

High nitrate content in the surface water of Balipara, North Brahmaputra river basin, Sonitpur district, Assam, India: a multivariate approach

Richa Chaturvedi^{1,3}, Saumen Banerjee^{1,*},
Bodhaditya Das¹, Pronobesh Chattopadhyay²,
Chira R. Bhattacharjee³ and Vijay Veer¹

¹Department of Chemistry and

²Department of Pharmaceutical Technology,
Defence Research Laboratory, Tezpur 784 001, India

³Department of Chemistry, Assam University, Silchar 788 001, India

The present study is an evaluation of ground and surface water quality at Balipara, North Brahmaputra river basin, Sonitpur district, Assam, India using multivariate statistical methods. The results show high concentration of Fe, Mn, Pb and Cr in groundwater. Arsenic was observed in both ground and surface water. In the surface water, nitrate content was also found to be high. Ward's method was used for hierarchical agglomerative cluster analysis. A close relationship between electrical conductivity (EC) and total dissolved solids (TDS) was established by the smallest proximity distance between these parameters. Group 1 comprised of TDS, EC, total alkalinity (TA), F, Ca, Pb, Cr and Cl based on proximity distances. Group 2 consisted of Fe, Mn, As and Group 3 of TH, Mg, pH, Zn, SO₄ and NO₃ in groundwater. In surface water, Group 1 comprised of TDS, EC, SO₄, NO₃, Cl, Zn, pH and Ca. Group 2 accounted for Mg and F and Group 3 for Fe, Mn, TA, Pb, As and Cr. The interrelationships between the contaminants depicted by cluster analysis, categorize the contamination levels. Factor analyses were applied for understanding the interrelationships between the variables and for identifying probable source components. Six factors justifying 83.64% of the total variance in groundwater and five factors describing 81.92% of the total variance in surface water were found responsible for variation in the data structure. The relative contribution of all the water-quality parameters was best explained by discriminant analysis.

Keywords: Contaminants, groundwater and surface water, multivariate statistical techniques, water quality parameters.

WATER quality is a major environmental concern across the world and is considered to be the main factor influencing health and the state of disease in both men and animals. Anthropogenic activity as well as natural processes degrade ground and surface water, and restrict their use for drinking, industrial, agricultural, recreational or

other purposes^{1,2}. Human activities degrade the ground and surface water quality through atmospheric pollution, effluent discharge, use of agricultural chemicals, soil erosion and land use³. Annually about 5 million people die in the developing world due to water-borne diseases, though much of these can be prevented with adequate water supply⁴.

Rural communities around the world, get their water supply directly from rivers or from shallow dug wells. In the absence of surface water, groundwater is the main source and serves an estimated 20% of the global population who live in arid and semi-arid regions⁵.

In the North East states of India, groundwater is highly ferruginous^{6,7} and scanty information is available from the water quality of the North Brahmaputra basin, Sonitpur district, Assam. There is dearth of data pertaining to analysis of water-quality parameters.

In the present study, data obtained for 17 parameters were analysed during multivariate statistics to find out the suitability and potability of ground and surface water.

For a better understanding of the water quality and ecological status of the studied systems, identification of possible factors/sources that influence water systems, multivariate statistical techniques such as cluster analysis (CA), principal component analysis (PCA), factor analysis (FA) and discriminant analysis (DA) plays an important role by helping in the interpretation of complex data matrices and offering a valuable tool for reliable management of water resources as well as rapid solution to pollution problems⁸⁻¹⁵. These multivariate statistical techniques have been applied successfully to characterize and evaluate surface water and groundwater quality¹⁵⁻¹⁹.

For the present study, the area chosen was Balipara, Sonitpur district, Assam, India (Figure 1). The district is situated over an area of 5324 sq. m on the north bank of the Brahmaputra River. Land use in this area is divided primarily among tropical semi-evergreen, moist deciduous, grassland, agricultural land and tea garden. The district lies between 26°30'N and 27°01'N lat. and 92°16' and 93°43'E long. Agriculture is the main occupation of people of this district; tea gardens are the next major controlling factor of the economy of the district. The district has practically no large-scale industries and is economically backward. The primary drinking water sources in the rural areas of Sonitpur district are shallow wells, shallow tube wells and deep tube wells. Lifestyle of the inhabitants and their economic positions affect the water used within the home in different parts of the district.

Table 1 summarizes the water-quality parameters, units of measurement and analysis methodology. Seventy water samples, 35 each of ground and surface water respectively, were collected from Balipara town. These include water samples from shallow wells, shallow tube wells and the deep tube wells. The groundwater samples from deep tube wells were operated at least 5 min before collection to flush out the stagnant water inside the tube

*For correspondence. (e-mail: saumenbanerjee2004@gmail.com)

