Research Journal of Chemical Sciences ISSN 2231-606X
Vol. **3(5),** 32-42, May **(2013)** *Res. J. Chem. Sci.*

Characterization and Classification of Hydrochemistry using Multivariate Graphical and Hydrostatistical Techniques

Manoj K., Ghosh S. and Padhy P.K.*
Department of Environmental Studies, Institute of Science, Visva-Bharati, Santiniketan, 731235, Birbhum, West Bengal, INDIA

Available online at: www.isca.in
Received 21st March 2013, revised 31st March 2013, accepted 20th April 2013

Abstract

Multivariate data analyses systems provide simultaneous inspection of several variables both in space and time. This study was undertaken to present the utility of multivariate data analyses models in characterization and classification of surface water chemistry. A case study was developed, where hydrochemistry of some water bodies (ponds) in the Santiniketan-Bolpur-Sriniketan zone, Birbhum district, West Bengal, India, was investigated followed by the assessment of impact of the well known annual fair Pous Mela on the chemistry of surface waters present near the vicinity of the fair ground. Santiniketan has historical importance in India and the world, made famous by versatile Nobel Laureate Gurudev Rabindra Nath Tagore. The Pous Mela, which was started by Maharshi Debendranath Tagore in 1894, is an annual fair organized on the campus of Visva-Bharati University in the month of December and is visited by thousands of people every year. Water samples collected, both spatially and temporally, were analyzed for sixteen parameters and multivariate Piper trilinear diagram was constructed for examination of their chemistry. Two multivariate data analyses techniques viz. agglomerative hierarchical cluster analysis (AHCA) and discriminant analysis (DA) were further applied for intelligent interpretation of water quality data matrix. Piper diagram classified all water samples into 'Mixed Ca^{2+}—Na^{+}—HCO_3^{-} Type'. No change in water type was recorded temporally suggesting the ionic stability of water bodies with respect to Na^+, K^+, Ca^{2+}, Mg^{2+}, HCO_3^-, Cl^- and SO_4^{2-} and further indicating the impact of the Pous Mela on major ion chemistry of nearby waters as negligible. However, hydrochemical data analyses revealed organic wastes along with PO_4^{3-} as the most serious pollutants. AHCA grouped sampling sites into three groups viz. having less anthropogenic influence, medium anthropogenic influence and high anthropogenic influence. Application of DA on data matrix uncovered four predictor variables namely BOD, pH, Cl^- and total hardness as the most relevant discriminating parameters between three groups. This study showed that the multivariate models are highly effective in elucidating and illustrating the hydrochemical profile and framework assessment which can be applied on large scale for better interpretation of water chemistry and management of water resources of any region.

Keywords: Cluster analysis, discriminant analysis, piper trilinear diagram, Pous Mela, Santiniketan

Introduction

Water is central to the survival of human beings as it is needed in all aspects of life[1-3]. In addition to sources of available fresh surface water like rivers and lakes, pond water is also utilized for human consumption. In West Bengal, India, ponds are widely used by the local communities for domestic and agricultural activities. Characterization of water with respect to physical, chemical and biological parameters is essential to evaluate its ecological status[4,5]. For example, enhanced nitrate and phosphate levels can give rise to eutrophication characterized by algal and cyanobacterial blooms of many forms. The eutrophic conditions can severely impact other aquatic life forms including vertebrates like fishes. Alteration in pH and presence of ammonia can drastically affect the aquatic biota. Moreover, the polluted water becomes unfit for domestic, agricultural and industrial purposes. The assessment of water quality is not simply about laboratory analyses of hydrochemical parameters and comparing with known standards to draw some conclusions. Water chemistry is a multivariate data involving many variables and some sophisticated statistical modeling systems are required for their comprehensive evaluation. Many hidden phenomena and inherent complex water chemistry can be expressed through these modeling methods without loss of original information. Piper trilinear diagram, cluster analysis (CA) and discriminant analysis (DA) are such multivariate data analyses models that can be employed for the surface water characterization and classification. Piper diagram reveals hydrochemical regime with respect to the presence of ions viz. Na^+, K^+, Ca^{2+}, Mg^{2+}, HCO_3^-, CO_3^{2-}, Cl^- and SO_4^{2-}. Multivariate data analyses involve sequential application of statistical techniques. For example, CA uses unclassified data to reveal groups of observations, while DA employs data matrix that is pre-classified into groups[6]. The basic aim of CA is classifying variables like sampling sites or environmental quality monitoring parameters into mutually exclusive groups based on their similarity or dissimilarity trend. It distinguishes members of one group from the members of other groups and illustrates results in a graphical form called dendrogram which makes the data interpretation easy and understandable. This analysis of within-group and between-groups relationships is important for the sustainable management of precious natural resources. DA employs a set of methods to differentiate among groups in data and to assign new observations into the existing groups[7]. It identifies most significant parameters responsible for differentiating naturally occurring groups or clusters (minimum two) from a large dataset, and thus, brings about significant dimensionality or data reduction.

This study was carried out to demonstrate the utility of multivariate data analyses statistical techniques in characterization and classification of surface water chemistry. A case study was developed, where the hydrochemistry of some water bodies (ponds) in Santiniketan-Bolpur-Sriniketan zone, Birbhum district, West Bengal, India, was investigated followed by the assessment of impact of the well known annual fair Pous Mela on the chemistry of surface waters present near the vicinity of the fair ground. The Pous Mela, which was started by Maharshi Debendranath Tagore in 1894, is an annual three day fair on the campus of Visva-Bharati University that begins in the fourth week of December and is organized jointly by the University and the Santiniketan Tust. Many people, including visitors and persons running stalls use ponds located close to the Pous Mela ground in Bhubandanga for bathing, cleaning utensils, washing clothes and openly defecate on boundaries surrounding these ponds. Though, the local people use ponds' water for the above mentioned activities throughout the year, during the time of Pous Mela the pressure on water bodies escalates which provides a suitable opportunity to study the impact of the fair on the ecology of water bodies especially with respect to variation in water chemistry. Multivariate models like Piper trilinear graphical technique and hydrostatistical approaches viz. agglomerative hierarchical cluster analysis (AHCA) and DA were applied for characterization, classification and intelligent elucidation of water chemistry of the studied surface water bodies.

