Appl Water Sci (2017) 7:3511–3525
DOI 10.1007/s13201-017-0619-0

ORIGINAL ARTICLE

Assessment of roadside surface water quality of Savar, Dhaka, Bangladesh using GIS and multivariate statistical techniques

Fahad Ahmed[1] · A. N. M. Fakhruddin[1] · MD. Toufick Imam[1] · Nasima Khan[2] · Abu Tareq Mohammad Abdullah[2] · Tanzir Ahmed Khan[2] · Md. Mahfuzur Rahman[2] · Mohammad Nashir Uddin[3]

Received: 21 January 2015 / Accepted: 7 September 2017 / Published online: 18 September 2017

Abstract In this study, multivariate statistical techniques in collaboration with GIS are used to assess the roadside surface water quality of Savar region. Nineteen water samples were collected in dry season and 15 water quality parameters including TSS, TDS, pH, DO, BOD, Cl^-, F^-, NO_3^{2-}, NO_2^-, SO_4^{2-}, Ca, Mg, K, Zn and Pb were measured. The univariate overview of water quality parameters are TSS 25.154 $\pm$ 8.674 mg/l, TDS 840.400 $\pm$ 311.081 mg/l, pH 7.574 $\pm$ 0.256 pH unit, DO 4.544 $\pm$ 0.933 mg/l, BOD 0.758 $\pm$ 0.179 mg/l, Cl^- 51.494 $\pm$ 28.095 mg/l, F^- 0.771 $\pm$ 0.153 mg/l, NO_3^{2-} 2.211 $\pm$ 0.878 mg/l, NO_2^- 4.692 $\pm$ 5.971 mg/l, SO_4^{2-} 69.545 $\pm$ 53.873 mg/l, Ca 48.458 $\pm$ 22.690 mg/l, Mg 19.676 $\pm$ 7.361 mg/l, K 12.874 $\pm$ 11.382 mg/l, Zn 0.027 $\pm$ 0.029 mg/l, Pb 0.096 $\pm$ 0.154 mg/l. The water quality data were subjected to R-mode PCA which resulted in five major components. PC1 explains 28% of total variance and indicates the roadside and brick field dust settle down (TDS, TSS) in the nearby water body. PC2 explains 22.123% of total variance and indicates the agricultural influence (K, Ca, and NO_2^-). PC3 describes the contribution of nonpoint pollution from agricultural and soil erosion processes (SO_4^{2-}, Cl^-, and K). PC4 depicts heavy positively loaded by vehicle emission and diffusion from battery stores (Zn, Pb). PC5 depicts strong positive loading of BOD and strong negative loading of pH. Cluster analysis represents three major clusters for both water parameters and sampling sites. The site based on cluster showed similar grouping pattern of R-mode factor score map. The present work reveals a new scope to monitor the roadside water quality for future research in Bangladesh.

Keywords Surface water · Principal component analysis · Cluster analysis · Geographical information system (GIS) · Savar

✉ Abu Tareq Mohammad Abdullah
tareq_dubd@yahoo.com

1 Department of Environmental Sciences, Jahangirnagar University, Savar, Dhaka 1342, Bangladesh

2 Institute of Food Science and Technology, Bangladesh Council of Scientific and Industrial Research (BCSIR), Dhaka, Bangladesh

3 Bangladesh Council of Scientific and Industrial Research (BCSIR), Dhaka, Bangladesh

Introduction

Worldwide deterioration of surface water quality has been attributed to both natural processes and anthropogenic activities, including hydrological features, climate change, precipitation, agricultural land use, and sewage discharge (Ravichandran 2003; Gantidis et al. 2007; Kundewicz et al. 2007; Arain et al. 2008). Information on water quality and pollution sources is important for the implementation of sustainable water-use management strategies (Crosa et al. 2006; Sarkar et al. 2007; Zhou et al. 2007).Many different sources and processes are known to contribute to the deterioration in quality and contamination of water. Thus, a thorough understanding of the nature and extent of contamination in an area requires detailed hydro chemical data (Helena et al. 1999). Unfortunately, very few studies have so far been undertaken combining the effects of multiple water quality variables for evaluating the water quality and the extent and nature of contamination (Shuxia et al. 2003).

In Bangladesh, total environment, as well as economic growth and developments, are all highly influenced by water. In terms of quality, the surface water of the country is unprotected from untreated industrial effluents and municipal waste water, runoff pollution from chemical fertilizers, vehicle emission pollutants, pesticides, etc. (Bhuiyan et al. 2011). 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 system is regarded as a multivariate one (Marengo et al. 1995; Stefanov et al. 1999; Wunderlin et al. 2001; Lu and Lo 2002; Simeonov et al. 2003; Mendiguchía et al. 2004; Astel et al. 2006; Kowalkowski et al. 2006; Simeonova and Simeonov 2007; Astel et al. 2008).

Factor analysis, which includes principal component analysis (PCA) is a very powerful technique applied to reduce the dimensionality of a data set consisting of a large number of interrelated variables while remaining as much as possible the variability present in data set. This reduction is achieved by transforming the data set into a new set of variables, the principal components (PCs), which are orthogonal (non-correlated) and are arranged in decreasing order of importance (Panda et al. 2006). Principal component analysis provides information on the most meaningful parameters, which describe whole data set rendering data reduction with minimum loss of original information (Singh et al. 2004). PCA has allowed the identification of a reduced number of latent factors with pollution sources such as spatial (pollution from anthropogenic origin) and temporal (seasonal and climatic) sources of variation affecting quality and hydrochemistry of river water have been differentiated and assigned to polluting sources (Shrestha and Kazama 2007; Simeonov et al. 2003; Kowalkowski et al. 2006; Pekey et al. 2004; Vega et al. 1998). At the same time PCA has allowed the explaining of related parameters by only one factor (Boyacioglu and Boyacioglu 2006; Kannel et al. 2007; Kottı et al. 2005; Kowalkowski et al. 2006; Singh et al. 2004) and exposing the important factor responsible for seasonal changes in river water quality (Ouyang 2005; Ouyang et al. 2006). Cluster analysis helps in grouping objects (cases) into classes (clusters) on the basis of similarities within a class and dissimilarities between different classes. The class characteristics are not known in advance but may be determined from the analysis. The results of CA help in interpreting the data and indicate patterns (Vega et al. 1998). PCA, FA, and CA will be excellently used in future studies to find inter-parameter associations existing between different pollutants. This data-mining technique will further help in reducing the number of pollution parameters to be tested and subsequent cost of analysis (Ahmed et al. 2016).

Dhaka Aricha Highway, Savar has high traffic density and industrial influence. These factors are deteriorating the surface water quality which possesses a potential environmental risk. So far, no study regarding roadside surface water quality is conducted in Bangladesh. The study provides the information about roadside water state and its hydro chemical data to gain knowledge about the status of roadside surface water in the North-Western part of Dhaka city. The water quality data of the area is subjected to PCA, FA and CA to better interpret, understand and define the anthropogenic processes and specific source of water quality deterioration and contamination in the area.

Study area

The study area is along the Dhaka Aricha Road lies between latitudes from 23°47′45.84″N to 23°47′40.08″N and longitude 90°16′36.04″E to 90°19′33.80″E which is 5.04 km long (Fig. 1). The study area is situated near Savar which is 17 km north of Dhaka center runs northward (Aktaruzzaman et al. 2013). The study area was selected because it links with Dhaka city with comparatively high traffic density and has industrial influence. It carries, on an average, 9000 motor vehicles per day. The study area is surrounded by numerous brick fields and Landfill area near Aminbazar. Gabtoli-Amin bazar area is the transition point of Dhaka city, largest bus stand acting as entry and exit point from the city. The average elevation is 26.5 ft above sea level and perennially inundated by monsoon flood and roadside runoff during monsoon. The geology of this area is uplifted in Madhupur area which is covered by dark reddish-brown to brownish red, mottled, sticky and compact Madhupur Clay Residuum of Pleistocene age, underlain by Plio-Pleistocene Dupitila Sandstone Formation (Maitra and Akhter 2011).

Materials and methods

Sample collection

During dry seasons (January 2014), total of 19 water samples (prefixed S) were collected from roadside surface water of Hemayetpur to Gabtoli area, Dhaka Aricha Highway. Sample bottles were preconditioned with 5% nitric acid and rinsed with distilled deionized water. Each sample was collected from 0 to 10 cm of roadside surface water by 500 ml plastic bottle. Duplicate samples were taken per each sampling and later transported to the laboratory for analysis. Samples were transferred to the laboratory in coolers containing ice to reduce the degradation of samples before analysis. Samples were preserved at 4 °C for Anion determination.

Methods for chemical and physical analyses

The geographic positions of the sample sites were recorded by GPS (Explorist, model: 200). pH was measured in situ with a portable meter (HANNA Instruments, model: pH 211.Romannia). Clark cell probes method was used for DO and BOD measurements (Johnston and Williams 2006). TSS measurement was conducted by a simple gravimetric method. Anions (Cl^-, F^-, NO_3^{-1}, NO_2^{-1} and SO_4^{2-}) were analyzed by Ion Exchange Chromatography (DIONEX ICS-3000 Series, USA, Software version Chromeleon 6.80). Water quality parameters, their units and methods of analysis are summarized in Table 1.