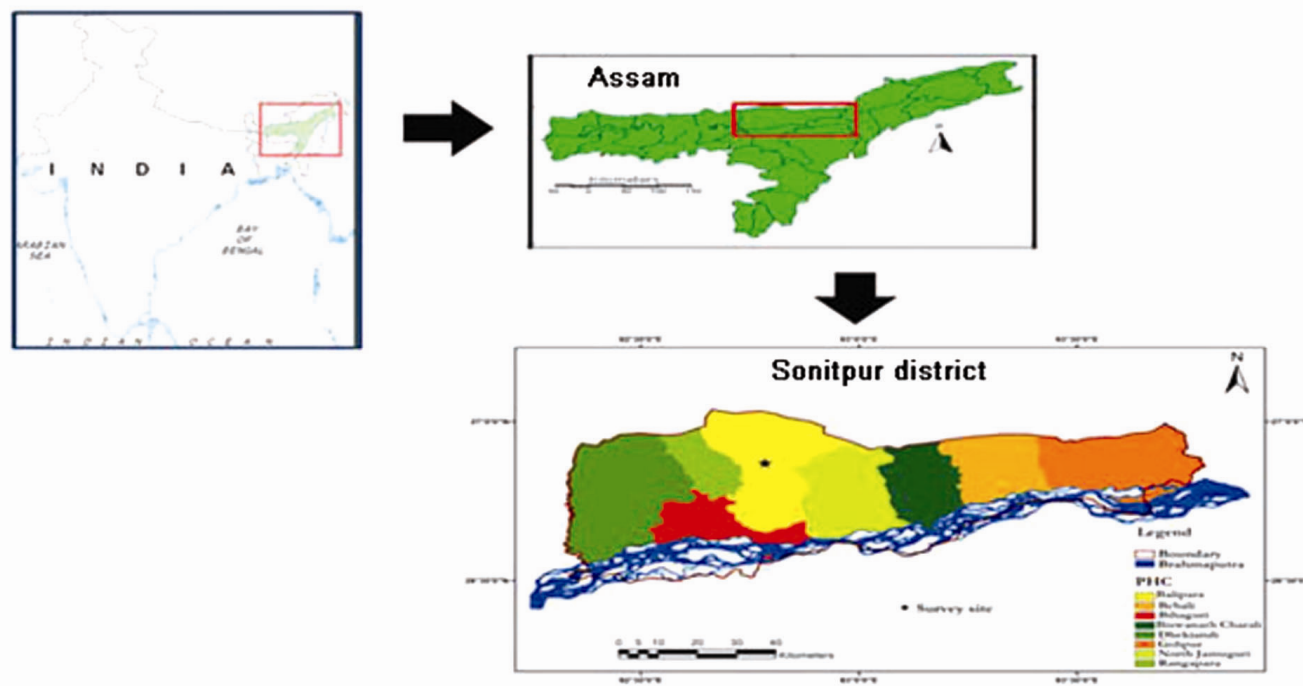


Figure 1. Location map of the study area.

Table 1. Abbreviations, units and analytical methods used in water quality parameters

Variables	Abbreviations	Units	Analytical methods
pH	pH	pH unit	pH-meter
Total dissolved solids	TDS	mg/l	TDS-meter
Electrical conductivity	EC	mS/cm	EC meter
Total alkalinity	TA	mg/l	Titrimetric
Total hardness	TH	mg/l	Titrimetric
Fluoride	F	mg/l	Ion selective electrode
Sulphate	SO ₄	mg/l	Spectrophotometric
Nitrate	NO ₃	mg/l	Ion selective electrode
Chloride	Cl	mg/l	Titrimetric
Calcium	Ca	mg/l	Titrimetric
Magnesium	Mg	mg/l	Titrimetric
Zinc	Zn	mg/l	AAS
Lead	Pb	mg/l	AAS
Chromium	Cr	mg/l	AAS
Iron	Fe	mg/l	AAS
Arsenic	As	mg/l	Hydride generation AAS
Manganese	Mn	mg/l	AAS

and get fresh groundwater. The containers were first washed-off with the water sample to be collected. The water samples were collected in one litre polyvinyl chloride container and were acidified with 10 N (10 ml) nitric acid for heavy metal analysis. pH was determined in the field at the time of sampling using a calibrated portable pH meter (Eutech Instruments, Singapore). Samples were analysed for total alkalinity (TA), total dissolved solids (TDS), electrical conductivity (EC), total hardness (TH) as CaCO₃, nitrate as NO₃, sulphate as SO₄, chloride as Cl, and fluoride as F following standard procedure²⁰.

Analytical grade reagents and type-I water (Elga Composite vent Millipore filter, UK) were used throughout for preparation of working standards. Prior to analysis, the instruments were calibrated and standardized as usual. TH, Ca and Mg ions were determined by titrimetry using ethylenediaminetetraacetic (EDTA) acid as titrant. HCl titrimetric method was followed to estimate total alkalinity and argentometric titration was used to quantify chloride. Fluoride and nitrate were measured using an ion analyser (Orion, Thermo Scientific Ltd) equipped individually with fluoride and nitrate electrode respectively.

Sulphate content was quantified turbidimetrically at 450 nm using a UV/Vis-spectrophotometer (SPECORD 40, Analytic Jena, Germany). EC and TDS were measured by EC-TDS meter. The heavy metals (Fe, As, Mn, Pb, Cr and Zn) were measured using atomic absorption spectrophotometer (AAS; Lab India AA 7000). The calibration curves were constructed for each of the determinants prior to analysis as usual. A blank sample was analysed between each reading to measure the baseline stability of the instrument. Calibration of the instrument was confirmed by additionally analysing standard solution after a batch of every ten samples. All the analysis was carried out in triplicate.

Arsenic (As) was quantified by AAS (detection limit 0.02 µg/l). Assuming As (V) may be present in the water samples along with As (III), reduction of As (V) to As (III) was performed using potassium iodide solution and ascorbic acid in moderately concentrated (5 mol/l) HCl solution. Time for reduction was 30 min. Next 10 ml of reduced water samples was analysed using AAS with MHS-15 (mercury hydride generation system). Hydride generation was done using argon gas and sodium borohydride. Oxyacetylene flame was used for quantification of heavy metals.

SPSS 16 statistical software package was used for performing all the statistical calculations. Standardization of the data was done by Z-score transformation which renders the data dimensionless^{21,22}. This transformation also increases the influence of variables whose variance is small and vice versa.

CA is a technique in which the variables are categorized based on their distance or similarity between the objects to be clustered²³. The analysis categorizes objects, so that each object is similar to others in the cluster with respect to a predetermined selection criterion²⁴. The resulting clusters of the objects should then exhibit high internal (within-cluster) homogeneity and high external (between clusters) heterogeneity²⁴. Clusters are formed sequentially starting with the most similar pair of variables and then forming higher clusters step by step in hierarchical agglomerative clustering²³, and are exemplified by a dendrogram (tree diagram)²⁴. Visual summary of the clustering processes as shown by the dendrogram, presents a picture of the groups and their proximity with a dramatic reduction in dimensionality of the original data²⁴. The similarity between the two samples is given by the Euclidean distance, which can be represented by the difference between analytical values from the samples^{23,25}. In this study, hierarchical agglomerative CA was performed on the normalized dataset by means of the Ward's method²³, using squared Euclidean distance as a measure of similarity.

