.129
TURB	-0.014	***-0.635***	0.184	0.503	0.418
PH	0.078	**0.902**	0.185	0.222	-0.065
ALKAL	**0.846**	0.206	-0.015	-0.288	0.215
EC	**0.952**	0.001	-0.014	0.110	0.029
DO	-0.322	0.079	-0.126	0.050	-0.908
BOD	0.053	**0.946**	-0.115	0.021	-0.092
TEMP	0.017	-0.096	**0.887**	0.045	-0.048
FC	**0.823**	0.063	0.172	0.204	0.435
TC	0.315	**0.878**	-0.172	-0.144	0.179
Eigen value	4.156	3.586	2.766	1.891	1.721
% of total variance	25.976	22.414	17.287	11.819	10.757
Cumulative % of variance	25.976	48.390	65.678	77.496	88.253
R-mode					
Sampling site	VF1	VF2	VF3	VF4	VF5
TB	**0.996**	-0.084	**0.846**	**1.089**	0.083
AS	0.336	-0.658	-0.654	**1.747**	**1.368**
AMT	-0.523	-1.057	0.475	-0.814	**0.855**
NUMT	-0.886	-0.245	-0.091	-0.819	**0.994**
AC	-1.468	**0.660**	**0.854**	-0.448	0.139
AW	-1.070	0.063	-0.878	**0.971**	-0.994
BE	**0.854**	**2.287**	-0.996	-0.575	**0.632**
NP	0.245	-0.946	-1.661	-0.726	-1.362
BR	-0.134	**0.626**	**1.320**	**0.603**	-1.431
AB	**1.649**	-0.646	**0.786**	-1.027	-0.287

have little contribution to the variability of water quality.

Relatively high scores for VF1 occur at stations AB, TB and BE (Table 3), suggesting potential risk from inorganic, organic and mineral related parameters. High scores for VF2 are observed at BE, AC and BR, indicating pollution risks due to inputs from domestic and urban sewage sources. Stations BR, AC, TB and AB are at risk from pollution due to industrial and agricultural inputs, as they retain high scores for VF3. High scores for VF4 and VF5 are observed at stations AS, TB, AW and BR, and AS, NUMT, AMT and BE, respectively, suggesting pollution from agricultural and soil erosion processes.

The sites groupings have been explored in the plot of the first two principal components generated from the annual average parameters (Fig. 3). The sampling stations can be grouped into 3: Group 1 (AS, AB, TB, NP), Group 2 (AC, AW, BR, NUMT, AMT) and Group 3 (BE). The PCA groupings compare very well with the Q-mode cluster analysis.

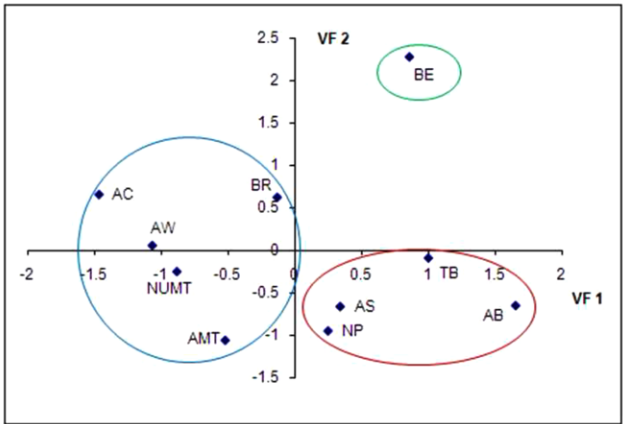

Fig. 3. Scores of Water Quality Parameters on the Bi-dimensional Plane Defined by the First Two Varifactors