Material and Methods

Synopsis of study area, water bodies and the annual fair Pous Mela: The study area, located between 23°40′07.97″N—87°39′58.53″E and 23°40′31.26″N—87°41′24.01″E, lies in Bolpur-Sriniketan block of the Birbhum district of the Indian state of West Bengal. The block carries high significance because of the presence of historical place Santiniketan and the Visva-Bharati University. Presence of Ballavapur Wildlife Sanctuary and the seasonal Kopai river are some other important features of the block. The area is covered with lateritic red soil with sandy texture. Five surface water bodies (ponds) lying in the Santiniketan-Bolpur-Sriniketan zone were selected for hydrochemical examination. Four ponds (Pond 2-5) are located to the right side of Santiniketan-Bolpur road in Bhubandanga and of them pond 5 is most extensively used by the local population. Pond 2 is located beside Nichupatty Bangalow Boys' hostel, pond 3 lies parallel to the roadside shops and Pond 4 is used for aquaculture. Pond 1 is situated to the left side of Sriniketan-Bolpur road in Kalisayer close to the Sriniketan market. The annual fair Pous Mela is organized in the fourth week of December jointly by the Visva-Bharati University and the Santiniketan Tust on the University campus. The fair ground is located nearby Bhubandanga market and Ponds 2, 3, 4 and 5. In 2012, though officially it was a four day event (adding one day extension), from 23.12.2012 to 26.12.2012, unofficially is continued up to the first week of January, 2013. Moreover, number of stalls was also increased to 1,200 in the year 2012 as compared to around 1000 stalls earlier. Figure-1 shows the study area and location of selected water bodies[8,9].

Sampling and analyses: Fifteen spots were selected, three spots for each pond, for the collection of water samples between 10:30 AM to 11:30 AM during third week of December, 2012 (prior to the Pous Mela), during the Pous Mela period and second week of January, 2013 (after the Pous Mela). A summary of the sampling sites is given in table-1. Collected water was immediately stored in acid cleaned and properly rinsed high density polythene (HDPE) bottles in order to minimize container pollution and better sample preservation[10]. The samples were stored in refrigerators at 4°C prior to analyses, and all analyses, except BOD, were finished within three days of samples' collection. Water quality of the collected samples was assessed to measure presence of some major cations like potassium (K^+), sodium (Na^+), magnesium (Mg^{2+}) and calcium (Ca^{2+}); some major anions like bicarbonate (HCO_3^-), chloride (Cl^-), nitrate (NO_3^-), phosphate (PO_4^{3-}) and sulphate (SO_4^{2-}); and other important parameters like temperature (Temp), pH, electrical conductivity (EC), total dissolved solids (TDS), total hardness (TH), dissolved oxygen (DO) and biochemical oxygen demand (BOD) following standard procedures[11]. Concentrations of Mg^{2+} were calculated from the values of TH and calcium hardness[11]. Temp, pH and DO were measured at the sampling sites. All chemicals were of analytical grade purchased from the Merck, India. Millipore water was used for preparation of all reagents and standards.

Piper trilinear diagram: Graphical methods, like Stiff diagram, Durov plot, Piper diagram etc. provide visible interpretation of water chemistry. In our study basic chemistry of surface water samples with respect to presence of some key cations like Na^+, K^+, Ca^{2+} and Mg^{2+} and some chief anions like HCO_3^-, CO_3^{2-}, Cl^- and SO_4^{2-} was determined constructing Piper trilinear diagrams. A Piper diagram consists of geometrical combination two outer triangles and a middle or inner diamond-shaped quadrilateral. Here, relative abundance of cations, in % meq/l, is plotted on left cations' triangle, while relative concentration of anions, also in % meq/l, is simultaneously plotted on right anions' triangle and the resultant points from the cation and anion triangular plots are placed over the inner diamond-like quadrilateral structure. Water types are designated according to the zones in which these points fall on the middle quadrilateral plot. Each point represents one water type.

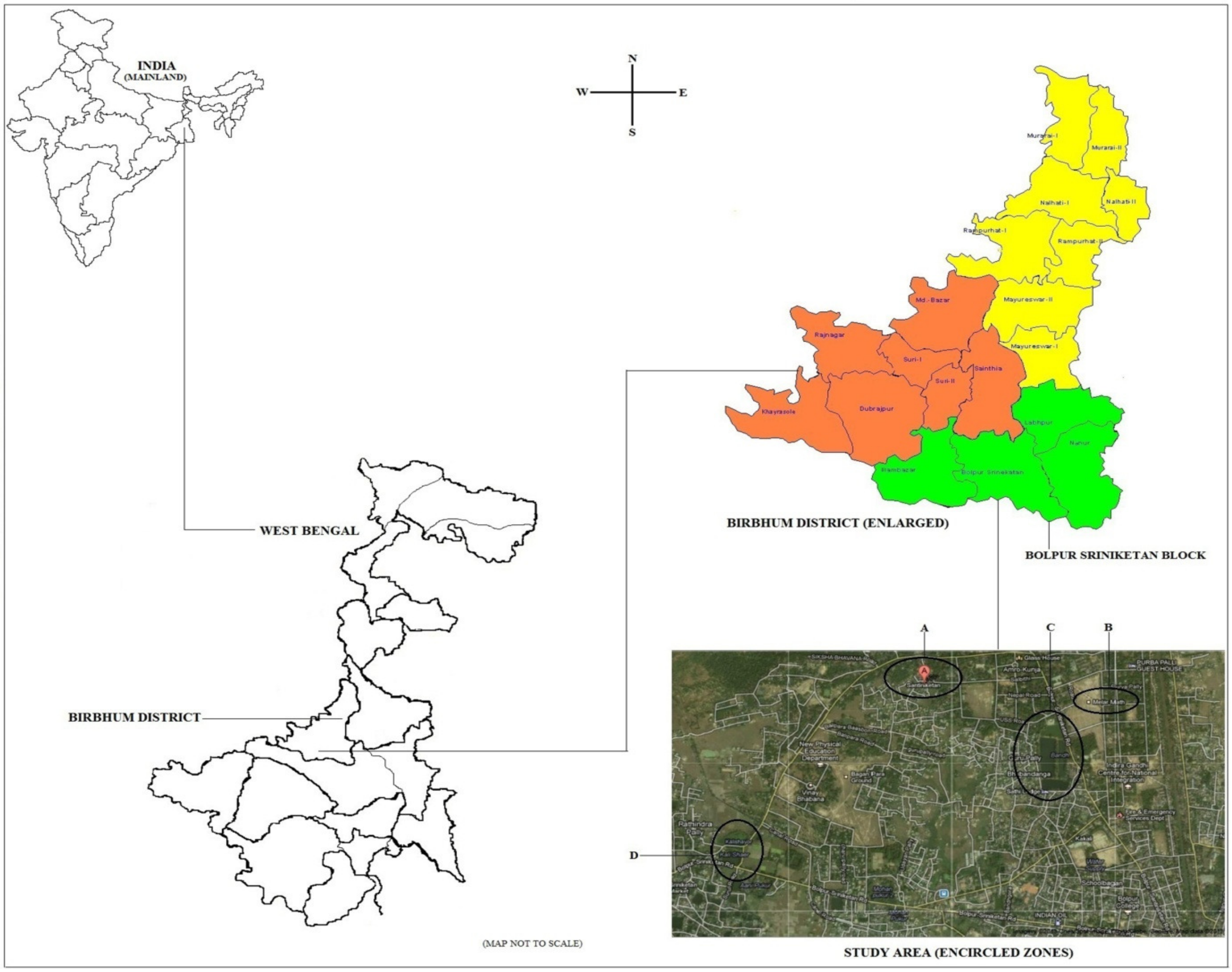

Figure-1
Study area and sampling sites; (Some parts of the diagram taken from other sources[8, 9]; A = Santiniketan, B = Pous Mela ground, C = Bhubandanga, D = Kalisayer)