In PCA, new, uncorrelated variable (axes) called principal components (PCs), are obtained from the original variables which are their linear combinations²⁴⁻²⁶. The new axes lie along the directions of maximum vari-

ance^{26,27}. PCA provides an objective way of finding indices of maximum variance²⁸, so that variation in the data can be transformed into a smaller set of independent (uncorrelated) variables²⁷. This analytical technique also exerts the Eigen values and Eigen vectors from the correlation matrix of original variables. PC provides information on the most meaningful parameters and describes a whole dataset affording data reduction with minimum loss of the original information^{8,16,28}. The mathematical expression of PC is as follows

$$Z = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}x_{3j} + \dots + a_{im}x_{mj}, \quad (1)$$

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

PCA is followed by FA. FA is to reduce the contribution of less significant variables to further simplify the data structure coming from PCA²⁸. This can be achieved by rotating the axis defined by PCA, according to the rules, and constructing new variables, also called varifactors (VFs)²⁸. PC is a linear combination of observable water quality variables, whereas VF can include unobservable, hypothetical, latent variables^{8,16}. To extract significant PCs and reduce the contribution of variables with minor significance, PCA was performed and then these PCs were subjected to varimax rotation (raw) generating VFs^{15,17,28-31}. As a result, approximately the same amount of information will be obtained by a small number of factors as does the much larger set of original observations²⁸. Mathematically, FA can be expressed as

$$Z = a_{f1}f_{1i} + a_{f2}f_{2i} + a_{f3}f_{3i} + \dots + a_{fm}f_{mi} + e_{fi}, \quad (2)$$

where z , a , f , e , i and m are the measured variable, factor loading, factor score, the residual term accounting for errors or other source of information, sample number and total number of factors respectively²⁸.

DA classifies cases into category-dependent values, usually a dichotomy²⁸. If DA is effective for a set of data, correct percentage will be yielded by the classification table²⁸. The most important application of DA is the description of group separation in which linear function of discriminant variables explains the variation between many groups (in this case, two) and relative contribution of all variables to separation of the groups is identified^{28,32}. Another aspect is the prediction of observations to groups in which linear or quadratic functions of the variable (classification functions (CFS)) are used to assign an observation to one of the groups^{28,33,34}. In the DA technique, a discriminant function (DF) is obtained which operates on raw data and for each group a discriminant function is generated as^{11,15,17,28,35}

$$F(G_i) = k_i + \sum_{j=1}^n w_{ij}p_{ij}, \quad (3)$$

where i is the number of groups (G), k_i is a constant inherent to each group, n is the number of parameters used to categorize a set of data into a given group and w_{ij} is the weight coefficient assigned by DA to a given parameter (p_{ij}). The distance between the means of the criterion (dependent) variable is maximized by the weight coefficient²⁸. The performance of DA is assessed by the classification table²⁸.

The present study shows that the surface water is more acidic (5.22) than groundwater (6.36). The low pH indicates the leaching probability of humic acid from the soil to the surface water³⁶. Higher concentration of iron (5.34 mg/l), manganese (2.12 mg/l), chromium (0.16 mg/l) and lead (0.07 mg/l) was observed in groundwater than surface water. An interesting observation of the coexistence of Fe and Mn has been reported³⁷. However, where this occurs, the concentration of iron is generally found greater because it has greater crustal abundance³⁸. All these values are higher than the limits set by WHO³⁹ and BIS⁴⁰. The concentration of arsenic (0.02 mg/l) also exceeded the desirable limit of WHO³⁹ and BIS⁴⁰ both in ground and surface water. Studies also reflect high concentrations of arsenic in the tea gardens of Assam⁴¹. On the contrary, surface water recorded higher values of nitrate (154.49 mg/l) than groundwater. This value is high compared to that prescribed by WHO³⁹ and BIS⁴⁰. The high concentration of nitrate in surface water of this area would be probably due to the use of chemical fertilizers in agricultural farmlands and decomposition of organic matter. The source of nitrate contamination in surface water may be due to leaching of these plant nutrients from the agricultural farmlands⁴². Table 2 provides the descriptive statistics (range, mean and standard error of mean (SEM)) for ground and surface water of Balipara town.

As mentioned earlier, CA was carried out to categorize the contamination levels. FA was applied separately to ground and surface water datasets to understand the inter-relationships between the variables and identify probable source components. DA was applied to identify the relative contribution of water quality parameters in discriminating the ground and surface water.

CA was used to determine the pattern identification of water quality parameters. Proximity matrix was used to determine the distances between the 17 studied parameters in both ground and surface water. Proximity distances range from 0.061 to 115.73 in groundwater (Table 3). The proximity was smallest between EC and TDS (0.061), and largest between pH and Fe (115.73). Grouping of parameters was performed based on proximity between the parameters. This yielded a dendrogram (Figure 2 a), grouping all the 17 water quality parameters of the groundwater into three statistically significant clusters (groups).

Group 1: TDS, EC, TA, F, Ca, Pb, Cr and Cl

Group 2: Fe, Mn, As

Group 3: TH, Mg, pH, Zn, SO₄ and NO₃

The proximity between Zn and Mn was 99.0, which is larger than that between Fe and Mn (12.45). This indicates good correlation between Fe and Mn and poor correlation between Zn and Mn. The close relationship between the grouped parameters reveals similar anthropogenic sources, mainly the use of fertilizers for irrigation; which are discharged into the groundwater. As shown in Figure 2 a, TDS, EC, TA, F, Ca, Pb, Cr and Cl form a common group. Similarly, Fe, Mn and As form one group, while TH, Mg, pH, Zn, SO₄ and NO₃ form another group, showing that the parameters are related to each other.

In case of surface water, proximity distances range from 1.304 to 123.61 (Table 4). The smallest value was between EC and TDS (1.30), while maximum proximity was found between pH and Fe (123.61). Figure 2 b. shows the dendrogram of surface water, grouping all the 17 parameters into three statistically significant clusters (groups).

Group 1: TDS, EC, SO₄, NO₃, Cl, Zn, pH and Ca

Group 2: TH, Mg and F.

Group 3: Fe, Mn, TA, Pb, As and Cr.

As shown in Figure 2 b, TDS, EC, SO₄, NO₃, Cl, Zn, pH and Ca form a common group, Similarly, TH, Mg and F formed one group, while Fe, Mn, TA, Pb, As and Cr form another group, revealing their close relationship.

Complex linear correlation between metal concentration in surface and groundwater of the study area was further studied by factor analysis. After varimax rotation, elements which belonged to a given factor were defined by the factor matrix, while those which had strong correlations were grouped into factors. The identification of factors is based on the dominant influence of contaminant present in the aquifer (Tables 5 and 6 respectively). Based on varimax rotation matrix, six factors were extracted from groundwater data explaining 83% of the total variance; in case of surface water, five factors were extracted explaining 81% of the total variance. Tables 5 and 6 give the Eigen values, total variance and cumulative variance for ground and surface water respectively.