Elemental analysis

Metals (Ca, Mg, K, Zn and Pb) were determined by Flame Atomic Absorption Spectrometer (AAS, Thermo scientific iCE-3000 Series, USA, Software version SOLAAR Data Station V11.02). The accuracy and precision of the AAS analytical method were validated by duplicate analyses of AA standard: Ultra Scientific Analytical Solution, a standard reference material, Matrix water with dilute 2% HNO_3. Sample analysis procedure is adopted from Thermo Scientific Cook Book, according to application note of Thermo Fisher scientific supplied along with AAS, iCE-3000 Series.

Chemicals and reagents

Reference standard for ion exchange chromatography (IC) were Dionex Seven Anion Standard 2, Deionized Water, part #57590, approx. Volume 100 ml, Dionex Corporation, USA.

Statistical analysis

For analyzing water quality data were subjected to descriptive multivariate analysis: cluster analysis (CA) and PCA/factor analysis (FA) and CA using SPSS of its version 22.0.

Principal component analysis and factor analysis

PCA is designed to transform the original variables into new, uncorrelated variables (axes), called the 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 (Sarbu and Pop 2005). PC provides information on the most meaningful parameters, which describes a whole data set affording data reduction with minimum loss of original information (Helena et al. 2000). The principal component (PC) can be expressed as:

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

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.

Table 1 Water quality parameters associated with their abbreviations, units and analytical methods

Parameters	Abbreviations	Units	Analytical methods
Total suspended solids	TSS	mg/l	Gravimetric method
Total dissolved solids	TDS	mg/l	Instrumental method
Acidity/alkalinity	pH	pH unit	Instrumental method
Dissolved oxygen	DO	mg/l	Clark cell probes method
Biochemical oxygen demand	BOD_5	mg/l	Clark cell probes method
Chloride	Cl^-	mg/l	Ion exchange chromatography
Fluoride	F^-	mg/l	Ion exchange chromatography
Nitrate	NO_3^{-1}	mg/l	Ion exchange chromatography
Nitrite	NO_2^{-1}	mg/l	Ion exchange chromatography
Sulfate	SO_4^{2-}	mg/l	Ion exchange chromatography
Calcium	Ca	mg/l	Flame atomic absorption spectrometer
Magnesium	Mg	mg/l	Flame atomic absorption spectrometer
Potassium	K	mg/l	Flame atomic absorption spectrometer
Zinc	Zn	mg/l	Flame atomic absorption spectrometer
Lead	Pb	mg/l	Flame atomic absorption spectrometer

The main purpose of FA is to reduce the contribution of less significant variables to simplify even more of the data structure coming from PCA. This purpose can be achieved by rotating the axis defined by PCA, according to well-established rules, and constructing new variables, also called varifactors (VF). PC is a linear combination of observable water quality variables, whereas VF can include unobservable, hypothetical, latent variables (Vega et al. 1998; Helena et al. 2000). PCA of the normalized variables was performed to extract significant PCs and to further reduce the contribution of variables with minor significance; these PCs were subjected to varimax rotation with Kaiser Normalization generating VFs (Brumelis et al. 2000; Singh et al. 2004, 2005; Love et al. 2004; Abdul-Wahab et al. 2005). As a result, a small number of factors will usually account for approximately the same amount of information as do the much larger set of original observations. The FA can be expressed as:

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

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.

Cluster analysis (CA)

The purpose of cluster analysis is to identify groups or clusters of similar sites on the basis of similarities within a class and dissimilarities between different classes (Sparks 2000). It 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, to classify the objects of the system into categories or clusters based on their nearness or similarity (Panda et al. 2006). Cluster analysis is a group of multivariate techniques whose primary purpose is to assemble objects based on the characteristics they possess. Cluster analysis classifies objects, so that each object is similar to the others in the cluster with respect to a predetermined selection criterion. The resulting clusters of objects should then exhibit high internal (within-cluster) homogeneity and high external (between clusters) heterogeneity. Hierarchical agglomerative clustering is the most common approach, which provides intuitive similarity relationships between any one sample and the entire data set, and is typically illustrated by a dendrogram (McKenna 2003). The dendrogram provides a visual summary 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 gives the similarity between two samples and a distance can be represented by the difference between analytical values from the samples (Otto 1998). In this study, hierarchical agglomerative CA was performed on the normalized data set by means of the Ward's method, using squared Euclidean distances 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 (SS) of any two clusters that can be formed at each step. Agglomerative hierarchical clustering is the most commonly used method where clusters are formed sequentially, by starting with the most similar pair of objects and forming higher clusters step by step. Cluster analysis was applied on experimental data standardized through z-scale transformation to avoid misclassification due to wide differences in data dimensionality (Liu et al. 2003). Standardization tends to increase the influence of variables whose variance is small and reduce the influence of variables whose variance is large (Singh et al. 2004).

Inverse distance weighting

The factor scores from the R-mode PCA were used with ArcGIS to determine the spatial variations of the dominant processes influencing the surface water hydrochemistry in the area using inverse distance weighting (IDW) method. The IDW method estimates the values of an attribute at unsampled points using a linear combination of values at sampled points weighted by an inverse function of the distance from the point of interest to the sampled points. The assumption is that sampled points closer to the unsampled point are more similar to it than those further away in their values. The weights can be expressed as:

$$\lambda_i = \frac{1/d_i^p}{\sum_{i=1}^{n} 1/d_i^p},$$

where d_i is the distance between x_0 and x_i, p is a power parameter, and n represents the number of sampled points used for the estimation. The main factor affecting the accuracy of IDW is the value of the power parameter (Isaak and Srivastava 1989). Weights diminish as the distance increases, especially when the value of the power parameter increases, so nearby samples have a heavier weight and have more influence on the estimation, and the resultant spatial interpolation is local (Isaak and Srivastava 1989). The choice of power parameter and neighborhood size is arbitrary (Webster and Oliver 2001). The most popular choice of p is 2 and the resulting method is often called inverse square distance or inverse distance squared (IDS).

Results and discussion

Descriptive study of surface water quality

The univariate overview of water quality parameters are presented in Table 2. The mean concentration of TSS, TDS, pH, DO, BOD, Cl^-, F^-, NO_3^{2-}, NO_2^-, SO_4^{2-}, Ca,

Table 2 Descriptive statistics of surface water quality

	Bangladesh standard	WHO	Minimum	Maximum	Mean	SD
TSS	10	–	10.800	39.280	25.154	8.674
TDS	1000	1000	340.000	1340.000	840.400	311.081
pH	6.5–8.5	6.5–8.5	7.098	7.890	7.574	0.256
DO	6	–	2.298	5.408	4.544	0.933
BOD	0.2	6	0.390	1.120	0.758	0.179
Cl^-	150–600	250	16.363	129.224	51.494	28.095
F^-	–	–	0.431	1.032	0.771	0.153
NO_3^{2-}	10	50	1.243	4.417	2.211	0.878
NO_2^-	–	–	0.874	21.376	4.692	5.971
SO_4^{2-}	400	250	17.165	251.816	69.545	53.873
Ca	75	100	20.740	97.240	48.458	22.690
Mg	–	50	7.660	39.390	19.676	7.361
K	12	12	1.350	44.220	12.874	11.382
Zn	5	0.01	0.013	0.145	0.027	0.029
Pb	0.05	0.01	0.039	0.730	0.096	0.154

Table 3 Rotated component matrix of five factor model with moderate to strong loadings in bold typeface

Parameters	PC1	PC2	PC3	PC4	PC5
TDS	**0.945**	0.064	−0.001	0.009	0.226
pH	0.199	−0.128	−0.282	0.018	**−0.825**
DO	0.406	**−0.666**	−0.456	−0.315	−0.159
BOD	0.106	−0.188	−0.064	0.082	**0.899**
Ca	−0.076	**0.877**	0.049	0.129	0.149
Mg	**−0.706**	0.239	0.351	0.137	0.376
K	−0.155	**0.648**	**0.563**	−0.245	−0.102
Zn	0.056	0.162	−0.007	**0.970**	0.040
Pb	0.082	0.166	−0.010	**0.973**	0.036
Cl^-	−0.247	0.180	**0.877**	0.058	0.007
F^-	−0.177	**−0.669**	0.422	−0.338	0.200
NO_3^{2-}	**−0.607**	0.105	0.086	−0.025	0.134
NO_2^-	0.215	**0.749**	0.303	0.278	−0.303
SO_4^{2-}	0.189	0.017	**0.865**	−0.025	0.183
TSS	**0.844**	0.178	0.145	0.281	0.009
Eigen value (total)	4.255	3.318	2.066	1.922	1.066
% of total variance	28.365	22.123	13.770	12.811	7.104
Cumulative % of variance	28.365	50.488	64.258	77.069	84.173

Mg, K, Zn and Pb are 25.154 ± 8.674, 840.400 ± 311.081 mg/l, 7.574 ± 0.256 pH unit, 4.544 ± 0.933, 0.758 ± 0.179, 51.494 ± 28.095, 0.771 ± 0.153, 2.211 ± 0.878, 4.692 ± 5.971, 69.545 ± 53.873, 48.458 ± 22.690, 19.676 ± 7.361, 12.874 ± 11.382, 0.027 ± 0.029, 0.096 ± 0.154 mg/l, respectively.