Table-1
Sampling sites and their co-ordinates

Water body	Place	Code 1	Code 2	Code 3	Co-ordinate
Pond 1	Kalisayer (near Sriniketan – Bolpur road)	P1S1	1	S1	23°40′075″ N; 87°40′135″ E
		P1S2	2	S2	23°40′099″ N; 87°40′204″ E
		P1S3	3	S3	23°40′067″ N; 87°40′216″ E
Pond 2	Bhubandanga (near Santiniketan – Bolpur road)	P2S1	4	S4	23°40′562″ N; 87°41′465″ E
		P2S2	5	S5	23°40′590″ N; 87°41′495″ E
		P2S3	6	S6	23°40′570″ N; 87°41′442″ E
Pond 3	Bhubandanga (near Santiniketan – Bolpur road)	P3S1	7	S7	23°40′593″ N; 87°41′503″ E
		P3S2	8	S8	23°40′481″ N; 87°41′508″ E
		P3S3	9	S9	23°40′391″ N; 87°41′507″ E
Pond 4	Bhubandanga (near Santiniketan – Bolpur road)	P4S1	10	S10	23°40′342″ N; 87°41′466″ E
		P4S2	11	S11	23°40′332″ N; 87°41′434″ E
		P4S3	12	S12	23°40′372″ N; 87°41′515″ E
Pond 5	Bhubandanga (near Santiniketan – Bolpur road)	P5S1	13	S13	23°40′439″ N; 87°41′422″ E
		P5S2	14	S14	23°40′395″ N; 87°41′423″ E
		P5S3	15	S15	23°40′593″ N; 87°41′489″ E

(P = pond, S = sampling site)

Data exploration and checking: Prior to performing statistical analyses the dataset was examined for its normal distribution characteristics. Statistical plots like P-P plot, Q-Q plot and Box and Whisker plot followed by skewness and mean-median comparison were used for checking normal distribution behaviour of the studied parameters. Presence of outliers was also scrutinized using the Box and Whisker plot. A dataset not normally distributed is first transformed utilizing some mathematical tools, like log transformation, to make the values approaching symmetrical or normal distribution. New sets of values are formed after the transformations which are further processed for multivariate data analyses.

Correlation and multivariate statistical analyses: Interrelationships between the parameters were determined through correlation analysis applying Pearson correlation matrix. Two multivariate statistical data analyses techniques viz. cluster analysis (CA) and discriminant analysis (DA) were performed on the data matrix for advanced characterization of sampling sites and chemistry of the studied surface waters. CA is an exploratory data analysis tool for classifying objects into homogeneous groups or clusters where the scale of relationship is strong among individuals of the same cluster than the individuals of different clusters[12]. It can be done employing both hierarchical and non-hierarchical methods with former having some advantage over the latter, as hierarchical clustering doesn't require the number of clusters to be fixed in advance[13]. In the present study agglomerative hierarchical cluster analysis (AHCA) was carried out using Ward's method and squared Euclidean distance, as a measure of proximity, for the grouping of sampling sites. Euclidean distance is a common distance measure used and is given by[6]:

$$d_{ij} = \left[\sum_{l=1}^{q}\left(x_{il} - x_{jl}\right)^2\right]^{1/2} \quad (1)$$

Where, d_{ij} = Euclidean distance between two members or individuals i and j (d_{ij}^2 = squared Euclidean distance), each calculated on q variables, x_{il}, x_{jl}, l = 1, …., q. DA is another important multivariate technique for data classification along with dimensionality reduction[14]. It is commonly performed to examine dissimilarities between groups, to distinguish or discriminate groups effectively and to discover chief discriminating variables[15]. One or more discriminant functions (DF), defined as a linear combination of the weightings and scores on used variables, are generated through DA[16]. DFs are calculated as provided below[17,18]:

$$f(G_i) = k_i + \sum_{j=1}^{n} w_{ij} P_{ij} \quad (2)$$

Where, i = number of groups (G); k_i = constant inherent to each group; n = number of parameters used to classify a set of data into a given group; w_j = weight coefficient assigned by DA to a given parameter (P_j)[17,18]. In addition, DA can also be applied to test hypothesis on disparities between the expected combinations or groupings and to categorize new observations into already existing groups[15]. In our study DA was performed to uncover the most significant water quality components responsible for spatial differences between clusters obtained by CA. Sampling sites grouped according to the degree of anthropogenic influence through CA were taken as spatial grouping variables. Since, DA uses criteria of dependent and independent variables, spatial grouping variables were considered as dependent variables and water quality components were taken as independent variables. The DA was applied on the raw data and stepwise method was followed to identify the discriminating parameters.

Results and Discussion

Water chemistry: Hydrochemical characterization of water body presents the condition of water with respect to its quality measuring parameters considered under the study. We monitored and assessed sixteen parameters both spatially and temporally for the evaluation of water quality of the selected water bodies. Most of the parameters were found to be within the recommended standards, indicating their presence not harmful to the living organisms, especially the aquatic life[19]. However, DO and BOD values failed the prescribed standards at all the sampling sites spatio-temporally. Central Pollution Control Board (CPCB) of India classifies waters based on designated best use into five different classes[20]. From table-2 we can see that minimum recommended value for DO is 4 mg/l, and the maximum value of BOD in surface waters should be 3 mg/l. DO and BOD values recorded at different sampling sites, figure-2, showed deviations from these regulatory benchmarks. This trend analysis points to the presence of organic pollution in water bodies. Both natural and anthropogenic causes can be attributed to the presence of organic load in the surface waters. The water bodies are surrounded by natural vegetation and the fall and decomposition of plant litters make the water rich in organic matter. Use of water for domestic purposes like bathing and washing clothes and utensils throughout the year increases the organic pollution load. Moreover, these water bodies are also used for idol immersions which further enhance their organic matter contents. To investigate the influence of anthropogenic activities on surface waters we used Pous Mela as our sampling time as sites of pond 5 are extensively used during this period. Relatively stable water chemistry including DO and BOD values was noted at most of the sampling sites. However, Pond 5 displayed fluctuations in almost all important parameters showing increased concentration during the time of Pous Mela as compared to before Mela period. An increase in organic pollution indicated by DO and BOD values, Figure-2, was noticed during the Pous Mela time, specifying that anthropogenic activities are responsible for deterioration of water quality. Organic pollution can severely affect the water chemistry because of reduction in oxygen level and accumulation of degradation products like ammonia and hydrogen sulphide. Organic load can also increase the phosphorus content as it is released in the water medium after decomposition[21]. We noticed a sudden increase in phosphate (P) concentration in Pond 5 during the Mela period, Figure-3, whereas the rest displayed stability in its concentration.