For groundwater, factor 1 contributes 26% of the total variance of 83% with positive loading on TDS, EC, TA and moderate loading on As and Cr. It indicates a strong association ($r = 0.572-0.894$) among TDS, EC, TA, As and Cr. EC and TDS control the overall mineralization. The influence of EC and TDS is augmented by the ions in the following order: TA > As > Cr (ref. 20). Mn and Fe were found to have a positive loading contributing to 20% of the variance²⁰. The data support the coexistence of Fe and Mn³⁸. Factor 3 exhibits 13% of the total variance with positive loading on Ca and F and low loading for the

Table 2. Descriptive statistics for groundwater and surface water (mg/l) in Sonitpur district, Assam

Parameters	WHO/BIS limits		Groundwater	Surface water
pH	6.5–8.5	Range	6.36–7.72	5.22–7.74
		Mean	7.00	6.80
		SEM	0.06	0.11
TDS (mg/l)	500–2000	Range	1.5–57.00	5.09–46.80
		Mean	20.07	16.93
		SEM	1.94	1.96
EC (mS/cm)	1.5	Range	3.01–114.0	9.0–93.7
		Mean	40.34	33.15
		SEM	3.92	4.00
TA (mg/l)	200–400	Range	8.0–172.0	7.5–92.0
		Mean	64.23	39.25
		SEM	5.75	3.83
TH (mg/l)	200–600	Range	12–150	21.6–240
		Mean	61.71	75.30
		SEM	6.76	8.37
F (mg/l)	1–1.5	Range	0.005–0.92	0.075–0.92
		Mean	0.12	0.19
		SEM	0.03	0.03
SO ₄ (mg/l)	200–400	Range	3.43–23.10	2.70–48.44
		Mean	11.26	11.32
		SEM	0.74	1.78
NO ₃ (mg/l)	45-NR	Range	8.19–41.86	8.19–154.49
		Mean	18.58	45.24
		SEM	1.48	6.96
Cl (mg/l)	250–1000	Range	28–204	16.32–164.0
		Mean	70.91	60.71
		SEM	7.19	5.67
Ca (mg/l)	75–200	Range	2.4–119.2	0.8–49.6
		Mean	18.00	14.33
		SEM	3.27	2.26
Mg (mg/l)	30–100	Range	0.49–21.87	1.93–49.57
		Mean	7.15	9.59
		SEM	1.21	1.68
Zn (mg/l)	5–15	Range	0.0–0.40	0.052–0.62
		Mean	0.11	0.246
		SEM	0.02	0.02
Pb (mg/l)	0.01-NR	Range	0.02–0.07	0.0001–0.05
		Mean	0.045	0.02
		SEM	0.003	0.0034
Cr (mg/l)	0.05-NR	Range	0.00–0.16	0.0001–0.096
		Mean	0.04	0.025
		SEM	0.01	0.0055
Fe (mg/l)	0.3–1.0	Range	0.421–5.34	0.112–3.32
		Mean	1.86	1.27
		SEM	0.20	0.19
As (mg/l)	0.01-NR	Range	0.0001–0.02	0.0001–0.02
		Mean	0.004	0.002
		SEM	0.0007	0.00
Mn (mg/l)	0.1–0.3	Range	0.001–2.12	0.001–1.70
		Mean	0.94	0.68
		SEM	0.09	0.12

Table 3. Proximity matrix showing proximity distances between the 17 parameters for groundwater

Proximity matrix																	
Case	Fe	As	Mn	pH	TDS	EC	TA	F	SO ₄	NO ₃	Cl	TH	Ca	Mg	Zn	Pb	Cr
Fe	0																
As	48.737	0															
Mn	12.454	41.213	0														
pH	115.727	72.358	121.423	0													
TDS	75.783	46.161	60.917	68.475	0												
EC	75.504	44.776	59.733	68.814	0	0											
TA	85.220	51.415	76.714	48.051	22.151	22.207	0										
F	75.996	53.061	62.187	70.998	49.889	49.411	57.019	0									
SO ₄	88.591	84.026	84.088	57.872	52.818	53.422	53.253	75.271	0								
NO ₃	73.991	78.346	76.964	66.032	60.917	61.323	61.405	74.420	51.626	0							
Cl	68.857	84.361	58.829	82.260	30.092	30.681	62.516	41.395	45.733	56.894	0						
TH	78.042	87.249	71.061	75.991	45.990	46.737	58.174	81.380	43.831	61.548	41.661	0					
Ca	82.340	73.338	65.600	72.030	45.490	45.405	49.097	11.864	58.299	77.266	32.287	58.741	0				
Mg	67.822	71.744	63.571	76.840	51.442	51.498	71.656	80.230	58.840	58.126	54.106	11.472	71.411	0			
Zn	92.805	76.759	99.001	42.037	68.312	68.898	72.286	77.899	63.445	58.563	70.213	65.024	86.302	64.025	0		
Pb	39.480	58.318	36.380	92.461	48.861	48.149	79.177	47.164	74.319	66.743	29.819	78.832	53.053	70.985	76.059	0	
Cr	55.422	38.245	38.323	89.752	25.932	24.931	53.573	37.942	62.786	70.964	33.229	56.718	41.221	49.842	86.742	23.920	0

Table 4. Proximity matrix showing proximity distances between the 17 parameters for surface water

Proximity matrix																	
Case	Fe	As	Mn	pH	TDS	EC	TA	F	SO ₄	NO ₃	Cl	TH	Ca	Mg	Zn	Pb	Cr
Fe	0																
As	33.748	0															
Mn	1.629	32.305	0														
pH	121.578	98.838	123.613	0													
TDS	98.883	92.928	98.330	45.553	0												
EC	102.473	91.758	101.240	40.273	1.304	0											
TA	22.353	40.602	17.092	104.880	80.551	80.786	0										
F	61.705	76.952	63.715	64.721	69.301	67.848	58.838	0									
SO ₄	102.506	89.959	101.406	46.736	5.121	4.948	82.770	67.389	0								
NO ₃	100.258	89.464	101.280	39.667	6.192	5.079	87.198	68.019	7.010	0							
Cl	65.057	70.377	63.367	70.712	37.743	37.254	64.659	72.876	37.752	35.026	0						
TH	93.622	79.147	90.805	43.998	23.564	22.016	70.911	64.763	23.711	33.166	46.601	0					
Ca	99.998	85.765	99.990	32.579	20.134	19.173	81.156	67.786	26.231	25.988	50.914	29.058	0				
Mg	72.835	66.964	69.439	67.923	53.415	52.327	60.786	64.253	48.596	60.241	56.103	17.590	76.545	0			
Zn	72.992	65.524	69.713	66.438	37.610	35.037	49.642	70.799	28.873	30.088	45.540	39.298	46.068	51.235	0		
Pb	20.886	33.298	18.121	118.559	97.171	98.507	21.395	72.138	93.039	98.221	62.926	84.004	102.794	58.917	58.948	0	
Cr	45.080	69.049	48.696	96.834	75.021	79.886	64.286	77.177	76.826	83.442	62.455	73.208	73.740	69.618	68.330	40.914	0