The results show that most of the parameters (TSS, TDS, Cl^-, SO_4^{2-}, Ca and K) express significant changes, whereas pH, DO, BOD, F^-, NO_3^2, Zn and Pb show minimum changes in all cases. Among all the parameters TSS, BOD, NO_2^-, Pb were above the standard values (DoE 1997 international; WHO 2004) (Table 2). Noticeable depletion of DO is recorded at all points, indicative of potential ecological and environmental risk. The sharp decline in DO may have resulted from the introduction of organic matter in the water which consumed oxygen during decomposition (Masamba and Mazvimavi 2008).

Table 4 Rotated component matrix of five factor model for sampling site with moderate to strong loadings in bold typeface

Sample location	PC1	PC2	PC3	PC4	PC5
S1	**−1.352**	−0.349	−0.577	0.045	0.150
S2	**−0.906**	**−0.793**	**−0.607**	−0.116	**1.547**
S3	**−2.531**	**1.184**	0.388	−0.146	**1.621**
S4	**−1.240**	**−1.151**	0.256	−0.060	**−1.009**
S5	−0.056	−0.525	0.457	−0.082	0.028
S6	0.218	−0.585	0.033	0.025	0.276
S7	−0.006	**2.437**	−0.073	−0.430	**−1.450**
S8	−0.091	−0.395	−0.463	−0.317	**−1.363**
S9	**−0.814**	−0.536	**−0.619**	−0.066	**−1.396**
S10	−0.124	−0.484	**−1.068**	−0.058	**−0.651**
S11	0.182	**0.730**	**3.385**	−0.508	0.179
S12	**1.223**	−0.147	**−0.788**	−0.198	**1.522**
S13	**1.356**	−0.311	−0.331	−0.382	**1.375**
S14	0.279	−0.055	0.012	−0.377	−0.165
S15	**0.879**	0.567	−0.305	**−0.700**	**0.692**
S16	**0.709**	**−0.648**	0.187	−0.080	−0.256
S17	**1.016**	**−1.440**	**1.214**	0.061	**−0.718**
S18	**0.878**	**1.839**	**−1.156**	**−0.631**	−0.443
S19	0.381	**0.660**	0.056	**4.022**	0.060