Concentrations of 0.1 mg total P/l are unacceptably high and concentrations of 0.02 mg/l are often a problem[22]. During the study period, water bodies displayed phosphate levels undesirable to their proper ecological functioning. The situation may give rise to eutrophic conditions resulting in fish mortality (fishes are found in all surface water bodies taken in this study) and major shifts in the species composition at all trophic levels[22]. Some variations in water chemistry at P4S2 was also noticed during the Pous Mela time owing to the presence and use of diesel run pumping machine and generator and some construction works at the site.

Piper trilinear diagram: Some major ions determined in water samples were plotted on the Piper trilinear diagram and compared with other reported literature in order to classify and designate ionic nature of water[23]. This illustration of ionic signature helps in uncovering the principal ions controlling the water chemistry. Using the classification scheme given in figure-4, Piper diagram classified all water samples into 'Mixed Ca^{2+}—Na^{+}—HCO_3^{-} Type'. Classification of water types is displayed in figure-5 and figure-6. No change in water types was recorded temporally suggesting the ionic stability of water bodies with respect to Na^+, K^+, Ca^{2+}, Mg^{2+}, HCO_3^-, Cl^- and SO_4^{2-} and further indicating the impact of the Pous Mela on major ion chemistry of nearby waters as negligible. One peculiar observation noticed was that, though, pond 1 and other ponds are separated by a distance of about 5 km they displayed same water type. This suggests geogenic activities are primarily responsible for the occurrence of above mentioned ions in waters, and which also indicates similar geogenic processes control the ionic signature of water bodies in this area. Thus, through Piper diagram, apart from identifying the nature of water samples, their relationships among each other can also be uncovered. Geologic units with chemically similar water can be recognized and classified followed by trend analysis of water chemistry along the flow path[24].

Table-2
Established classification of surface waters by CPCB (India)

Designated best use	Class of water	Criteria
Source of drinking water (after disinfection, but exclusive of conventional treatment)	A	i. pH = 6.5 – 8.5 ii. DO = 6 mg/l or more iii. BOD_5 at 20°C = 2 mg/l or less
Outdoor bathing (organized)	B	i. pH = 6.5 – 8.5 ii. DO = 5 mg/l or more iii. BOD_5 at 20°C = 3 mg/l or less
Source of drinking water (after proper conventional treatment and disinfection)	C	i. pH = 6 – 9 ii. DO = 4 mg/l or more iii. BOD_5 at 20°C = 3 mg/l or less
For the propagation purposes of wild life and fisheries	D	i. pH = 6.5 – 8.5 ii. DO = 4 mg/l or more iii. Free ammonia (as N) = 1.2 mg/l or less
Controlled waste disposal, industrial cooling, irrigation	E	i. pH = 6.0 – 8.5 ii. EC = 2250 µmhos/cm, maximum, at 25°C

(Source: Guidelines for water quality management[20]; CPCB = Central Pollution Control Board)

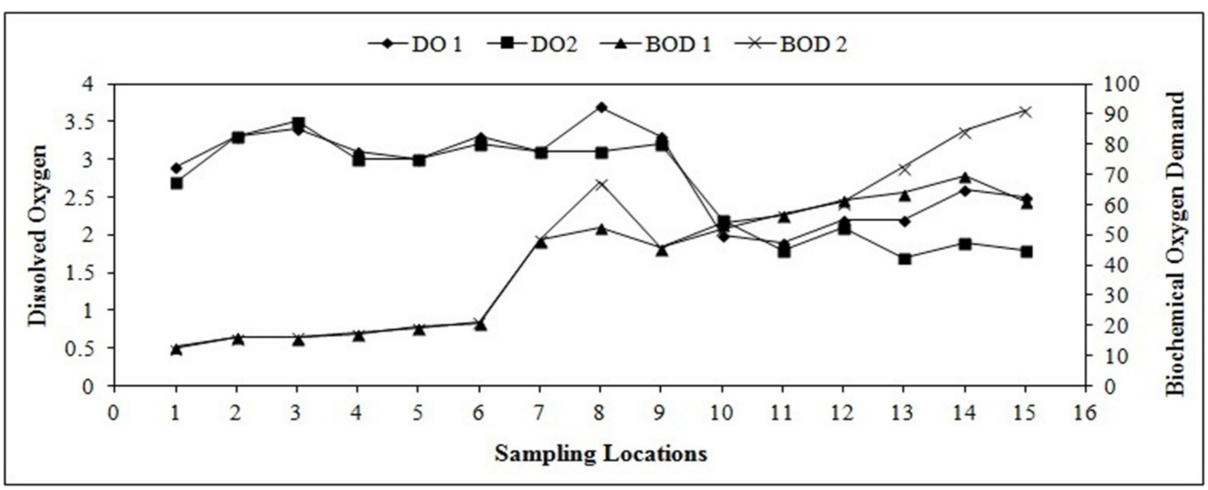

Figure-2
DO and BOD values recorded spatially and temporally;
(DO 1 = before mela, DO 2 = during mela; BOD 1 = before mela, BOD 2 = during mela)

Correlation analysis: The result of correlation analysis (full analysis not shown) performed on the combined dataset of three sampling periods is presented in table-3. Most of the parameters displayed significant correlation among each other indicating high interactions of the chemical constituents in surface waters. TDS showed strong positive correlation with all measured cations and anions, whereas it was negatively correlated with DO. Significant positive correlation was also observed between total hardness, ionic species and BOD. DO showed negative but significant correlation with BOD and PO_4^{3-}, while BOD was strongly and positively correlated with PO_4^{3-}. Similarly, very high relationship was noticed between NO_3^- and PO_4^{3-}.