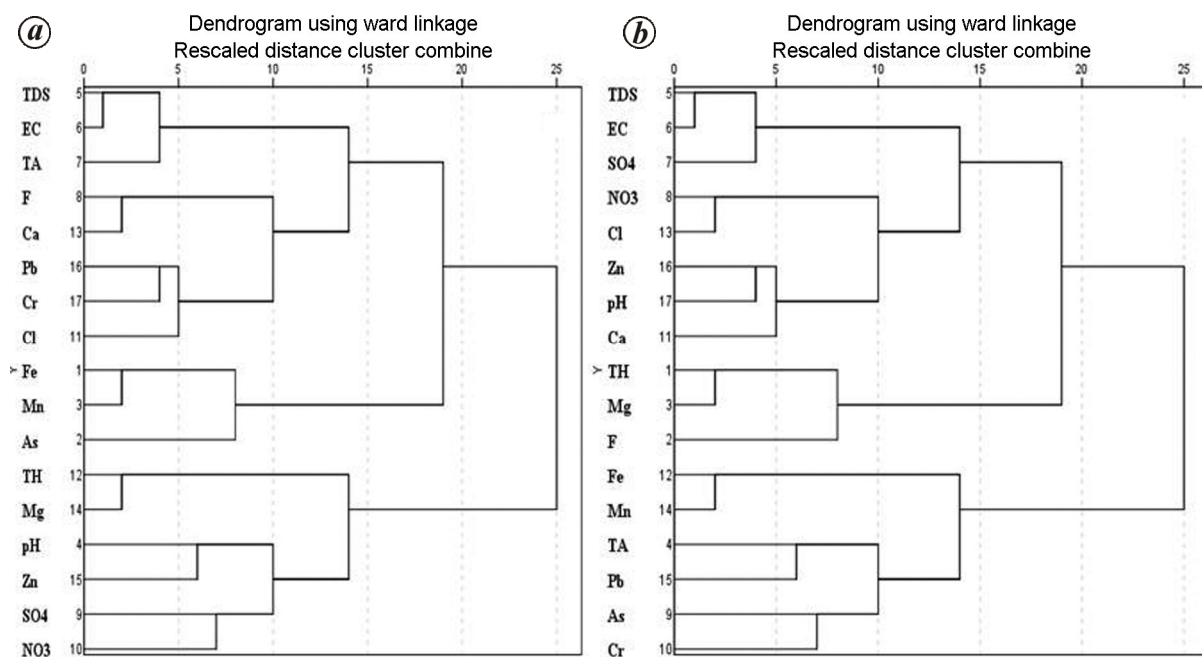


Figure 2. Dendrogram showing grouping of parameters in (a) groundwater (b) surface water of Balipara, North Brahmaputra river basin, Sonitpur district, Assam, India.

Table 5. Factor analysis of groundwater in PHC 1

Parameters	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
pH	0.112	-0.852	-0.059	-0.162	-0.171	-0.008
TDS	0.886	-0.017	0.195	0.222	0.201	0.150
EC	0.894	-0.007	0.195	0.215	0.204	0.135
TA	0.804	-0.181	0.128	-0.068	-0.358	0.176
TH	0.115	-0.016	0.056	0.940	-0.094	0.224
F	0.200	-0.029	0.812	-0.219	0.225	-0.191
SO ₄	0.146	-0.149	0.109	0.206	-0.134	0.695
NO ₃	0.096	-0.051	-0.274	-0.064	0.192	0.700
Cl	0.205	0.100	0.527	0.298	0.531	0.427
Ca	0.163	-0.005	0.947	0.039	0.008	0.044
Mg	0.109	0.050	-0.135	0.916	0.084	-0.017
Zn	-0.060	-0.679	-0.349	0.083	0.427	-0.020
Pb	0.139	0.397	0.230	-0.124	0.781	0.030
Cr	0.572	0.352	0.350	0.182	0.430	-0.097
Fe	-0.062	0.860	-0.229	-0.088	0.173	-0.146
As	0.621	0.234	-0.180	-0.246	0.042	-0.519
Mn	0.134	0.902	-0.005	0.003	0.168	-0.178
Eigen values	4.512	3.435	2.245	1.755	1.238	1.035
Percentage of variance	26.538	20.207	13.208	10.323	7.280	6.086
Cumulative variance	26.538	46.746	59.954	70.277	77.557	83.642

other anions ($r = 0.812-0.947$), indicating significant complexation of Ca and F in groundwater. The complexation of F with Ca may be due to the geochemical disposition influenced by relatively high temperature and association of Ca (ref. 43). Factor 4 exhibits 10% of the total variance with positive loading on Mg and TH ($r = 0.916-0.940$), showing Mg to be a major contributor to TH in the studied area³⁹. Pb and Cl were found to be positively loaded and Zn was moderately loaded with 7%

of the total variance in factor 5 ($r = 0.427-0.781$). SO₄ and NO₃ were found to be positively loaded with 6% of the total variance for factor 6 ($r = 0.695-0.700$). The maximum loading ($r = 0.700$) of NO₃ in this factor may be reiterating the association of anthropogenic activities. This is also supported by the use of NPK fertilizers, manure, etc. in the adjoining tea gardens⁴². Figure 3a shows the loading exhibited by first two principal components (PC1 and PC2), while Figure 4a represents the