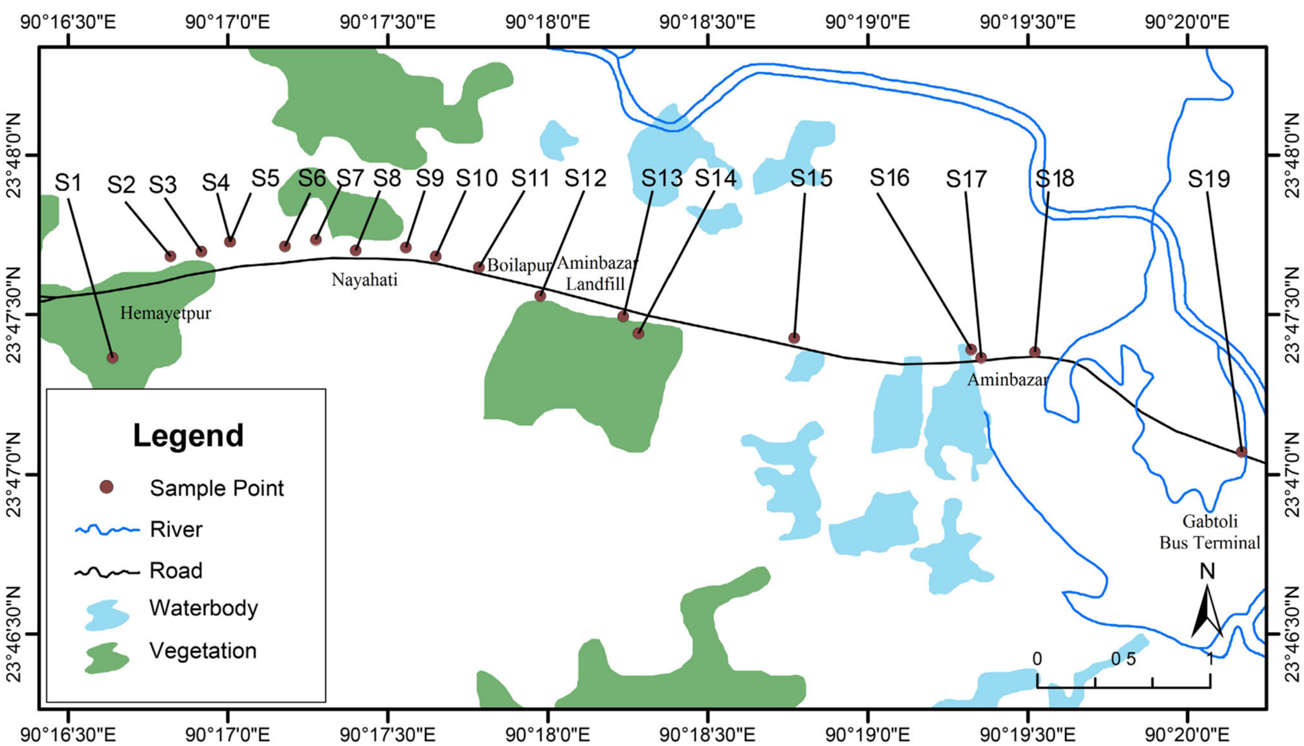

Fig. 1 Sampling location map of the study area

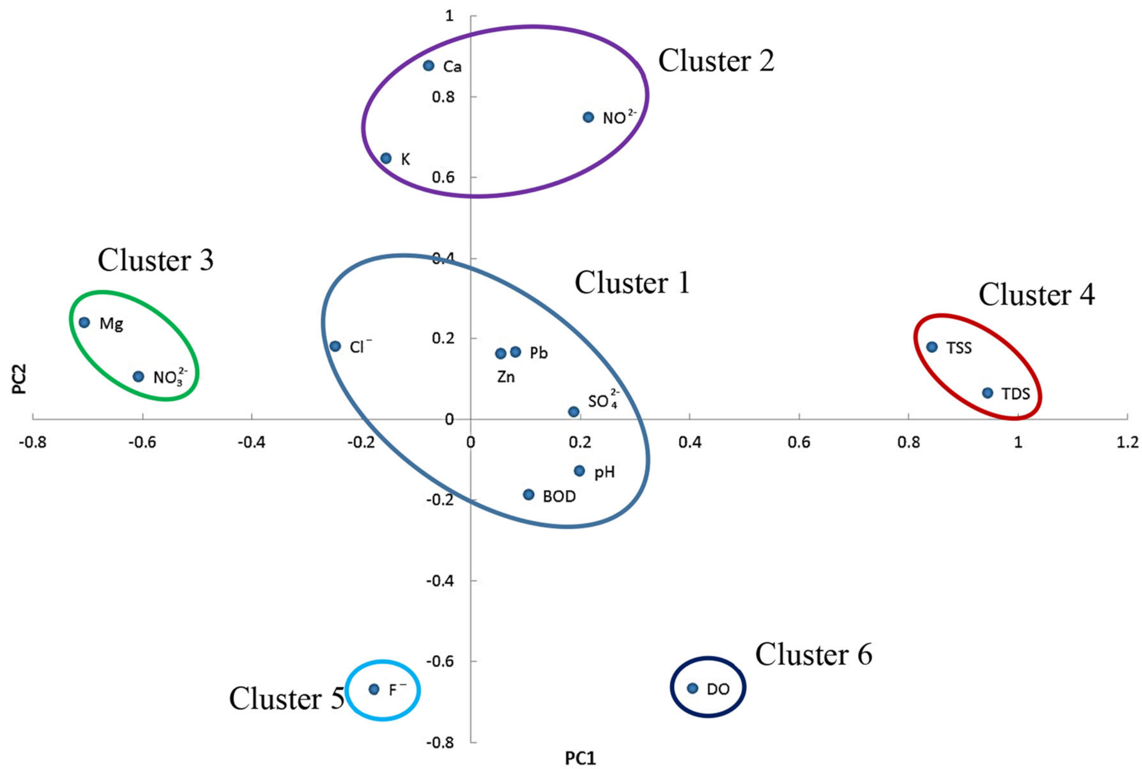

Fig. 2 Plots of PC1 vs. PC2 showing all analyzed parameters

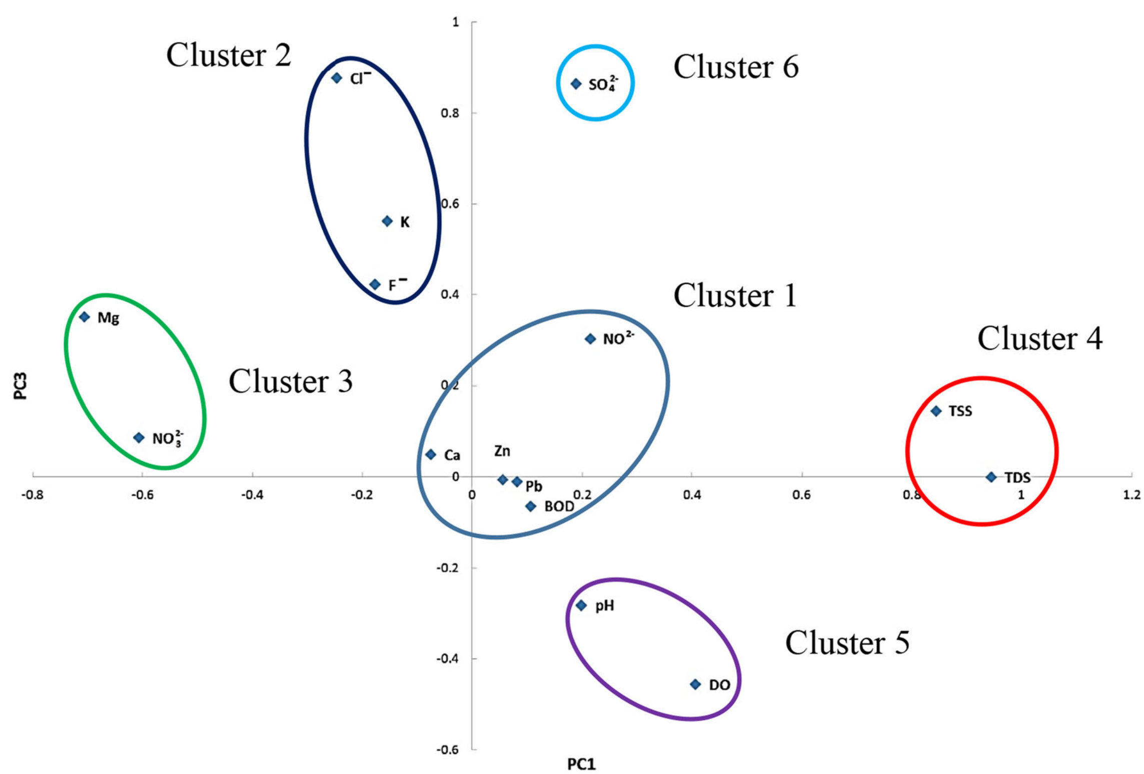

Fig. 3 Plots of PC1 vs. PC3 showing all analyzed parameters

Principal component analysis (PCA) and factor analysis (FA)

The calculated factor loadings, cumulative percent, and percentages of variance explained each factor in R-mode PCA are listed in Table 3. Varimax rotation was used to maximize the sum of variance of the factor coefficients (Bhuiyan et al. 2011). Five factors with eigenvalues >1 were extracted from the varimax-rotated factor analysis of all parameters in the dataset. The retained factors which led to a reduction of the initial dimension of the data set and explained about 84.173% of the total variance.

The terms 'strong', 'moderate', and 'weak' as applied to factor loadings, refer to absolute loading values of >0.75, 0.75–0.5 and 0.5–0.3, respectively (Liu et al. 2003).

The first principal component (PC1) in the data sets explains 28% of total variance and is heavily positively loaded on TDS and TSS which indicates the roadside and brick field dust settle down in nearby water body. Moderately negatively loaded on Mg, NO_3^{2-} indicates the nutrient uptake by aquatic organisms. These parameters retain high positive scores in S12, S13, S15, S16, S17, S18 and negative scores in S1, S2, S3, S4, S9 (Table 4).

PC2 explains 22.123% of total variance, and is heavily positively loaded on Ca, NO_2^- and moderately positively loaded on K which indicates the agricultural influence (fertilizer utilization and irrigation). Moderately negatively loaded on DO and F^- suggests utilization of dissolved oxygen to decompose the organic matter by bacteria (FC) function (Singh et al. 2004). These factors are the input from anthropogenic (industrial and agricultural) and natural sources. These parameters retain high positive scores in S3, S7, S11, S18, S19 and negative scores S2, S4, S16, S17.

PC3 describes 13.770% of total variance, and is heavily positively loaded on SO_4^{2-}, Cl^- and moderately positively loaded on K, these factors represent the contribution of nonpoint pollution from agricultural and soil erosion processes. In these areas, farmers use ammonium sulfate

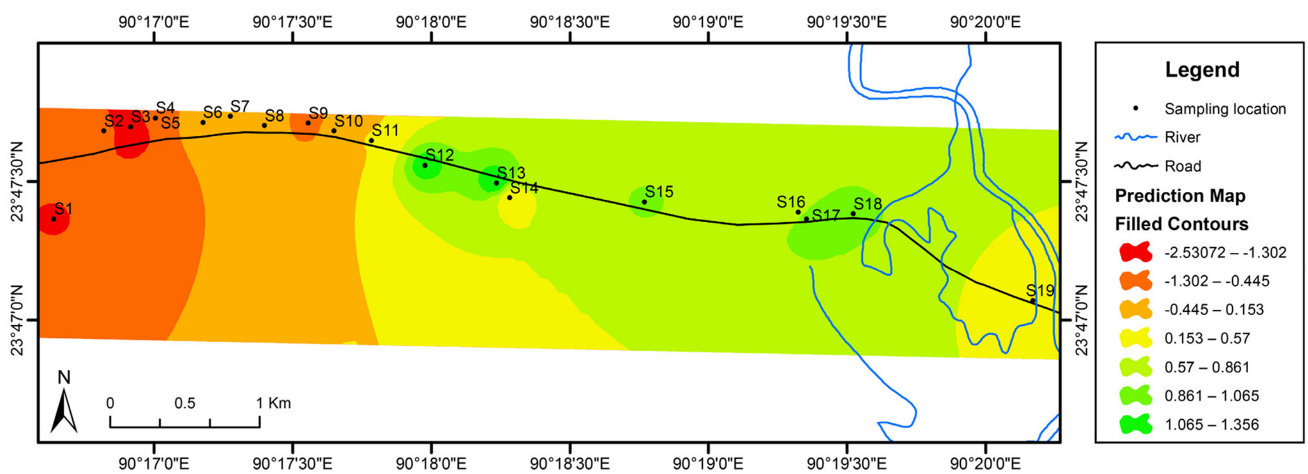

Fig. 4 Factor score map of principle component (PC1)

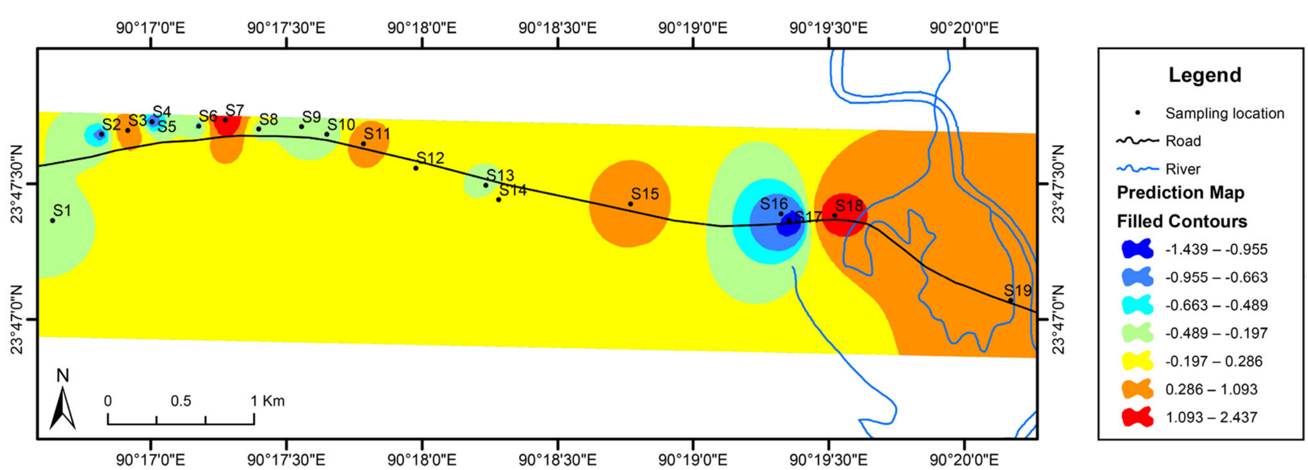

Fig. 5 Factor score map of principle component (PC2)