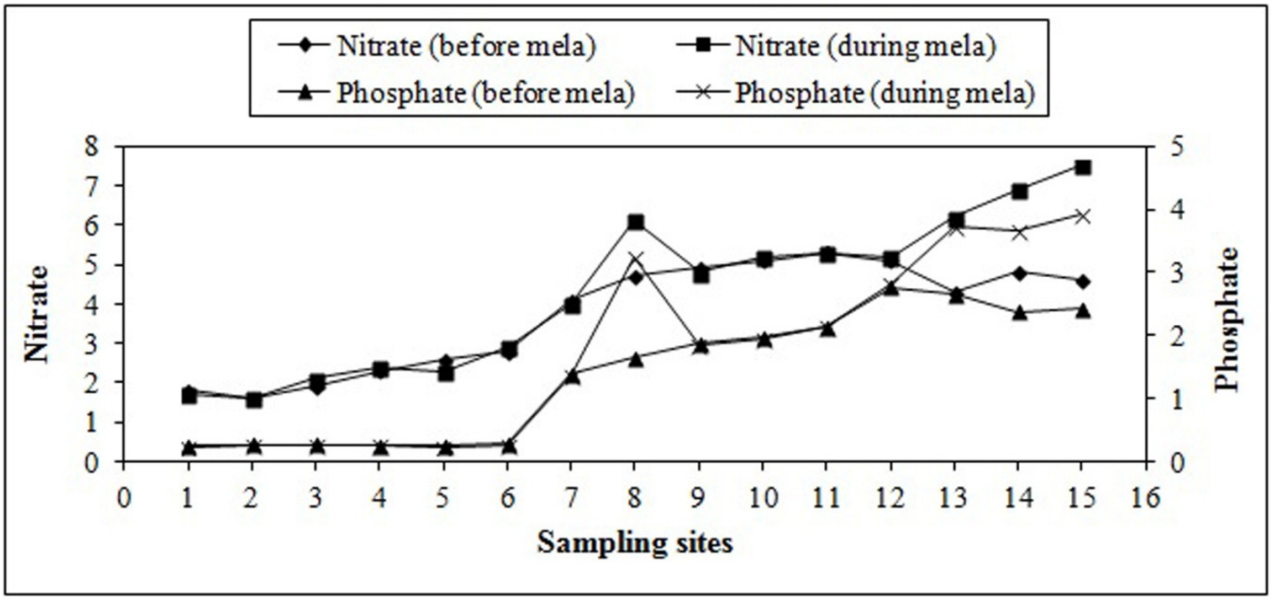

Figure-3
NO_3^- and PO_4^{3-} values recorded spatially and temporally

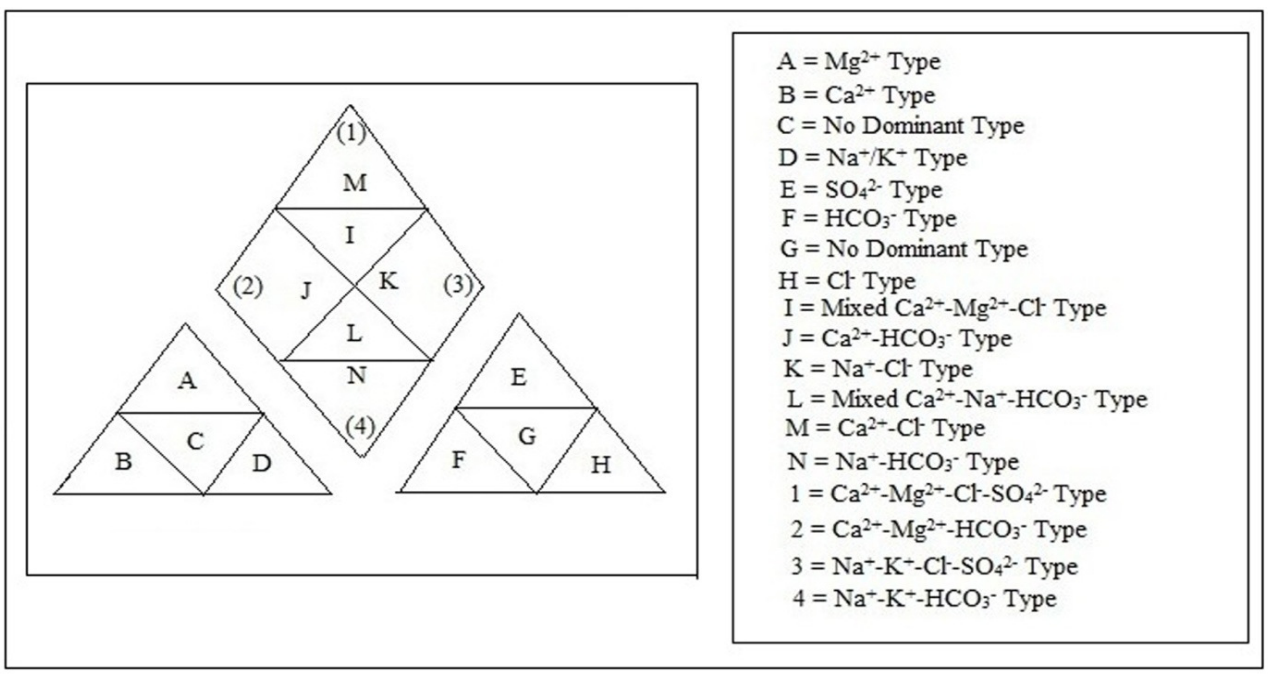

Figure-4
Reference Piper trilinear diagram; Left and right triangles designate cations and anions respectively (modified from elsewhere[23])