Table 6. Factor analysis of surface water in PHC 1

Parameters	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
pH	0.173	-0.803	0.063	-0.337	0.074
TDS	0.904	-0.289	0.131	0.030	0.058
EC	0.906	-0.312	0.155	-0.056	0.058
TA	0.037	0.852	0.076	-0.165	0.160
TH	0.559	-0.193	0.732	-0.055	0.058
F	-0.030	0.065	0.046	-0.111	0.940
SO ₄	0.880	-0.277	0.218	-0.012	0.022
NO ₃	0.882	-0.314	0.019	-0.091	0.021
Cl	0.633	0.188	0.054	0.222	-0.087
Ca	0.707	-0.368	-0.116	-0.104	0.058
Mg	0.097	0.067	0.980	0.019	0.022
Zn	0.662	0.251	0.223	-0.148	-0.109
Pb	-0.203	0.832	0.134	0.207	-0.145
Cr	-0.041	0.224	-0.007	0.882	-0.109
Fe	-0.237	0.877	-0.145	0.181	0.098
As	-0.177	0.634	0.010	-0.364	-0.370
Mn	-0.217	0.921	-0.092	0.126	0.068
Eigen values	7.129	3.211	1.370	1.185	1.032
Percentage of variance	41.936	18.889	8.058	6.972	6.069
Cumulative variance	41.936	60.825	68.883	75.854	81.924

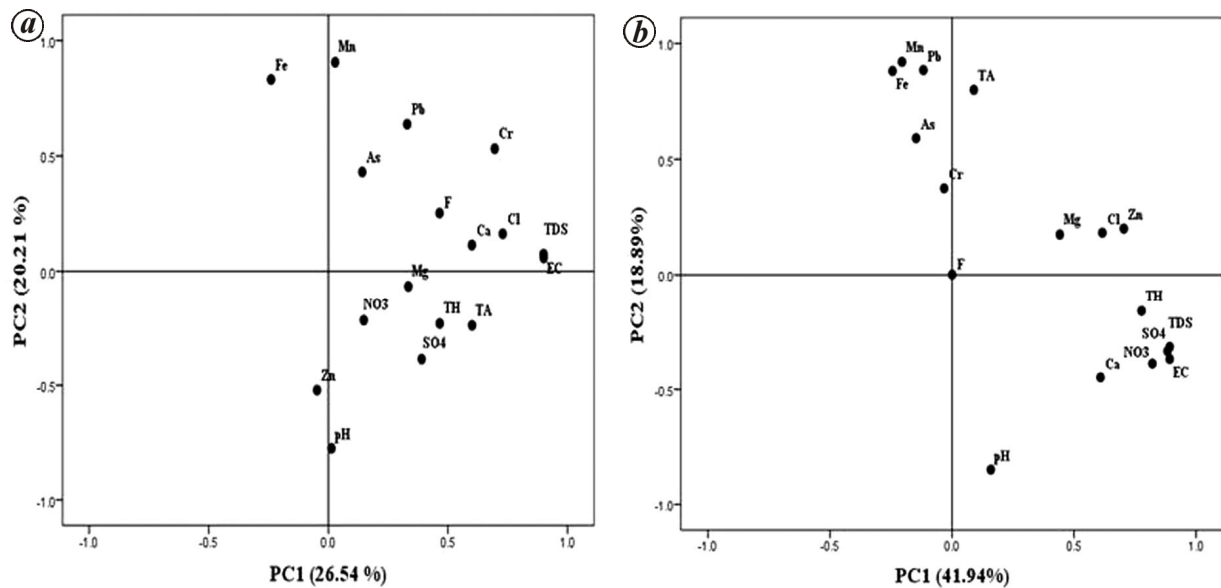


Figure 3. Loading of the first two principal components (PC1 and PC2) explaining (a) 46.75% of the total variance in groundwater and (b) 60.83% of the total variance in surface water of Balikpapan.

screen plot showing the variation of eigen values with component number in groundwater of Balikpapan.

For surface water, factor 1 contributes 41% of the total variance of 81% with positive loading on TDS, EC, SO₄ and NO₃. Moderate loading was observed for Ca, Zn, Cl and TH. Mn, Fe, TA and Pb were found to be positively loaded, As moderately loaded and pH negatively loaded with 18.8% of the total variance for factor 2. Factor 3 contributes 8% of the total variance with positive loading on Mg. Factor 4 is found to be positively loaded with Cr contributing 6% of the total variance. Factor 5 also con-

tributes 6% with positive loading on F. Figure 3 b shows the loading exhibited by the first two PCs (PC1 and PC2), while Figure 4 b represents the scree plot showing the variation of Eigen values with component number in surface water of Balikpapan.

DA was used to evaluate the variation in water quality parameters. It was applied to raw data consisting of 17 parameters. Table 7 shows that only one DF discriminates between the ground and surface water. Wilk's lambda test shows that DF is statistically significant (Table 8). Furthermore, 100% of the total variance between ground and

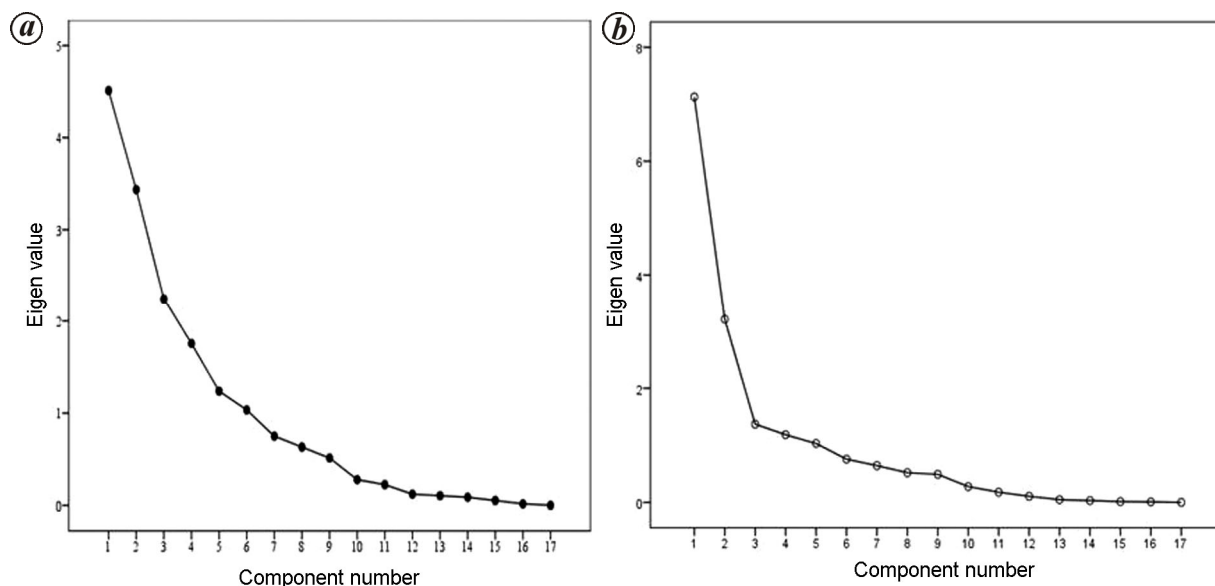


Figure 4. Scree plot showing variation of Eigen values with the component number in (a) groundwater and (b) surface water of Balipara.

