Determination of processes affecting groundwater quality in the coastal aquifer beneath Puri city, India: a multivariate statistical approach

P. K. Mohapatra, R. Vijay, P. R. Pujari, S. K. Sundaray and B. P. Mohanty

ABSTRACT

Variability of groundwater quality parameters is linked to various processes such as weathering, organic matter degradation, aerobic respiration, iron reduction, mineral dissolution and precipitation, cation exchange and mixing of salt water with fresh water. Multivariate statistical analyses such as principal component analysis (PCA) and hierarchical cluster analysis (HCA) were applied to the standardized data set of eleven groundwater quality parameters (i.e. pH, Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, Fe^{3+}, alkalinity, NO_3^{-}, Cl^{-}, SO_4^{2-}, TDS) collected during the post-monsoon and the summer seasons in order to elicit hydrologic and biogeochemical processes affecting water quality in the unconfined aquifer beneath Puri city in eastern India. The application of PCA resulted in four factors explaining 73% variance in post-monsoon and 81% variance in summer. The HCA using Ward's method and squared Euclidean distance measure classified the parameters into four clusters based on their similarities. PCA and HCA allowed interpretation of processes. During both post-monsoon and summer seasons, anthropogenic pollution and organic matter degradation/Fe(III) reduction were found dominant due to contribution from on-site sanitation in septic tanks and soak pits in the city. Cation exchange and mineral precipitation were possible causes for increase in Na^{+} and decrease in Ca^{2+} concentration in summer. Fresh water recharge during monsoon and Sea water intrusion in summer are attributed as significant hydrologic processes to variations of the groundwater quality at the study site.

Key words | coastal aquifer, hierarchical cluster analysis, multivariate statistical analysis, principal component analysis, Puri city

P. K. Mohapatra (corresponding author)
Orissa Water Supply and Sewerage Board,
Water Works Road,
Puri - 752002,
Orissa,
India
E-mail: *prasant_mohapatra@hotmail.com*

R. Vijay
P. R. Pujari
National Environmental Engineering Research Institute,
Nagpur - 440020,
India

S. K. Sundaray
Institute of Oceanography,
National Taiwan University,
Taipei - 10617,
Taiwan

B. P. Mohanty
Department of Biological and Agricultural Engineering,
2117 TAMU,
Texas A&M University,
College Station,
Texas,
USA

INTRODUCTION

Groundwater quality parameters are controlled by many factors such as rainfall, composition of aquifer material, topography, hydrologic fluctuation and climate. These factors interact in a complex way and result in spatial and temporal variation in water quality parameters. Determination of processes affecting groundwater quality in a coastal aquifer is very complex. The variability of the parameters is linked to various biological, physical and chemical processes taking place in the aquifer such as: weathering, organic matter degradation, aerobic respiration, iron reduction, mineral dissolution and precipitation, cation exchange and mixing of salt water with fresh water. Due to simultaneous effects of multiple hydrologic and biogeochemical processes, it becomes difficult to interpret relationships among the water quality parameters and the governing biogeochemical processes from graphical techniques (Guler *et al.* 2002). Multivariate statistical analyses such as principal component analysis (PCA) and hierarchical cluster analysis (HCA) have been used to provide a quantitative measure of relatedness of water quality parameters and to suggest the underlying natural and anthropogenic processes in groundwater aquifers. Recent studies have confirmed the usefulness of multivariate analysis techniques for (i) evaluation and interpretation of groundwater quality data sets (Singh *et al.* 2004, 2009), (ii) providing insight into the processes (Ruiz *et al.* 1990; Jayakumar & Siraz 1997; Papatheodorou *et al.* 2007; Báez-Cazull *et al.* 2008; Machado *et al.* 2008), (iii) identifying critical water

doi: 10.2166/wst.2011.605

quality issues and possible sources of pollution/polluting processes (Singh *et al.* 2005; Kumar & Riyazuddin 2008; Sargaonkar *et al.* 2008; Rao *et al.* 2010) and (iv) interaction of river water/groundwater and groundwater mixing (Reghunath *et al.* 2002).

In this paper, PCA and HCA were applied to the physico-chemical properties of groundwater collected during post-monsoon and summer seasons to elicit dominant processes affecting groundwater quality in the coastal unconfined aquifer of Puri city in India.

MATERIALS AND METHODS

Study area

The study area is a 16 km^2 urban catchment (Figure 1). Two groundwater well fields were located on the eastern and western boundaries of the Puri city for domestic water supply (CDP 2006). The city is located on the coast of Bay of Bengal in Orissa state in India. The aquifer beneath the city is sandy and unconfined up to 40 m depth and lies above a thick clay layer. Thin clay layers are also found as isolated lenses at shallow depths within the unconfined sandy aquifer. The area receives rainfall from southwest monsoon from second week of June to October (IMD 2008). After the monsoon season the groundwater table in the area gradually lowers and reaches lowest level in the first week of June. The water table fluctuation, i.e., the difference in maximum and minimum water level was observed to be 3 m during the sampling year 2006 and 2007 (NEERI 2008).