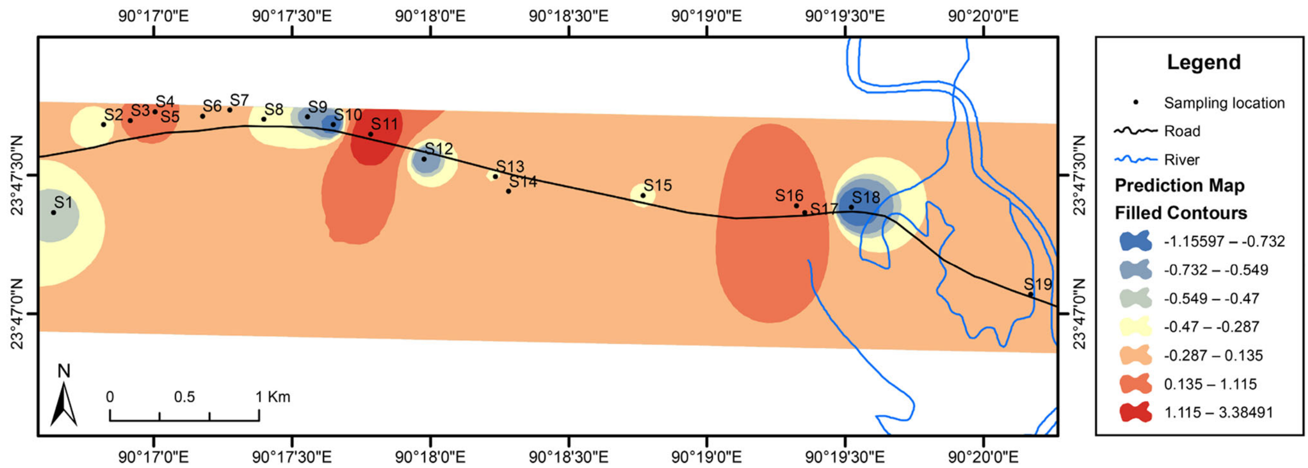

Fig. 6 Factor score map of principle component (PC3)

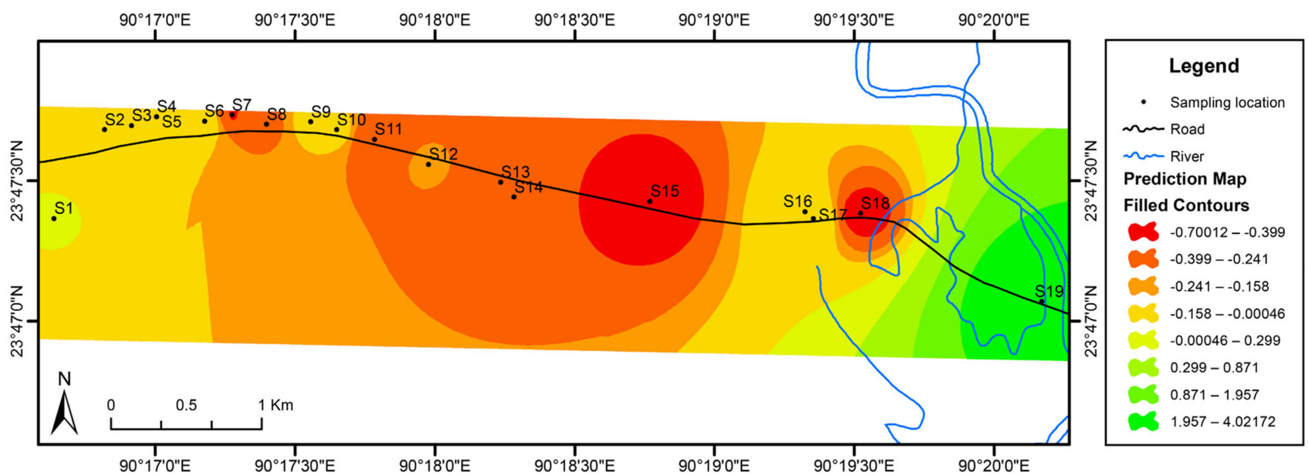

Fig. 7 Factor score map of principle component (PC4)

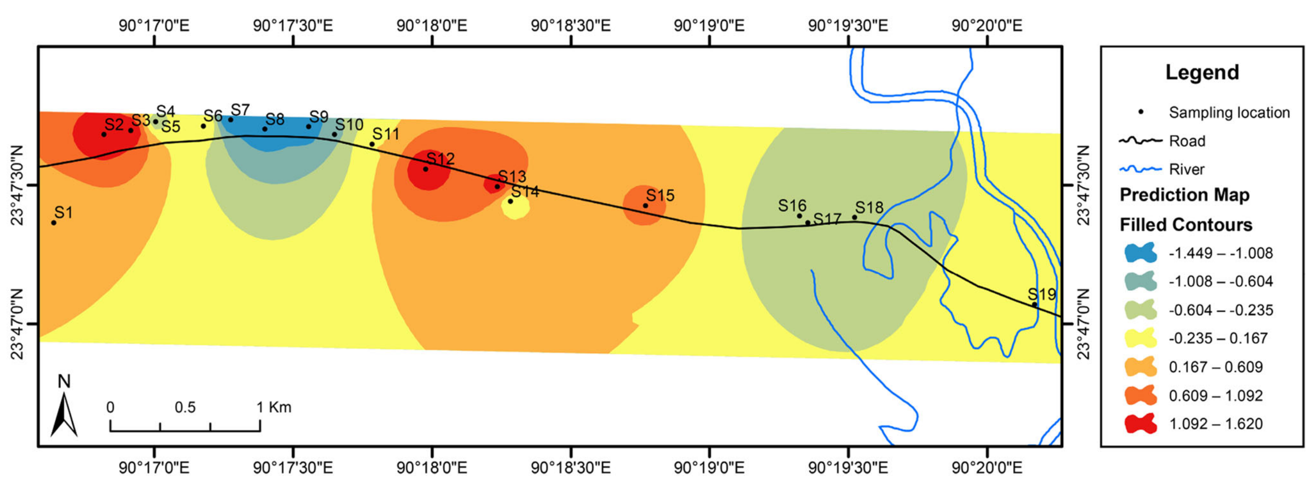

Fig. 8 Factor score map of principle component (PC5)