Table 7. Eigen-value of DF for groundwater and surface water

Function	Eigen value	Percentage of variance	Cumulative Percentage
1	3.356 ^a	100.0	100.0

^a98.6% of original grouped cases correctly classified.

Table 8. Wilk's Lamda for testing discriminant function validity

Test of function(s)	Wilks' lambda	Chi-square	P-value
1	0.230	87.553	0.000

Table 9. Classification results for discriminate analysis of the ground and surface water^a

Water	% Correct	Predicted group	
		Groundwater	Surface water
Groundwater	100	97.1	2.9
Surface water	100	0	100

^a98.6% of original grouped cases correctly classified.

surface water is explained by only one DF. The relative contribution of each parameter is given in eq (4)

$$\begin{aligned}
 Z = & 0.439 \text{ Fe} - 0.074 \text{ As} + 0.217 \text{ Mn} \\
 & + 0.999 \text{ pH} + 0.189 \text{ TDS} + 0.420 \text{ EC} - 0.017 \text{ TA} \\
 & - 0.359 \text{ F} + 1.166 \text{ SO}_4 - 1.254 \text{ NO}_3 - 0.117 \\
 & \text{Cl} - 0.395 \text{ TH} + 0.227 \text{ Ca} + 0.098 \text{ Mg} \\
 & - 0.504 \text{ Zn} + 0.823 \text{ Pb} - 0.463 \text{ Cr}.
 \end{aligned}
 \tag{4}$$

It can be seen that nitrate and sulphate exhibit strong contribution in discriminating between the ground and sur-

face water and account for most of the expected variation in water quality, with moderate contribution from pH, lead and zinc, and meagre contribution by other parameters in explaining the variation. Based on these findings, the order of the relative contribution for water quality parameters is as follows:

$$\text{NO}_3 > \text{SO}_4 > \text{pH} > \text{Pb} > \text{Zn} > \text{Cr} > \text{Fe} > \text{EC} > \text{TH} > \text{F} > \text{Ca} > \text{Mn} > \text{TDS} > \text{Cl} > \text{Mg} > \text{As} > \text{TA}.$$

The classification matrix (Table 9) shows that more than 98.6% of the cases are correctly classified in their respective groups. The results which are classified in terms of one DF show that significant differences exist between the ground and surface water.

In conclusion, various multivariate statistical techniques such as CA, FA, PCA and DA were applied to assess the variation of ground and surface water quality at Balipara. Hierarchical clustering grouped the ground and surface water each into three clusters of similar water quality characteristics. FA/PCA analysis extracted six factors in groundwater explaining more than 83% of the total variance, while more than 81% of the total variance in surface water was explained by five factors. Analysis revealed that TDS, EC, TA, As and Cr were the main contaminants in groundwater with the loading of 26% in factor 1. Factor 2 contributes 20% of the total variance with major contribution of Fe and Mn. In case of surface water, factor 1 contributes 41% of the total variance with positive loadings on TDS, EC, SO₄ and NO₃, Ca, Zn, Cl and TH. In factor 2, Mn, Fe, TA and Pb are the main contributors to the loading value of 18.8% of the total variance. This study helps identifying the factors responsible for variation in water quality parameters. Hydrogeochemistry,

sedimentology, lithology and plausible correlation between the water quality parameters are likely to be the controlling factors responsible for their variation. The best result to identify the relative contribution of water quality parameters was provided by DA which afforded 100% correct assignments in discriminating the ground and surface water. It was observed that nitrate and sulphate exhibited strong contribution in discriminating the ground and surface water. This accounted for most of the expected variation in water quality; moderate contribution from pH, lead and zinc was observed, while meagre contribution was shown by other parameters to explain the variation. The prolonged consumption of high nitrate content in the drinking water at Balipara may cause blue-baby syndrome in the near future. This calls for a toxicological study in the locality to prevent the disease.