Sampling and data analysis

The water quality data were obtained from two sampling seasons. In the post-monsoon season during November 2006 groundwater samples were collected from 51 tube wells/production wells across the city. In the second sampling season during summer (first week of June 2007) water samples were collected from 43 wells. The water samples were analysed for phyico-chemical parameters following Standard Methods (NEERI 2008). The eleven water quality parameters subjected to statistical analysis are pH, Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, Fe^{3+}, alkalinity, NO_3^{-}, Cl^{-}, SO_4^{2-} and total dissolved solid (TDS). Other geochemical concentrations (fluoride and phosphate) were found to be too low during the water sampling, and thus were not used for further analyses. While this work is focused on dominant hydrologic and geochemical processes, microbiological water quality data were reported separately in Vijay *et al.* (2011). The variables (water quality parameters) were standardized using z-score: $z = (y_i - \hat{y})/s$, where '$\hat{y}$' is the average value of a parameter in a data set and 's' is its standard deviation to avoid the problem of difference in scale, i.e., range of values and variances. The PCA and HCA were carried out on the standardized data sets. The software SPSS Statistics 17.0 was used for data standardization, PCA and HCA.

Principal component analysis

The method of principal components is a special case of the more general method of Factor Analysis. The aim of PCA is

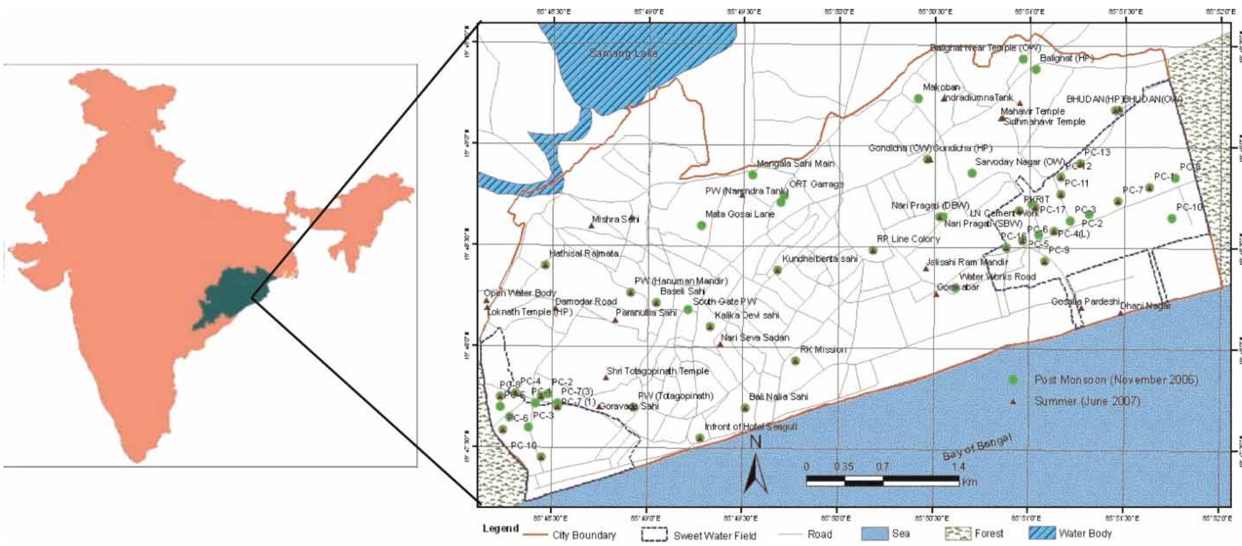

Figure 1 | Study area.

construction of new variables called principal components out of a set of existing original variables. The new variables are a linear combination of the existing variables. The PCA is performed to reduce the large data set of variables into few factors called the principal components which can be interpreted to reveal underlying data structure. The characteristics of principal components are that they are uncorrelated, i.e., orthogonal. The first principal component (F_1) absorbs and accounts for the maximum possible proportion of the total variance in the data set and the second component (F_2) absorbs the maximum of the remaining variance and so on. The maximum number of principal components is equal to the number of variables. The total variance accounted by all the F_i's will be equal to the number of variables because each standardized variable has variance equal to 1. For interpretation, only a few numbers of F_i are retained in the analysis. The number of principal components to be retained in the analysis is based on Kaiser criterion. As per this criterion, principal components (F_i) having latent root or Eigen value (denoted by λi) greater than one are considered essential and are retained in the analysis. The latent root is also expressed as percentage of total variance in the data set. For example, the percentage contribution of F_i in the total variance in the standardized data set is defined by the following expression:

$$\text{Percentage variance accounted by } F_i = \frac{\lambda i}{\sum \lambda i} \times 100$$

The sum of latent roots of all the principal components is equal to the number of variables in the data set, i.e., $\Sigma\lambda i =$ number of variables. The values of latent roots become smaller for subsequent F_i's because the principal component procedure extracts the maximum possible variance for each previous F's.