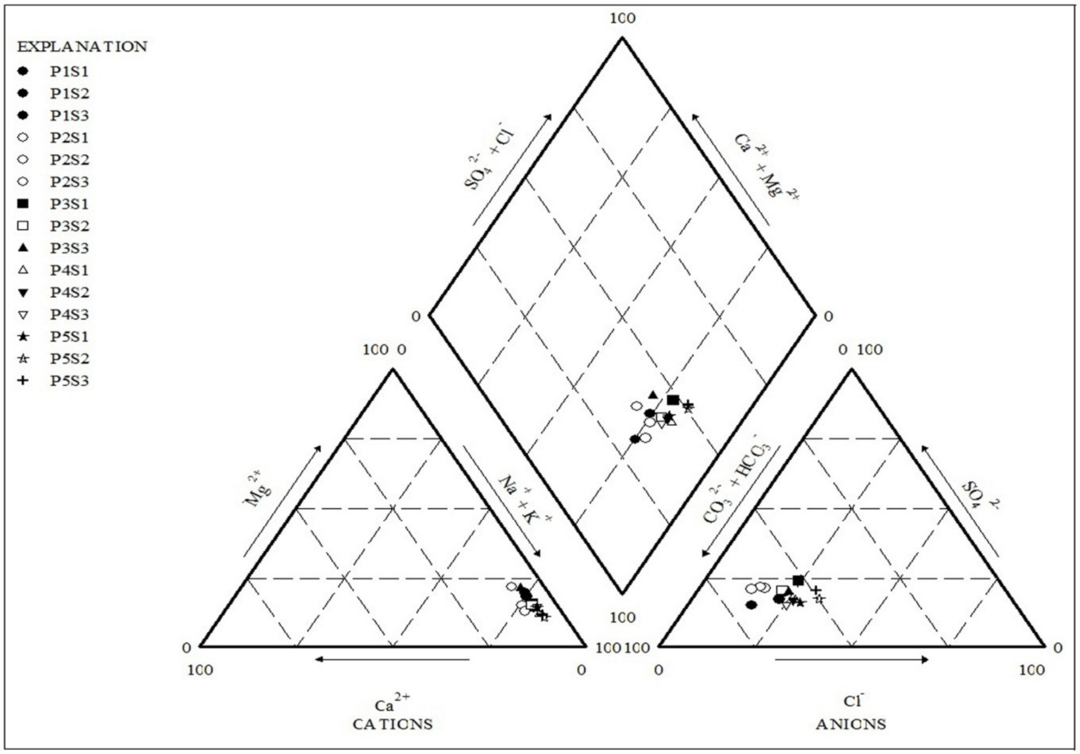

Figure-5
Piper trilinear diagram illustrating hydrochemical regime (before Pous Mela)

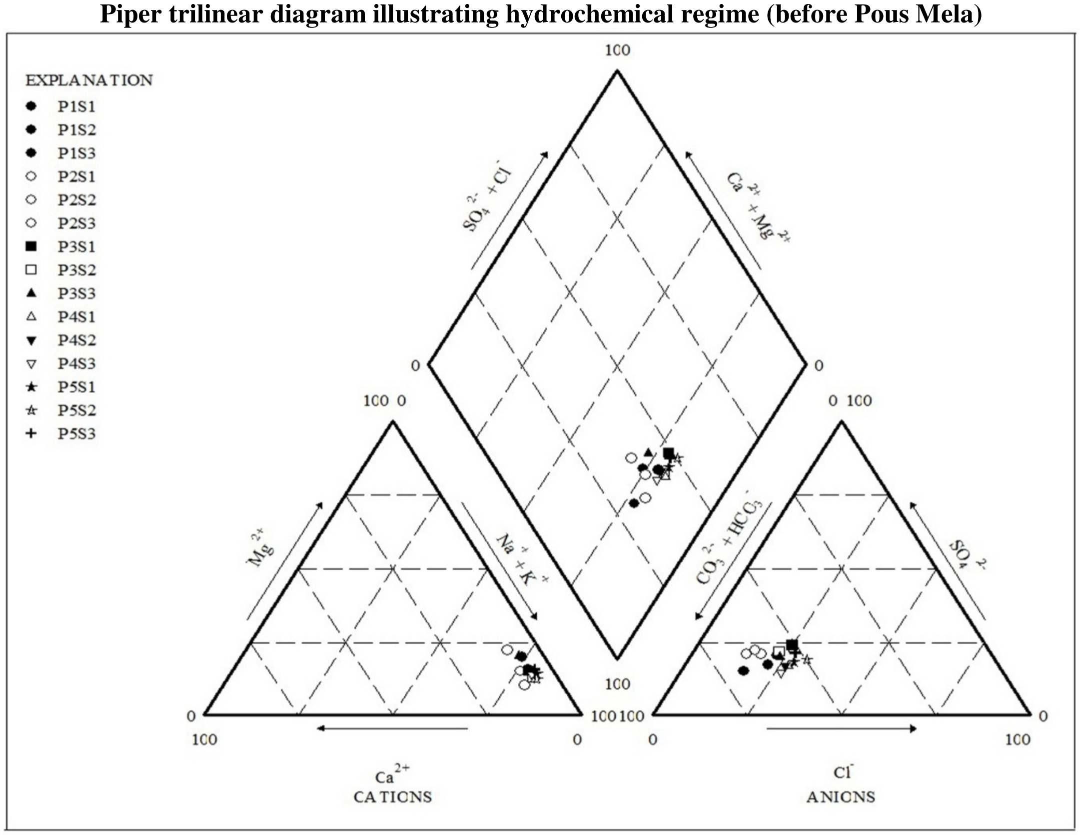

Figure-6
Piper trilinear diagram illustrating hydrochemical regime (during Pous Mela)

Cluster analysis (CA): The vital objective of doing CA was to classify the sampling sites based on their spatial similarities. CA was performed on the dataset of all three sampling campaigns which revealed similar dendrogram pattern before and after the Pous Mela. CA classified fifteen sampling sites into two clusters/ groups during these periods. Cluster 1 contained all sites belonging to ponds 1 and 2, while the rest were included in Cluster 2. However, a different scenario was noted during the Pous Mela period when the use of water of pond 5 suddenly escalated. The results of agglomerative hierarchical cluster analysis (AHCA) performed on the dataset after standardization is displayed as dendrogram in figure-7. The clustering pattern in figure-7 shows three statistically significant groups according to spatial resemblance. Here again, Cluster 1 includes all the sampling sites of ponds 1 and 2, whereas all sampling sites of pond 3 and two sampling sites of pond 4 have gone to Cluster 3. Cluster 2 includes sites where sudden increase in anthropogenic activities was noticed during the Pous Mela period. Accordingly, the three clusters can be categorized as: having less anthropogenic influence (Cluster 1), high anthropogenic influence (Cluster 2) and moderate anthropogenic influence (Cluster 3). CA emerged as a powerful exploratory data analysis technique which can be applied on large scale for spatial classification and categorization of natural resources. Since, sample collection and their laboratory analyses procedures are tedious and time consuming tasks, number of sampling sites can be effectively reduced in further analysis if they are organized into one group. For proper management of natural resources their large scale and continuous monitoring and assessments are needed. CA can be an effective statistical modeling tool for management of natural resources.

Discriminant analysis (DA): A stepwise DA was performed on three groups identified through AHCA involving fifteen sampling sites and sixteen water quality parameters. Stepwise DA was undertaken to find out the best combinations of independent or predictor variables differentiating three site clusters. The analysis generated two latent discriminant functions (DF) as linear combinations of indicator independent (predictor) variables and revealed that out of sixteen hydrochemical parameters only four parameters viz. BOD, pH, Cl^- and TH differentiated less, moderate and high anthropogenic influence sites. This also showed DA as a considerable dimension reduction technique. Wilks' Lambda table, presented as table-4, disclosed that all four above mentioned predictors were significant in adding some predictive power to the DF. With only four selected parameters all samples were correctly classified. Leave-one-out cross validation classification procedure, table-5, showed 93.30% of correctly classified cases. Table-6 displays the relative importance of DF1 compared to DF2 in explaining the differences among three groups. DF1 elucidates 61.10% of the variance among the three groups, while DF2 explains 38.90%. The degree of relationship between selected predictors and groups is also slightly higher for DF1 than DF2 as shown by their respective canonical correlation values of 0.973 and 0.959.