The factor analysis (supported by cluster analysis) has led to the preliminary observations discussed below. On the basis of the water quality parameters determined in this study, the sampling sites can basically be categorized into 3 groups. Group 1 [clustered by Tongi Bridge (TB), Ashulia (AS) and Amin Bazar (AB)] is loaded with PO_4^{3-}, NO_3^-, Cl^-, COD, SO_4^{2-}, EC, TDS, alkalinity, temperature, turbidity and FC and corresponds to relatively high polluted (HP) sites. The stations are situated at extreme upstream and downstream sites of the stream and receive pollution from point sources, i.e., mostly from industrial and municipal activities. Group 2 [Ashulia Model Town (AMT), New Uttara Model Town (NUMT), Ashulia Channel (AC), Ashulia Wetland (AW), National Park (NP) and Bangshi River (BR)] is loaded with PO_4^{3-}, Cl^-, COD, temperature and corresponds to relatively moderate pollution (MP) sites. In group 2, stations are situated in between upstream and downstream sites of the study area, and are contaminated with industrial effluents. Group 3 (Berulia; BE) is loaded with PO_4^{3-}, NO_3^-, EC, pH, TSS alkalinity, BOD, TC and FC and also corresponds to relatively High Polluted sites (HP). This station is highly influenced by urban sewage system of Dhaka City Corporation (DCC). These stations receive pollution from point and nonpoint sources (i.e., agricultural and livestock farms, domestic wastewater, surface runoff from villages).

Similar studies have been carried out on river systems in other countries with different extract of principal components and overall

Table 4. Pearson Correlation Matrix (CM) of Water Quality Parameters

Parameter	TSS	TDS	TURB	PH	ALKAL	EC	DO	BOD	COD	PO4	NO3	SO4	CL	TEMP	FC	TC
TSS	1															
TDS	-0.019	1														
TURB	-0.31	-0.084	1													
pH	0.014	-0.241	0.288	1												
ALKAL	0.038	0.42	0.135	-0.353	1											
EC	-0.04	0.766**	-0.481	-0.075	0.231	1										
DO	0.11	0.693*	-0.438	0.051	0.033	0.832**	1									
BOD	-0.091	-0.499	0.023	-0.482	-0.039	-0.546	-0.373	1								
COD	-0.447	0.499	0.339	-0.472	0.733*	0.131	0.013	0.007	1							
PO4	0.66*	0.167	-0.025	0.554	-0.118	0.178	0.422	-0.57	-0.368	1						
NO3	0.582	0.097	-0.534	0.215	-0.019	0.441	0.534	-0.37	-0.367	0.611	1					
SO4	0.264	0.255	0.266	0.311	0.208	-0.011	0.438	-0.04	0.172	0.56	0.248	1				
CL	0.15	0.43	-0.337	-0.013	0.66*	0.6	0.569	-0.31	0.361	0.266	0.612	0.399	1			
TEMP	0.656*	-0.042	-0.447	0.078	0.281	0.102	0.009	-0.15	-0.318	0.358	0.277	-0.015	0.293	1		
FC	0.574	0.459	0.094	0.011	0.073	0.179	0.427	-0.38	0.079	0.728*	0.486	0.573	0.231	-0.05	1	
TC	-0.378	-0.027	0.418	-0.475	0.089	-0.362	-0.361	0.69*	0.316	-0.707*	-0.748*	-0.117	-0.498	-0.4	-0.332	1

**Correlation is significant at the 0.01 level, and *is significant at the 0.05 level.

component loadings with respect to the physical parameters analyzed (e.g., Vega *et al.,* 1998; Ouyang *et al.,* 2006). Discrepancies in results could be attributed to the different river environments and different water quality parameters as well as to the different time periods (i.e., seasonal vs. overall) used in each study. Results suggest that water quality variables that play important roles in influencing river water quality in one place may not be important in another place or one season may not be the same as another season.

4.4 Correlation Matrix (CM)

Pearson's correlation matrix (CM) derived from water quality data agreed with the results obtained by PCA. CM reveals some new associations between the parameters (Table 4) that are not adequately reported in the PCA.