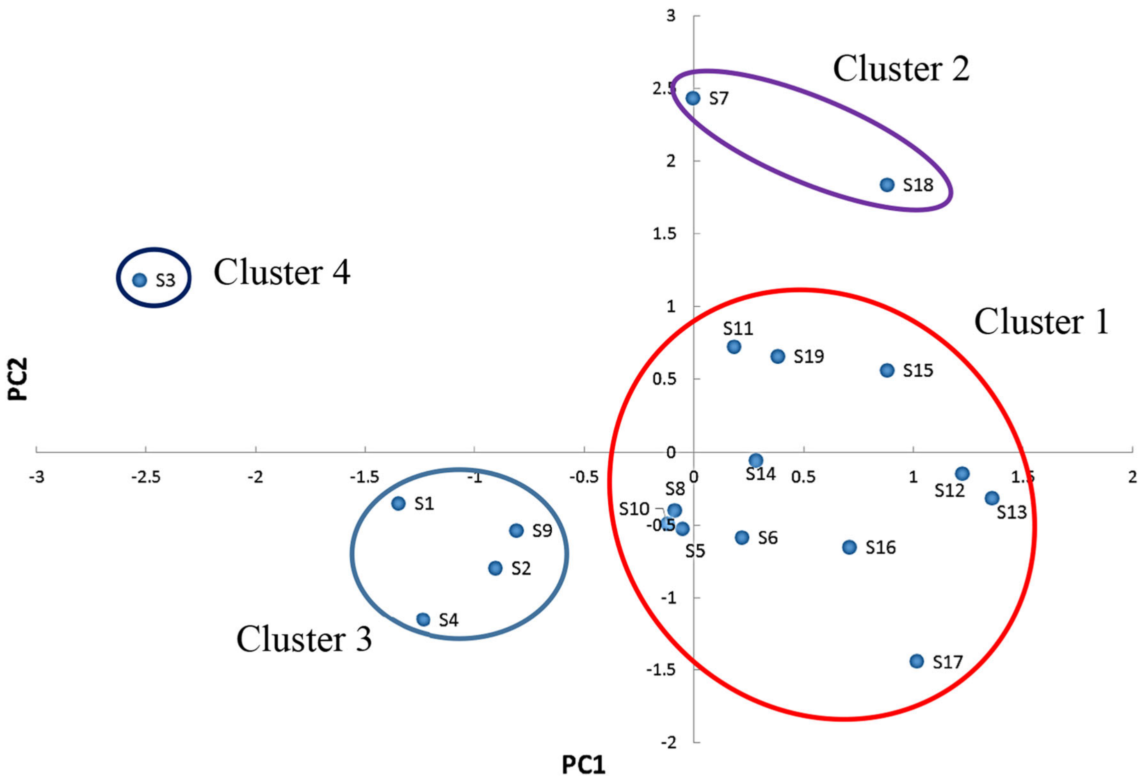

Fig. 9 Plots of PC1 vs. PC2 showing groupings among sampling sites

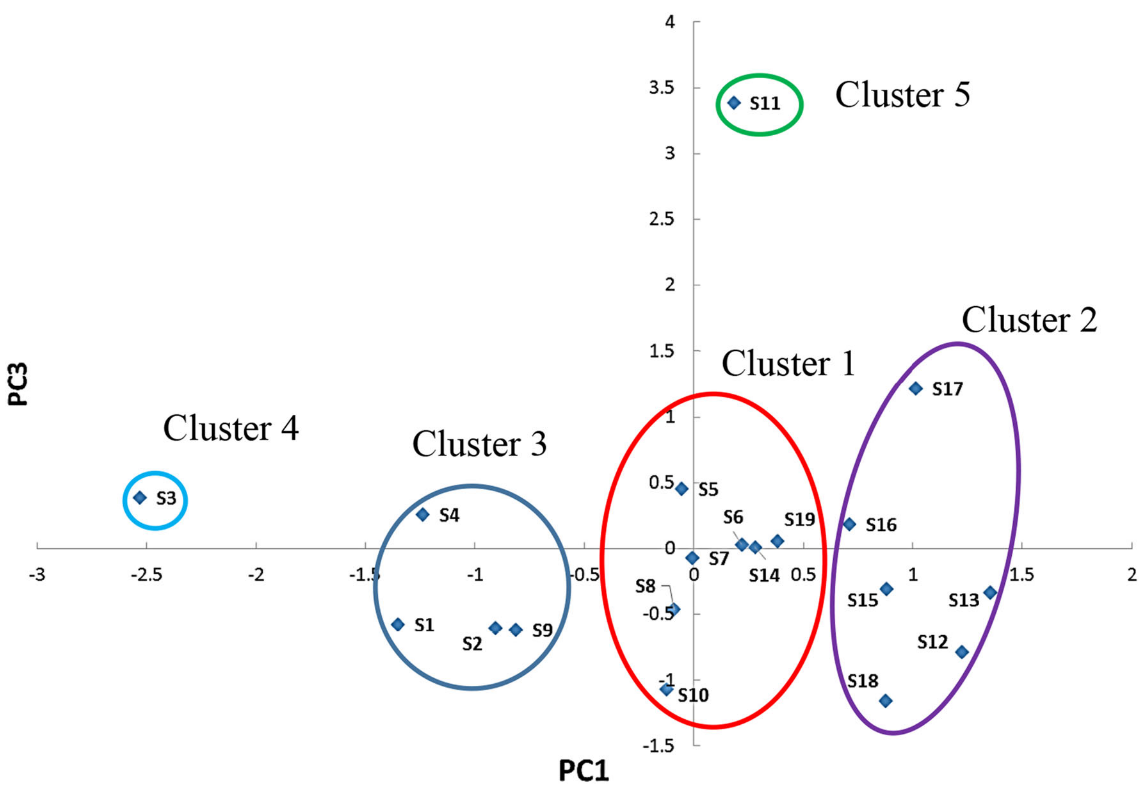

Fig. 10 Plots of PC1 vs. PC3 showing groupings among sampling sites

fertilizers, and the surface runoff receives sulfate via surface runoff and irrigation waters (Bhuiyan et al. 2011). These parameters retain high positive scores in S11, S17 and negative scores S9, S10, S12, S18.

PC4 depicts heavy positively loaded on Zn and Pb which denotes 12.811% of total variance. This represents vehicle emission and diffusion from battery stores. These parameters retain high positive scores in S19 and negative scores S15, S18.

PC5 shows only 7.104% of total variance and it depicts strong positive loading of BOD and strong negative loading of pH. These parameters retain high positive scores in S2, S3, S12, S13, S15 and negative scores S4, S7, S8, S9, S10, S17.

The relationships among the analyzed parameters are also visualized in the factor loadings plots of PC1 vs. PC2 and PC1 vs. PC3 (Figs. 2, 3).

For all parameters, six main clusters are obtained from the plotting of PC1 vs. PC2 (Fig. 2). Cluster 1 contains pH, BOD, Zn, Pb, Cl^- and SO_4^{2-}parameters and cluster 2 consists of Ca, K and NO_2^-. Cluster 3 includes Mg and NO_3^{2-}. Cluster 4 includes TSS and TDS. Whereas F^- and DO independently remain in cluster 5 and 6, respectively.

For PC1 vs. PC3 plot (Fig. 3) similarly six main clusters are obtained. Cluster 3 and 4 of both plots show similar grouping.

Spatial similarities and site grouping

GIS-based factor score maps were developed using IDW for each five factors/principal components extracted from the PCA of the data set. The factor scores for each factor and the corresponding coordinates of the sampling points were used to create interpolation surfaces. The power value was set 2, the Standard neighborhood was used instead of smooth neighborhood and sector type was 4 sector with 45° offset. These interpolation maps show spatial variations of the influence of each five dominant processes in the study area.

Figures 4, 5, 6, 7, and 8 represent factor score maps for principal component 1, 2, 3, 4 and 5, respectively.

In Fig. 4, the area has factor scores ranging from −2.530 to 1.356. Within this range of scores, about 68.16% of the study area has positive factor scores, and about 31.83% has scored in the range of −2.53072 to 0.153. This suggests

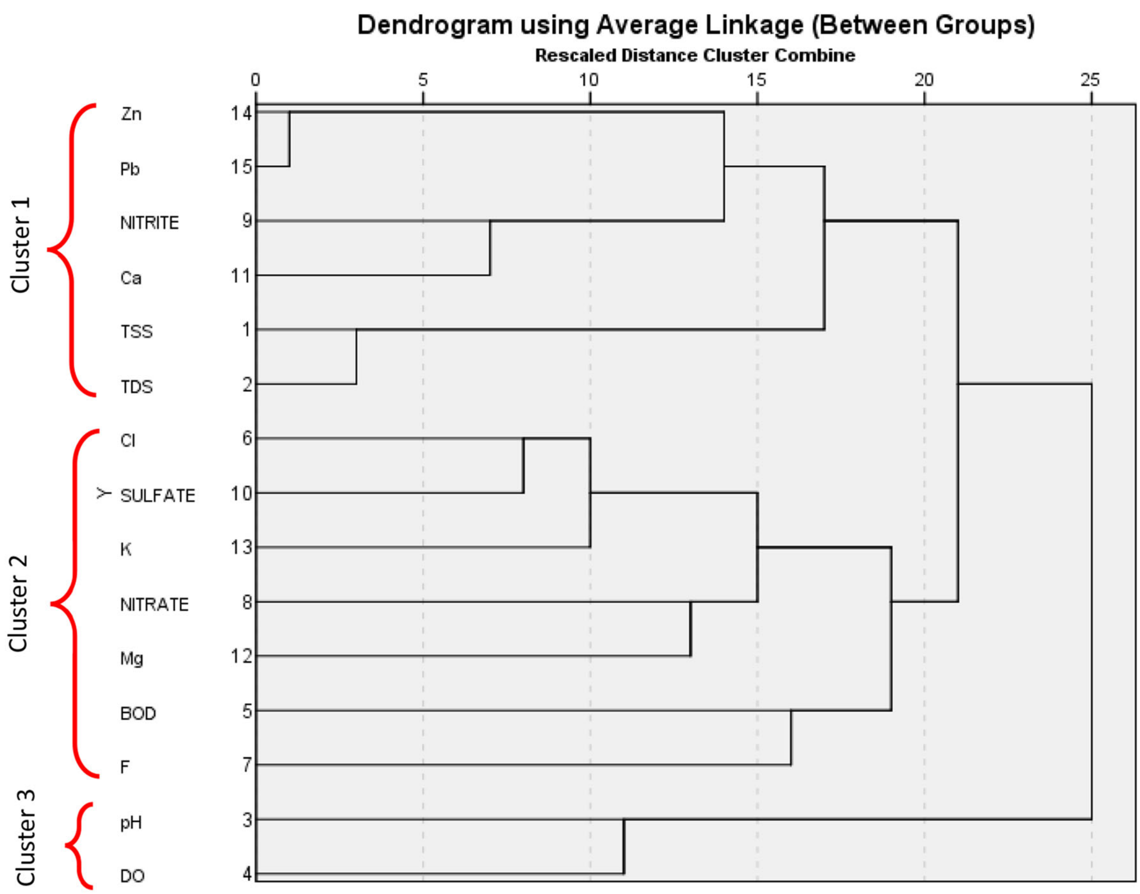

Fig. 11 Dendrogram showing the hierarchical clusters of analyzed parameters

that the processes involved in PC1. Generally, the effects of PC1 increase from northwestern to southeastern parts of the study area. The highest positive impact of PC1 occurs in S12 and S13.

In Fig. 5, the factor scores of PC2 ranging from −1.439 to 2.437. Within this range of scores, about 27.720% of the study area has positive factor scores, and about 72.279% has scored in the range of −1.439 to 0.286. The highest positive impact of PC2 occurs in S7, S18 and S19 which is on the eastern side of the study area near Gabtoli bus terminal.

The PC3 factor score map (Fig. 6) ranges from −1.155 to 3.384. Only 11.316% of the area covers the positive factor loadings. Most of the area covers the range of −0.287 to 0.135. This area is utilized for agricultural purpose mostly.