- Carpenter, S. R., Caraco, N. F., Correll, D. L., Howarth, R. W., Sharpely, A. N. and Smith, V. H., Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecol. Appl.*, 1998, **8**, 559–568.
- Fergusson, J. E. (ed.), The heavy elements: chemistry. In *Environmental Impact and Health Effects*, Pergamon, New York, 1990, p. 720.
- Niemi, G. J., Devore, P., Detenbeck, N., Taylor, D. and Lima, A., Overview of case studies on recovery of aquatic systems from disturbance. *Environ. Manage.*, 1990, **14**, 571–587.
- Pruss-Ustun, A., Bos, R., Gore, F. and Bartram, J., Safer water, better health, 2008; http://www.who.int/quantifying_chimpacts/publications/saferwater/en
- United Nations Environment Programme, Vital water graphics: problems related to freshwater resources. 2002; <http://www.unep.org/dewa/assessments/ecosystems/water/vitalwater/20.html>
- Owal, A., A. F. S. A., Design of iron eliminator for hand tube well. *J. Indian water Works Assoc.*, 1981, **13**, 65.
- Das, H. B., Goswami, D. C., Borah, K., Kalita, H. and Saikia, L. B., Water quality of Northeastern region. Tech. Bull. No. DRDO-DRLT-TB-1-1996, 1996, pp. 1–24.
- Vega, M., Pardo, R., Barrado, E. and Deban, L., Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Res.*, 1998, **32**, 3581–3592.
- Mustapha, A. and Aris, A. Z., Spatial aspects of surface water quality in the Jakara Basin, Nigeria using chemometric analysis. *J. Environ. Sci. Health Part A*, 2012, **47**, 1455–1465.
- Adams, S., Titus, R., Pietesen, K., Tredoux, G. and Harris, C., Hydrochemical characteristics of aquifers near Sutherland in the Western Karoo, South Africa. *J. Hydrol.*, 2001, **241**, 91–103.
- Wunderlin, D. A., Diaz, M. P., Ame, M. V., Pesce, S. F., Hued, A. C. and Bistoni, M. A., Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. a case study: Suquia river basin (Cordoba, Argentina). *Water Res.*, 2001, **35**, 2881–2894.
- Reghunath, R., Murthy, T. R. S. and Raghavan, B. R., The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. *Water Res.*, 2002, **36**, 2437–2442.
- Simeonova, P., Simeonov, V. and Andreev, G., Environmetric analysis of the Struma River water quality. *Cent. Eur. J. Chem.*, 2003, **4**, 121–126.
- Shirodkar, P. V., Mesquita, A., Pradhan, U. K., Verlekar, X. N., Babu, M. T. and Vethamony, P., Factors controlling physico-chemical characteristics in the coastal waters off Mangalore – a multivariate approach. *Environ Res.*, 2009, **109**, 245–257.
- 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*, 2005, **538**, 355–374.
- Helena, B., Pardo, R., Vega, M., Barrado, E., Fernández, J. M. and Fernández, L., Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis. *Water Res.*, 2000, **34**, 807–816.
- Singh, K. P., Malik, A., Mohan, D. and Sinha, S., Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India): a case study. *Water Res.*, 2004, **38**, 3980–3992.
- Subramanian, S. and Balasubramanian, A., Hydrogeological studies of Tiruchendur coast, Tamil Nadu, India. In *Regional Workshop on Environmental Aspects of Groundwater Development*, 17–19 October 1994, pp. 26–32.
- Lucho-Constantino, C. A., Álvarez-Suárez, M., Beltrán-Hernández, R. I., Prieto-García, F. and Poggi-Varaldo, H. M., A multivariate analysis of the accumulation and fractionation of major and trace elements in agricultural soils in Hidalgo State, Mexico irrigated with raw wastewater. *Environ. Int.*, 2005, **31**, 313–323.
- APHA, *Standard Methods for Examinations of Water and Wastewater*, American Public Health Association, Washington, DC, 2005.
- Johnson, R. A. and Wichern, D. W., *Applied Multivariate Statistical Analysis*, Pearson Prentice Hall International, USA, 2007, 6th edn.
- Zhou, F., Liu, Y. and Guo, H., Application of multivariate statistical methods to the water quality assessment of the watercourses in Northwestern New Territories, Hong Kong. *Environ. Monit. Assess.*, 2007, **132**, 1–13.
- Kumari, M., Tripathi, S. and Pathak, V., Chemometric characterization of river water quality. *Environ. Monit. Assess.*, 2013, **185**, 3081–3092.
- McKenna Jr, J. E., An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. *Environ. Model. Softw.*, 2003, **18**, 205–220.
- Otto, M., *Chemometrics: Statistics and Computer Application in Analytical Chemistry*, Wiley, Germany, 2007.
- Kazi, T. G. *et al.*, Assessment of water quality of polluted lake using multivariate statistical techniques: a case study. *Ecotoxicol. Environ. Saf.*, 2009, **72**, 301–309.
- Krishna, A. K., Satyanarayanan, M. and Govil, P. K., Assessment of heavy metal pollution in water using multivariate statistical techniques in an industrial area: a case study from Patancheru, Medak District, Andhra Pradesh, India. *J. Hazard. Mater.*, 2009, **167**, 366–373.
- Shrestha, S. and Kazama, F., Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. *Environ. Model. Softw.*, 2007, **22**, 464–475.
- Brumelis, G., Lapina, L., Nikodemus, O. and Tabors, G., Use of an artificial model of monitoring data to aid interpretation of principal component analysis. *Environ. Model. Softw.*, 2000, **15**, 755–763.
- Love, D., Hallbauer, D. K., Amos, A. and Hranova, R. K., Factor analysis as a tool in groundwater quality management: two southern African case studies. *Phys. Chem. Earth*, 2004, **29**, 1135–1143.
- Abdul-Wahab, S. A., Bakheit, C. S. and Al-Alawi, S. M., Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environ. Model. Softw.*, 2005, **20**, 1263–1271.
- Alkarkh, A. F. M., Assessment of surface water through multivariate analysis. *J. Sustain. Dev.*, 2008, **1**, 27–33.
- Johnson, R. A. and Wichern, D. W., *Applied Multivariate Statistical Analysis*, Prentice-Hall, London, 2002.

RESEARCH COMMUNICATIONS

34. Alvin, C. R., *Methods of Multivariate Analysis*, John Wiley, USA, 2002.
35. Johnson, R. A. and Wichern, D. W., *Applied Multivariate Statistical Analysis*, Prentice-Hall, NJ, 1992.
36. Krug, E. C. and Frink, C. R., Acid rain on acid soil: a new perspective. *Science*, 2010, **221**, 520–525.
37. Banerjee, S. *et al.*, Heavy metal contaminants of underground water in Indo Bangla border districts of Tripura, India. *Int. J. ChemTech. Res.*, 2011, **3**, 516–522.
38. Post, J. E., Manganese oxide minerals: crystal structures and economic and environmental significance. *Proc. Natl. Acad. Sci., USA*, 1999, **96**, 3447–3454.
39. Guidelines for Drinking Water Quality, WHO, World Health Organisation, 2011, 4th edn; www.who.int/water_sanitation_health/publications/2011/...guidelines/en/
40. BIS, Drinking water specification IS: 10500, Bureau of Indian standards, 2012, pp. 1–11; <https://law.resource.org/pub/in/bis/S06/is.10500.2012.pdf>
41. Karak, T., Abollino, O., Bhattacharyya, P., Das, K. K. and Paul, R. K., Fractionation and speciation of arsenic in three tea gardens soil profiles and distribution of As in different parts of tea plant (*Camellia sinensis* L.), *Chemosphere*, 2011, **85**, 948–960.
42. Ishaku, J. M., Kaigama, U. and Onyeka, N. R., Assessment of groundwater quality using factor analysis in Mararaba-mubi area, northeastern Nigeria. *J. Earth Sci. Geotech. Eng.*, 2011, **1**, 9–33.
43. Kotoky, P., Geoenvironmental perspectives. *Sci. Cult.*, 2011, **77**, 474–479.

ACKNOWLEDGEMENTS. We thank DRDO Headquarters, New Delhi, for providing financial support and Mrs Kavita Yadav (Defence Research Laboratory, Tezpur) for mapping of the study area.

Received 18 July 2015; revised accepted 8 October 2015

doi: 10.18520/cs/v110/i7/1350-1360