TSS has a significant positive relationship with PO_4^{3-} (r = 0.660; P<0.05) and temp (r = 0.656; P<0.001). The strong correlation of parameters indicates a common origin especially from industrial, agricultural and municipal sources. Alkali is significantly correlated with COD(r = 0.733; P<0.05) and Cl^- (r = 0.660; P<0.05) which indicate similar sources of industrial pollution. BOD is positively correlated with TC (r = 0.690; P<0.05), indicating municipal inputs. TDS is strongly correlated with EC (r = 0.766; P<0.01) and significantly correlated with DO (r = 0.693; P<0.05) which suggest that the main contributing factors for dissolved oxygen are photosynthesis activities, decomposition rate of organic matter and assimilation of sewage and urban runoff. NO_3^- is negatively correlated with TC (r = 0.748; P<0.05) which indicates that due to municipal and agricultural input the aquatic environment is significantly polluted and the surface is not conducive for the survival of TC.

5. Conclusions

Surface water quality assessment has been conducted for the Turag River and its surrounding wetlands. In terms of pollution, the analyzed parameters exceed the Bangladesh standard limits except pH, SO_4^{2-} and Cl^-. The water quality data have been analyzed by multivariate statistical techniques (hierarchical cluster analysis and principal component analysis) to extract correlations and similarities between variables and to classify river water samples in groups of similar quality. Q-mode CA has identified three clusters among the 10 sampling sites. Cluster 1 consists of TB, AB and AS; cluster 2 consists of AMT, NUMT, AC, AW, NP and BR; cluster 3 consists of only BE. R-mode CA characterizes two groups among the water quality variables which move together. PCA has found a reduced number of "latent"' variables (principal components) that explain most of the variance of the experimental data set. A varimax rotation has been able to reduce the number of varifactors, each of them relating to a small group of experimental variables with a hydrological meaning. The varifactors have identified pollution sources affecting water quality in the study area as anthropogenic (i.e., industrial, agricultural, urban sewage) and natural (soil erosion, aquatic hyacinths and weeds) inputs.

The study has demonstrated the usefulness of multivariate statistical techniques (cluster and principal component analyses) in water quality monitoring that may play significant role in water management. It is, therefore, recommended that multivariate statistical methods are integrated in future studies on pollution risk assessment of water environment in Bangladesh.

Acknowledgements

The first author gratefully acknowledges the Japan Ministry of Education, Culture, Sports, Science and Technology (Monbukagakusho Scholarship 2007-2010), for financial support. Earnest gratitude goes to the anonymous reviewers whose constructive remarks significantly improved the manuscript.

References

APHA (1995). *Standard methods for the examination of water and wastewater,* 19th Ed., Washington, D.C.: American Public Health Association.

Aruga, R., Gastaldi, D., Negro, G., and Ostacoli, G. (1995). "Pollution of a river basin and its evolution with time studied by multivariate statistical analysis." *Anal. Chim. Acta*, Vol. 310, No. 1, pp. 15-25.

Astel, A., Glosiska, G., Sobczyski, T., Boszke, L., Simeonov, V., Siepak, J. (2006). "Chemometrics in assessment of sustainable development rule implementation." *CEJCh.*, Vol. 4, No. 3, pp. 543-564.

Astel, A., Tsakovski, S., Barbieri, P., and Simeonov, V. (2007). "Comparison of self- organizing maps classification approach with cluster and principal components analysis for large environmental data sets." *Water Res*., Vol. 41, No. 19, pp. 4566 - 4578.

Astel, A. Tsakovski, S., Simeonov, V., Reisenhofer, E., Piselli, S., and Barbieri, P. (2008). "Multivariate classification and modeling in surface water pollution estimation." *Anal. Bioanal. Chem.*, Vol. 390, No. 5, pp. 1283-1292.

Baghel, V. S., Gopal, K., Dwivedi, S., and Tripathi, R. D. (2005). "Bacterial indicators of faecal Contamination of the Gangetic river system right at its source." *Ecol. Indicat.,* Vol. 5, No. 1, pp. 49-56.