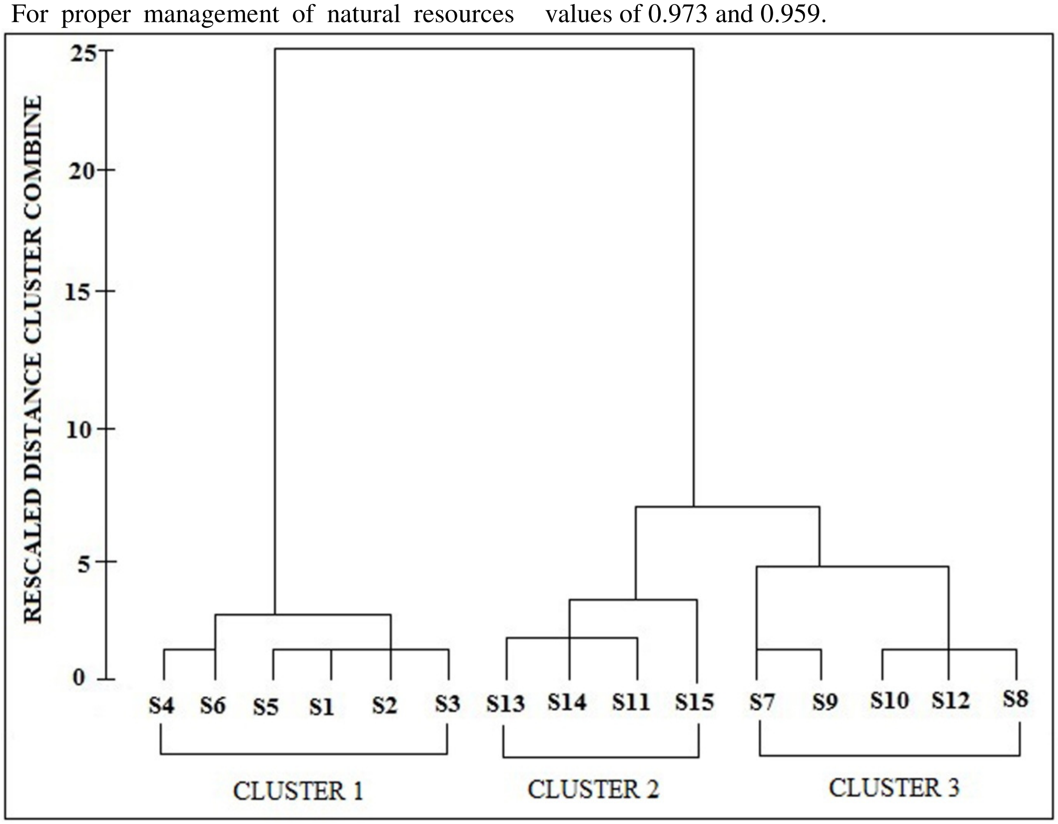

Figure-7
Agglomerative hierarchical cluster analysis displaying three distinct groups

Table-3
Correlation matrix of water chemistry

	TH	Mg^{2+}	Cl^-	HCO_3^-	Na^+	K^+	Ca^{2+}	TDS	DO	BOD	NO_3^-	PO_4^{3-}	SO_4^{2-}
TH	1	.718**	.751**	.813**	.805**	.689**	.887**	.818**	-.640*	.663**	.735**	.676**	.696**
Mg^{2+}		1	.849**	.760**	.843**	.876**	.817**	.862**	-.670**	.770**	.763**	.822**	.783**
Cl^-			1	.864**	.993**	.954**	.961**	.976**	-.850**	.916**	.878**	.931**	.894**
HCO_3^-				1	.896**	.823**	.889**	.938**	-.742**	.730**	.739**	.776**	.759**
Na^+					1	.948**	.978**	.987**	-.853**	.895**	.865**	.907**	.892**
K^+						1	.908**	.947**	-.726**	.861**	.806**	.878**	.873**
Ca^{2+}							1	.969**	-.814**	.890**	.896**	.894**	.877**
TDS								1	-.828**	.866**	.851**	.889**	.885**
DO									1	-.670**	-.650**	-.695**	-.640*
BOD										1	.973**	.982**	.937**
NO_3^-											1	.956**	.917**
PO_4^{3-}												1	.908**
SO_4^{2-}													1

**Significant correlation at the 0.01 level (two tailed), *Significant correlation at the 0.05 level (two tailed)

Table-4
Wilks' Lambda table

Step	Entered variable	Wilks' Lambda Statistics							
						Exact F			
		Statistic	Df1	Df2	Df3	Statistic	Df1	Df2	significance
1	BOD	0.105	1	2	12.000	51.379	2	12.000	0.000
2	pH	0.041	2	2	12.000	21.784	4	22.000	0.000
3	Cl^-	0.014	3	2	12.000	25.241	6	20.000	0.000
4	TH	0.004	4	2	12.000	32.113	8	18.000	0.000

Table-5
Classification results

		group	Predicted Group Membership			Total
			1	2	3	
Original	Count	1	6	0	0	6
		2	0	4	0	4
		3	0	0	5	5
	%	1	100.00	0	0	100.00
		2	0	100.00	0	100.00
		3	0	0	100.00	100.00
Cross-validated	Count	1	6	0	0	6
		2	1	3	0	4
		3	0	0	5	5
	%	1	100.00	0	0	100.00
		2	25.00	75.00	0	100.00
		3	0	0	100.00	100.00
Correctly classified original grouped cases = 100.00% Correctly classified cross-validated grouped cases = 93.30%						

Table-6
Outline of canonical discriminant functions

Eigenvalues				
Function	**Eigenvalue**	**Variance %**	**Cumulative %**	**Correlation**
1	17.858	61.10	61.10	0.973
2	11.369	38.90	100.00	0.959
Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	0.004	57.247	8	0.000
2	0.081	26.409	3	0.000
Discriminant function coefficients				
	Function			
	1		2	
TH	0.798		1.284	
Cl^-	-2.311		-1.474	
pH	1.661		0.931	
BOD	0.675		1.455	

Conclusion

Three multivariate models viz. Piper trilinear diagram, AHCA and DA used in this study successfully characterized and classified the water chemistry of selected water bodies of Santiniketan-Bolpur-Sriniketan zone in Birbhum district, West Bengal, India. Piper diagram revealed water type of each location and suggested that similar geogenic processes were controlling the ionic behaviour of the zone. AHCA grouped sampling sites, of similar characteristics, into two (before and after the Pous Mela) and three (during the Pous Mela) groups indicating the influence of anthropogenic activities on alteration of surface water chemistry. This also showed applicability of AHCA in other studies, like on river water, where numbers of sampling sites and so the cost of experiments can be brought down if the sampling sites form groups on the basis of their similar properties. DA applied on the data matrix classified 93.30% of cross-validated grouped cases correctly, and identified four principal hydrochemical parameters i.e. BOD, pH, Cl^- and TH, responsible for differentiating the groups. Thus, it rendered significant data reduction and recognized the most significant marker parameters explaining the variability of water chemistry. As this study shows, for a well-organized and effective assessment and monitoring, review and designing schemes of any water quality management programme multivariate models can be utilized. Multivariate graphical and hydrostatistical analyses should be inherent part of any water quality management strategy.