In Fig. 7, the factor scores of PC4 range from −0.700 to 4.021. Within this range of scores, about 19.685% of the study area has positive factor scores, and about 80.314% has scores in the range of −0.700 to 0.299. The highest positive impact of PC4 occurs in S19 which is in the eastern side of the study area near Gabtoli bus terminal that explains the heavy metal deposition.

The PC5 factor score map (Fig. 8) ranges from −1.449 to 1.620. About 36.973% of the area covers the positive factor loadings. The positive loading is highest in S2, S3, S12 and S13.

The factor scores from the R-mode PCA performed on the data set was also used to identify sampling site similarities/groupings (Fig. 9). On the plot of the first two principal components, PC1 vs. PC2 four main clusters are obtained. Cluster 1 contains S5, S6, S8, S10, S11, S12, S13, S14, S15, S16, S17 and S19. Cluster 2 consists of S7 and S18. Cluster 3 includes S1, S2, S4 and S9. Cluster 4 includes only S3. For PC1 vs. PC3 plot (Fig. 10) similarly five main clusters are obtained. Cluster 3 and 4 of both plots show similar grouping. Only S3 and S11 alone forms two individual clusters.

Cluster analysis

R-mode cluster analysis performed on the measured basic water quality parameters reveals three distinct groups or clusters. Cluster 1 contains Zn, Pb, Nitrite (NO_2^-), Ca, TSS and TDS. The interrelated association among Zn–Pb, TSS–

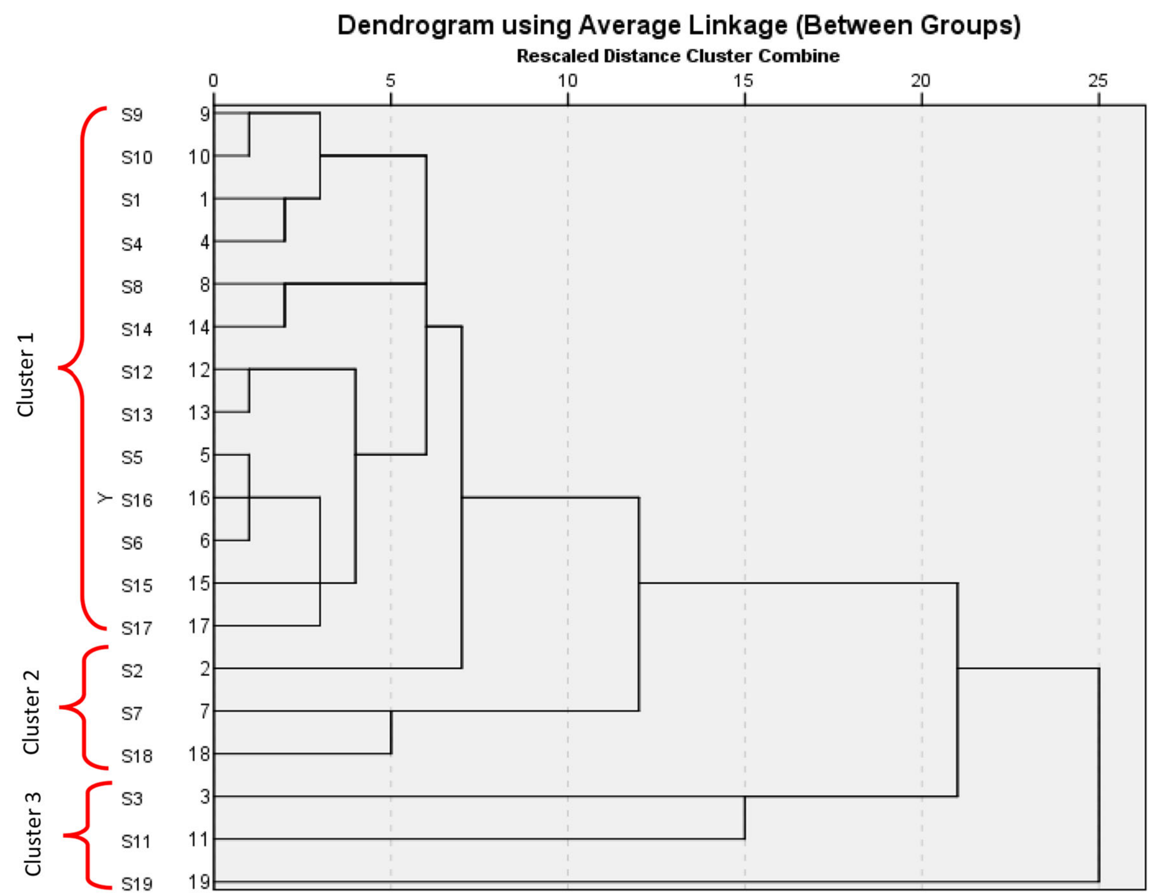

Fig. 12 Dendrogram showing the hierarchical clusters of sampling site

Table 5 Pearson correlation matrix (CM) of water quality parameters

	TSS	TDS	pH	DO	BOD	Cl^-	F^-	NO_3^{2-}	NO_2^-	SO_4^{2-}	Ca	Mg	K	Zn	Pb
TSS	1														
TDS	**0.873****	1													
pH	0.092	0.038	1												
DO	0.069	0.306	0.465*	1											
BOD	0.041	0.294	**−0.559***	0.054	1										
Cl^-	−0.007	−0.177	−0.280	**−0.651****	−0.067	1									
F^-	−0.308	−0.138	−0.207	0.277	0.223	0.259	1								
NO_3^{2-}	−0.222	−0.441	−0.247	−0.371	0.022	0.238	0.090	1							
NO_2^-	0.378	0.155	0.094	**−0.577****	−0.351	0.379	**−0.523***	−0.200	1						
SO_4^{2-}	0.199	0.153	−0.386	−0.345	0.104	**0.605****	0.291	−0.117	0.211	1					
Ca	0.061	0.029	−0.202	**−0.674****	−0.050	0.210	−0.444	0.003	**0.639****	0.090	1				
Mg	−0.481*	**−0.550***	**−0.508***	**−0.654****	0.223	**0.583****	0.152	0.391	0.087	0.223	0.418	1			
K	0.053	−0.120	−0.145	**−0.630****	−0.230	**0.578****	−0.108	0.405	**0.535***	0.459*	0.499*	0.317	1		
Zn	0.325	0.072	−0.020	−0.395	0.099	0.027	−0.409	−0.033	0.388	0.011	0.288	0.105	−0.117	1	
Pb	0.395	0.113	−0.005	−0.368	0.085	0.039	−0.441	0.001	0.374	−0.005	0.275	0.123	−0.115	**0.981****	1

* Correlation is significant at the 0.05 level (two-tailed)

** Correlation is significant at the 0.01 level (two-tailed)

TDS, and NO_2^--Ca shows similar positive loadings in PC4, PC1 and PC2, respectively (Fig. 11).

Cluster 2 includes Cl, sulfate (SO_4^{2-}), K, Nitrate (NO_3^{2-}), Mg, BOD and F. The interrelated association among Cl^-, Sulfate, K shows similar positive loadings in PC3 and NO_3^{2-}-Mg shows negative loadings in PC1.

In cluster 3 both pH and DO are slightly different from the other cluster members, as depicted by its long linkage distance.

To evaluate the spatial similarities and site groupings among the sampling sites R-mode cluster analysis is performed. Where sampling sites belonging to a particular cluster exhibit similar characteristics with respect to the analyzed parameters (Bhuiyan et al. 2011). Three major clusters were extracted from all analyzed parameters for the 19 sampling sites (Fig. 12).

In cluster 1 the similarities among sampling site S9–S10, S8–S14, S12–S13, and S5–S6–S16 is also observed in the factor score maps of PC2, PC4, PC1 and PC4, respectively.

Cluster 2 represents the similarities among sampling site S7–S18 which is noticed in factor score map of PC2.

The same observation is found in Cluster 3 which represents the similarities among sampling site S3–S11 in factor score map of PC.

Correlation matrix (CM)

Pearson's correlation matrix reveals some new associations between the parameters (Table 5) that are not adequately reported in the PCA. TSS has a strong positive relationship with TDS ($r = 0.873$; $P < 0.01$) indicates the roadside and brick field dust settle down in nearby water body. A similar result is also observed in PC1. An alike association is found between Pb–Zn ($r = 0.981$; $P < 0.01$) which is also observed in PC4. A positive correlation is also found among Cl^--SO_4^{2-}, Mg-Cl^-, K-Cl^-, Ca-NO_2^- and K-NO_2^-. Ca is negatively correlated with DO ($r = -0.674$; $P < 0.01$) which is described in PC2. A negative correlation is also found among Mg-TDS, pH-BOD, pH-Mg, DO-Cl^-, DO-NO_2^-, DO-Mg, DO-K and F-NO_2^-.