Bracken, C. L., Hendricks, C. W., and Harding, A. K. (2006). "Apparent bias in river water inoculum following centrifugation." *J. Microbio. Method*., Vol. 67, No. 2, pp. 304-309.

Clesceri, L. S., Greenberg, A. E., and Eaton, A. D. (1998). *Standard methods for the examination of water and wastewater*, 20th Ed. American Public Health Association, Washington, D.C.

DoE (1997). "*Industrial effluents quality standard for Bangladesh*." Bangladesh Gazette Additional.

Einax, J. W., Zwanziger, H. W., and Geib, S. (1997). *Chemometrics in environmental analysis*, Wiley, Weinheim.

Elmanama, A. A., Afifi, S., and Bahr, S. (2006). "Seasonal and spatial variation in the monitoring parameters of Gaza Beach during 2002-2003." *Environ. Res*., Vol. 101, No. 1, pp. 25-33.

Fatoki, O. S., Muyima, N. Y. O., and Lujiza, N. (2001). "Situation analysis of water quality in the Umtata River catchment." *Water SA,* Vol. 27, No. 4, pp. 467-473.

Helena, B., Pardo, R., Vega, M., Barrado, E., Fernández, J. M., Fernández, L. (2000). "Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis." *Water Res.,* Vol. 34, No. 3, pp. 807-816.

Howitt, D. and Cramer, D. (2005). *Introduction to SPSS in psychology: With supplement for releases 10, 11, 12 and 13*, Pearson, Harlow.

Jannasch, H. (1968). "Competitive elimination of Enterobacteriaceae from seawater." *Appl. Microbio.,* Vol. 16, No. 10, pp. 1616-1618.

Johnston, M. W. and Williams, J. S. (2006). *Field comparison of optical and clark cell dissolved oxygen sensors in the tualatin river*, Oregon, 2005, U.S. Geological Survey Open-File Report 2006-1047, p. 11.

Kowalkowski, T., Zbytniewski, R., Szpejna, J., and Buszewski, B. (2006). "Application of chemometrics in river water classification." *Water Res.,* Vol. 40, No. 1, pp. 744-752.

Liu, C. W., Lin, K. H., and Kuo, Y. M. (2003). "Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan." *Sci. Tot. Environ.,* Vol. 313, Nos. 1-3, pp. 77-89.

Lu, R. S., and Lo, S. L. (2002). "Diagnosing reservoir water quality using self-organizing maps and fuzzy theory." *Water Res.,* Vol. 36, No. 9, pp. 2265-2274.

Marengo, E., Gennaro, M. C., Giacosa, D., Abrigo, C., Saini, G., and Avignone, M. T. (1995). "How chemometrics can helpfully assist in evaluating environmental data Lagoon water." *Anal. Chim. Acta,* Vol. 317, Nos. 1-3, pp. 53-63.

Masamba, W. R. L. and Mazvimavi, D. (2008). "Impact on water quality of land uses along Thamalakane-Boteti River: An outlet of the Okavango Delta." *Physic. Chem. Earth*, Vol. 33, pp. 687-694.

McKenna, J.r., J. E. (2003). "An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis." *Environ. Mode. Softw.,* Vol. 18, No. 3, pp. 205-220.

Mendiguchía, C., Moreno, C., Galindo-Riaòo, D. M., and García-Vargas, M. (2004). "Using chemometric tools to assess anthropogenic effects in river water. A case study: Guadalquivir (Spain)." *Anal. Chim. Acta,* Vol. 515, No. 1, pp. 143-149.

Mvungi, A., Hranova, R. K., and Love, D. (2003). "Impact of home industries on water quality in a tributary of the Marimba River, Harare: implications for urban water management." *Physic. Chem. Earth,* Vol. 28, Nos. 20-27, pp. 1131-1137.