References

1. Tiwari M. and Ranga M.M., Assesment of Diurnal Variation of Physico Chemical Status of Khanpura Lake, Ajmer, India, *Res.J.Chem.Sci.,* **2** (**7**), 69-71 (**2012**)
2. Deshpande S.M. and Aher K.R., Evaluation of Groundwater Quality and its Suitability for Drinking and Agriculture use in Parts of Vaijapur, District Aurangabad, MS, India, *Res.J.Chem.Sci.,* **2** (**1**), 25-31 (**2012**)
3. Ranjan R., Water Quality Monitoring of Groundwater Resources around Sugar Factory, Near East-West Champaran Border, Bihar, India, *Res.J.Chem.Sci.,* **2** (**7**), 79-81 (**2012**)
4. Patil S.G., Chonde S.G., Jadhav A.S. and Raut P.D., Impact of Physico-Chemical Characteristics of Shivaji University lakes on Phytoplankton Communities, Kolhapur, India, *Res.J.Recent Sci.,* **1** (**2**), 56-60 (**2012**)
5. Parihar S.S., Kumar A., Kumar A., Gupta R.N., Pathak M., Shrivastav A. and Pandey A.C., Physico-Chemical and Microbiological Analysis of Underground Water in and Around Gwalior City, MP, India, *Res.J.Recent Sci.,* **1** (**6**), 62-65 (**2012**)
6. Landau S. and Everitt B.S., A Handbook of Statistical Analyses using SPSS, Chapman & Hall/CRC Press LLC, USA (**2004**)
7. Härdle, W. and Simar, L., Applied Multivariate Statistical Analysis. http://www.stat.wvu.edu/...mva.pdf, Accessed on 24. 03. 2012 (**2003**)
8. Birbhum District Map, Official Website of the Birbhum District, West Bengal, India, http://www.birbhum.gov.in/distmap.htm, Accessed on 20.02.2013 (**2013**)
9. Santiniketan – Google Maps, Imagery ©2013 Cnes/Spot Image, DigitalGlobe, GeoEye, Map data ©2013 Google, http://maps.google.co.in/maps...santiniketan&bav=on.2, Accessed on 20.02.2013 (**2013**)
10. Bhardwaj V., Singh D.S. and Singh A.K., Water Quality of the Choti Gandak River using Principal Component Analysis, Ganga plain, India, *J. Earth Syst. Sci.,* **119**, 117-127 (**2010**)

11. American Public Health Association (APHA)., Standard Methods for the Examination of Water and Wastewater, 21st Centennial Edition, APHA, AWWA, WPCF, Washington DC, USA **(2005)**

12. Kašparová M., Křupka J. and Chýlková J., Heavy Metals Contamination Analysis in Selected Czech Localities by Cluster Analysis, In: Recent Advances in Environment, Ecosystems and Development, Proceedings of the 7th WSEAS International Conference on Environment, Ecosystems and Development, December 14-16, WSEAS Press, Tenerife, Spain, 240-245 **(2009)**

13. Kašparová M., Křupka J. and Chýlková J., Comparison of Cluster Analysis Algorithms for Heavy Metals Contamination in Atmospheric Precipitation, *WSEAS Trans. Inf. Sci. Appl.,* **3** (**7**), 341-350 **(2010)**

14. Balakrishnama S. and Ganapathiraju A., Linear Discriminant Analysis - A Brief Tutorial, Institute for Signal and Information Processing, Mississippi State University http://www.music.mcgill.ca/....lda_theory.pdf, Accessed on 12.12.2012 **(2012)**

15. Fernandez G.C.J., Discriminant Analysis, A Powerful Classification Technique in Data Mining, Paper 247-27, In: Statistical and Data Analysis, SUGI 27 Proceedings, Orlando, Florida, April 14-17 **(2002)**

16. Ramayah, T., Ahmad, N.H., Halim, H.A., Zainal, S.R.M. and Lo, M.C., Discriminant Analysis: An Illustrated Example, *Afr. J. Bus Manageme.,* **4** (**9**), 1654-1667 **(2010)**

17. Singh K.P., Malik A. and Sinha S., Water Quality Assessment and Apportionment of Pollution Sources of Gomti River (India) using Multivariate Statistical Techniques-A Case Study, *Anal. Chim. Acta,* **538**, 355-374 **(2005)**

18. Juahir H., Zain S.M., Yusoff M.K., Hanidza T.I.T., Armi A.S.M., Toriman M.E. and Mokhtar M., Spatial Water Quality Assessment of Langat River Basin (Malaysia) using Environmetric Techniques. *Environ. Monit. Assess.,* **173**, 625-641 **(2011)**

19. Bureau of Indian Standards (BIS), Indian Standard Drinking Water Specification (second revision of IS: 10500), Manak Bhawan, New Delhi, India **(2004)**

20. Central Pollution Control Board (CPCB)., Use Based Classification of Surface Waters in India, In: Guidelines for water quality management, Parivesh Bhawan, Delhi, India **(2008)**

21. Goel P.K., Water Pollution: Causes, Effects and Control, Revised Second Edition, New Age International (P) Limited Publishers, New Delhi, India **(2009)**

22. Correll D.L., Phosphorus: A Rate Limiting Nutrient in Surface Waters, *Poultry Sci.,* **78**, 674-682, **(1999)**

23. Ravikumar P. and Somashekar R.K., Multivariate Analysis to Evaluate Geochemistry of Ground Water in Varahi River Basin of Udupi in Karnataka, India, *The Ecoscan*, **4 (2&3)**, 153-162 **(2010)**

24. Zhang P., EAS 44600 Groundwater Hydrology, Lecture 14: Water Chemistry 1, http://mail.sci.ccny.cuny.edu/~pzhang..14.pdf, Accessed on 10.03.2013 **(2013)**