Conclusion

This study was undertaken to evaluate the roadside surface water quality of Hemayetpur region. Descriptive statistical analysis showed TSS, BOD, NO_2^-, Pb was above the standard values. From PCA, five major principal components were extracted which perfectly reduced the data dimension and indicated possible anthropogenic sources. These components explain 84.173% of total variance. From factor score map high positive loading is found near Hemayetpur (PC5), near Nayahati (PC2), Boilapur area (PC3, PC5), Aminbazar Landfill site (PC1), Aminbazar (PC1, PC2) and Gabtoli Bus Terminal (PC4). Cluster analysis formed three major clusters for both water parameters and sampling sites. Anthropogenic sources are the main reason for water quality deterioration. This result regarding sources showed similarities among PCA and CA. From Pearson's correlation matrix significant positive relation was found between TDS–TSS and Zn–Pb that explained the anthropogenic activities (vehicle emission, atmospheric deposition from the brick field and industrial pollution) in that area.

Acknowledgements Authors are grateful to IFST, BCSIR laboratory and Department of Environmental Sciences, Jahangirnagar University, Bangladesh for instrumental support and other facility.

References

Abdul-Wahab SA, Bakheit CS, Al-Alawi SM (2005) Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. Environ Model Softw 20(10):1263–1271

Ahmed F, Fakhruddin ANM, Imam TMD, Khan N, Khan TA, Rahman MM, Abdullah ATM (2016) Spatial distribution and source identification of heavy metal pollution in roadside surface soil: a study of Dhaka Aricha highway, Bangladesh. Ecol Process 5:2

Aktaruzzaman M, Fakhruddin ANM, Chowdhury MAZ, Fardous Z, Alam MK (2013) Accumulation of heavy metals in soil and their transfer to leafy vegetables in the region of Dhaka Aricha Highway, Savar, Bangladesh. Pak J Biol Sci 16:332–338

Arain MB, Kazi TG, Jamali MK, Jalbani N, Afridi HI, Shah A (2008) Total dissolved and bioavailable elements in water and sediment samples and their accumulation in Oreochromis mossambicus of polluted Manchar Lake. Chemosphere 70:1845–1856

Astel A, Glosiska G, Sobczyski T, Boszke L, Simeonov V, Siepak J (2006) Chemometrics in assessment of sustainable development rule implementation. CEJCh 4(3):543–564

Astel A, Tsakovski S, Simeonov V, Reisenhofer E, Piselli S, Barbieri P (2008) Multivariate classification and modeling in surface water pollution estimation. Anal Bioanal Chem 390(5):1283–1292

Bhuiyan MAH, Rakib MA, Dampare SB, Ganyaglo S, Suzuki S (2011) Surface water quality assessment in the central part of Bangladesh using multivariate analysis. KSCE J Civ Eng 15(6):995–1003

Boyacıoğlu H, Boyacıoğlu H (2006) Water pollution sources assessment by multivariate statistical methods in the Tahtali Basin, Turkey. Environ Geol 54:275–282

Brumelis G, Lapina L, Nikodemus O, Tabors G (2000) Use of an artificial model of monitoring data to aid interpretation of principal component analysis. Environ Model Softw 15(8):755–763

Crosa G, Froebrich J, Nikolayenko V, Stefani F, Gallid P, Calamari D (2006) Spatial and seasonal variations in the water quality of the Amu Darya River (Central Asia). Water Res 40:2237–2245

DoE (1997) Industrial effluents quality standard for Bangladesh. Bangladesh Gazette Additional. http://extwprlegs1.fao.org/docs/pdf/bgd19918.pdf. Accessed 14 Dec 2014

Gantidis N, Pervolarakis M, Fytianos K (2007) Assessment of the quality characteristics of two lakes (Koronia and Volvi) of N, Greece. Environ Monit Assess 125:175–181

Helena BA, Vega M, Barrado E, Pardo R, Fernandez L (1999) A case of hydrochemical characterization of an alluvial aquifer influenced by human activities. Water Air Soil Pollut 112:365–387

Helena B, Pardo R, Vega M, Barrado E, Fernandez JM, Fernandez L (2000) Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis. Water Res 34:807–816

Isaak EH, Srivastava RM (1989) Applied geostatistics. Oxford University Press, New York, p 561

Johnston MW, Williams JS (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, pp 11

Kannel PR, Lee S, Kanel SR, Khan SP (2007) Chemometric application in classification and assessment of monitoring locations of an urban river system. Anal Chim Acta 582:390–399

Kottı ME, Vlessıdıs GA, Thanasoulıas NC, Evmırıdıs NP (2005) Assessment of river water quality in Northwestern Greece. Water Resour Manag 19:77–94

Kowalkowski T, Zbytniewski R, Szpejna J, Buszewski B (2006) Application of chemometrics in river water classification. Water Res 40(1):744–752

Kundzewicz ZW, Mata LJ, Arnell NW, Döll P, Kabat P, Jiménez B, Miller KA, Oki T, Sen Z, Shiklomanov IA (2007) Freshwater resources and their management. In: Parry ML, Canziani OF, Palutikof JP, van der Linden PJ, Hanson CE (eds) Climate change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, pp 173–210

Liu CW, Lin KH, Kuo YM (2003) Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. Sci Total Environ 313:77–89

Love D, Hallbauer D, Amos A, Hranova R (2004) Factor analysis as a tool in groundwater quality management: two southern African case studies. Phys Chem Earth 29:1135–1143

Lu RS, Lo SL (2002) Diagnosing reservoir water quality using self-organizing maps and fuzzy theory. Water Res 36(9):2265–2274

Maitra MK, Akhter SH (2011) Neotectonics in Madhupur tract and its surroundings floodplains. Dhaka Univ J Earth Environ Sci 1(2):83–89

Marengo E, Gennaro MC, Giacosa D, Abrigo C, Saini G, Avignone MT (1995) How chemometrics can helpfully assist in evaluating environmental data Lagoon water. Anal Chim Acta 317(1–3):53–63

Masamba WRL, Mazvimavi D (2008) Impact on water quality of land uses along Thamalakane-Boteti River: an outlet of the Okavango Delta. Phys Chem Earth 33:687–694

McKenna JE Jr (2003) An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. Environ Model Softw 18(3):205–220

Mendiguchía C, Moreno C, Galindo-Riaòo DM, García-Vargas M (2004) Using chemometric tools to assess anthropogenic effects in river water. A case study: Guadalquivir (Spain). Anal Chim Acta 515(1):143–149

Otto M (1998) Multivariate methods. In: Kellner R, Mermet JM, Otto M, Widmer HM (eds) Analytical Chemistry. WileyeVCH, Weinheim

Ouyang Y (2005) Evaluation of river water quality monitoring stations by principal component analysis. Water Res 39:2621–2635

Ouyang Y, Nkedi-Kizza P, Wu QT, Shinde D, Huang CH (2006) Assessment of seasonal variations in surface water quality. Water Res 40:3800–3810

Panda UC, Sundaray SK, Rath P, Nayak BB, Bhatta D (2006) Application of factor and cluster analysis for characterization of

river and estuarine water systems a case study: Mahanadi River (India). J Hydrol 331:434–445
Pekey H, Karakas D, Bakoglu M (2004) Source apportionment of trace metals in surface waters of a polluted stream using multivariate statistical analyses. Mar Pollut Bull 49:809–818
Ravichandran S (2003) Hydrological influences on the water quality trends in Tamiraparani Basin, South India. Environ Monit Assess 87:293–309
Sarbu C, Pop HF (2005) Principal component analysis versus fuzzy principal component analysis. A case study, the quality of Danube water (1985–1996). Talanta 65:1215–1220
Sarkar SK, Saha M, Takada H, Bhattacharya A, Mishra P, Bhattacharya B (2007) Water quality management in the lower stretch of the river Ganges, east coast of India: an approach through environmental education. J Clean Prod 15:1559–1567
Shrestha S, Kazama F (2007) Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. Environ Model Softw 22(4):464–475
Shuxia Y, Shang J, Zhao J, Guo H (2003) Factor analysis and dynamics of water quality of the Songhua River, Northeast China. Water Air Soil Pollut 144(1):159–169
Simeonov V, Stratis JA, Samara C, Zachariadis G, Voutsa D, Anthemidis A (2003) Assessment of the surface water quality in Northern Greece. Water Res 37(17):4119–4124
Simeonova P, Simeonov V (2007) Chemometrics to evaluate the quality of water sources for human consumption. Microchim Acta 156(3–4):315–320
Singh KP, 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 Res 38:3980–3992
Sparks T (2000) Statistics in ecotoxicology. Wiley, Chichester
Stefanov S, Simeonov V, Tsakovski S (1999) Chemometrical analysis of waste water monitoring data from Yantra river basin, Bulgaria. Toxicol Environ Chem 70:473–482
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 Res 32(12):3581–3592
Webster R, Oliver M (2001) Geostatistics for environmental scientists. Wiley, Chichester, p 271
WHO (2004) Guidelines for drinking-water quality, 3rd edn. WHO, Geneva
Wunderlin DA, Diaz MP, Ame MV, Pesce SF, Hued AC, Bistoni MA (2001) Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia river basin (Cordoba, Argentina). Water Res 35(12):2881–2894
Zhou F, Huang GH, Guo HC, Zhang W, Hao ZJ (2007) Spatio-temporal patterns and source apportionment of coastal water pollution in eastern Hong Kong. Water Res 41:3429–3439

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.