Nkansah, K., Dawson-Andoh, B., Slahor, J. (2010). "Rapid characterization of biomass using near infrared spectroscopy coupled with multivariate data analysis: Part 1 yellow-poplar (*Liriodendron tulipifera* L.)." *Bioresour. Technol*., Vol. 101, No. 2, pp. 4570- 4576.

Ntengwe, F. W. (2006). "Pollutant loads and water quality in streams of heavily populated and industrialized towns." *Physic. Chem. Earth*, Vol. 31, Nos. 15-16, pp. 832-839.

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 Res.,* Vol. 40, No. 20, pp. 3800-3810.

Palamuleni, L. G., Dolozi, M .B., Masamba, W. R. L., and Claudio-Jeke, A. (2004). "Bacteriological contamination of water in urban poor areas: A case study of South Lunzu Township, Blantyre Malawi." *Malawi J. Sci. Techno*., Vol. 7, No. 1, pp. 26-33.

Qadir, A., Malik, R. N., and Husain, S. Z. (2007). "Spatio-temporal variations in water quality of Nullah Aik-tributary of the river Chenab, Pakistan." *Environ. Monit. Assess.*, Vol. 140*,* Nos. 1-3, pp. 43-59.

Reghunath, R., Murthy, T. R. S., and Raghavan, B. R. (2002). "The utility of multivariate statistical techniques in hydrogeochemical studies: An example from Karnataka, India." *Water Res*., Vol. 36, No. 10, pp. 2437-2442.

Rozen, Y. and Belkin, S. (2001). "Survival of enteric bacteria in seawater." *FEMS Microbio. Rev.,* Vol. 25, No. 5, pp. 513-529.

Sarbu, C. and Pop, H. F. (2005). "Principal component analysis versus fuzzy principal component analysis. A case study: The quality of Danube water (1985e 1996)." *Talanta,* Vol. 65, No. 5, pp. 1215-1220.

Shrestha, S. and Kazama, F. (2007). "Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan." *Environ. Mode. Softw.,* Vol. 22, No. 4, pp. 464-475.

Simeonov, V., Stefanov, S., and Tsakovski, S. (2000). "Environmetrical treatment of water quality survey data from Yantra River, Bulgaria." *Microch. Acta*, Vol. 134, Nos. 1-2, pp. 15-21.

Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsa, D., and Anthemidis, A. (2003). "Assessment of the surface water quality in Northern Greece." *Water Res.*, Vol. 37, No. 17, pp. 4119-4124.

Simeonova, P., and Simeonov, V. (2007). "Chemometrics to evaluate the quality of water sources for human consumption." *Microch. Acta*, Vol. 156, Nos. 3-4, pp. 315-320.

Singh, K. P., Malik, A., Mohan, D., and 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 Res.,* Vol. 38, No. 18, pp. 3980-3992.

Stefanov, S., Simeonov, V., and Tsakovski, S. (1999). "Chemometrical analysis of waste water monitoring data from Yantra river basin, Bulgaria." *Toxicol. Environ. Chem.*,Vol. 70, pp. 473-482.

Tabachnick, B. G. and Fidell, L. S. (2007). *Using multivariate statistics*, Pearson/Allyn and Bacon, London.

USEPA (1985). *Test methods for Escherichia coli and enterococci in water by the membrane filter procedure (Method #1103.1)*, EPA 600/4-85-076, U.S. Environmental Protection Agency, Environmental Monitoring and Support Laboratory, Cincinnati, OH.

Vega, M., Pardo, R., Barrado, E., and Deban, L. (1998). "Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis." *Water Res.,* Vol. 32, No. 12, pp. 3581-3592.

WHO (2004). *Guidelines for drinking-water quality*, Third Ed. WHO, Geneva, Switzerland.

Wunderlin, D. A., Diaz, M. P., Ame, M. V., Pesce, S. F., Hued, A. C., and 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 Res.,* Vol. 35, No. 12, pp. 2881-2894.