The factor, F_i, is a linear combination of the original variables given by:

$$F_1 = a_{11}Z_1 + a_{12}Z_2 + \cdots + a_{1k}Z_k$$
$$F_2 = a_{21}Z_1 + a_{22}Z_2 + \cdots + a_{2k}Z_k$$
$$\cdots$$
$$\cdots$$
$$F_k = a_{k1}Z_1 + a_{k2}Z_2 + \cdots + a_{kk}Z_k$$

The latent root, λ_i, is given by:

$$\lambda_1 = (a_{11})^2 + (a_{12})^2 + \cdots + (a_{1k})^2$$
$$\lambda_2 = (a_{2_1})^2 + (a_{2_2})^2 + \cdots + (a_{2_k})^2$$
$$\cdots$$
$$\cdots$$
$$\lambda i = (a_{i1})^2 + (a_{i2})^2 + \cdots + (a_{ik})^2$$

In the above expressions, a's are called factor loadings, Z's are the standardized values of the original variable and 'k' is the number of variables. High factor loadings (value of a close to ±1) indicate strong relationship (positive or negative) between the variable and the factor. The factor loadings matrix is rotated using varimax orthogonal rotation to maximize the relationship between the variables and some of the factors. This rotation results in high factor loadings for the variables correlated in the factor and low loadings for the remaining variables. In the PCA, it is required to determine the values of the factor loadings, i.e., the value of a's from the standardized data set.

The suitability of the data set for PCA is tested by Kaiser–Meyer–Olkin (KMO) and Bartlett's tests. KMO is a measure of sampling adequacy. A value of KMO that is >0.5 indicates PCA can be performed. The Bartlett's test of sphericity tests the null hypothesis that the variables in the population correlation matrix are uncorrelated and the correlation matrix is an identity matrix. If the observed significance level is <0.05 then the null hypothesis is rejected. It indicates that there are significant relationships among variables. In this study PCA was carried out on the standardized data sets and sorted by using Eigen values greater than 1.0 as these are considered significant influences towards the hydro-geochemical processes (Sahu *et al.* 1998; Panigrahy *et al.* 1999; Sundaray *et al.* 2006; Sundaray 2010).

Hierarchical cluster analysis

HCA is useful to group water quality parameters into clusters so that parameters within a cluster are similar to each other but different from those in other clusters. HCA is an unsupervised pattern recognition technique and uncovers intrinsic structure or underlying pattern in a data set without making a priori assumption about the data in order to classify the parameters into clusters based on their similarities. In HCA, clusters are formed sequentially, starting with the most similar pair of variables and forming higher clusters step by step. The process of cluster formation is repeated until a single cluster containing all the variables are obtained. The result of clustering is a Dendrogram representing the nested grouping of patterns and similarity levels at which groupings change. The dendrogram can be broken at different levels to yield different clusters of the data set. It is most common to calculate the dissimilarity between two patterns using a distance measure. The most popular is the Euclidean distance. The HCA with Ward's method of linkages with squared Euclidean distance as dissimilarity measure was applied to detect multivariate similarities and to group parameters into clusters based on their similarities. The Ward's method of linkage uses the minimum variance approach to evaluate distance between clusters (Jain *et al.* 1999). The Ward's method with squared Euclidean distance as dissimilarity measure has been found to provide meaningful Dendrogram of clusters with the proximity or similarity of clusters measured with a rescaled distance. HCA has been successfully used in earlier studies for evaluation and interpretation of groundwater quality data set (Kumar & Riyazuddin 2008; Rani & Babu 2008). In this study, HCA was performed with standardized data set to eliminate the effect of scale of measurement of data.

RESULTS AND DISCUSSION

Hydro-geochemistry of major ions

The physico-chemical analysis of groundwater samples in the two sampling seasons indicates that the dominant major cations are Mg^{2+}, Ca^{2+}, Na^{+} and K^{+} and the dominant anions are HCO_3^{-}, Cl^{-} and SO_4^{2-}. The concentrations of water quality parameter vary in the two sampling seasons (NEERI 2008; Vijay *et al.* 2011). The Piper trilinear diagrams of post-monsoon and summer data sets (Figures 2(a) and (b)) show that in post-monsoon, major facies are: (i) Mg^{2+}–Ca^{2+}–HCO_3^{-}–Cl^{-} type followed by (ii) Mg^{2+}–Ca^{2+}–Na^{+}–K^{+}–HCO_3^{-}–Cl^{-} type, (iii) Mg^{2+}–Ca^{2+}–HCO_3^{-} type and (iv) Mg^{2+}–Ca^{2+}–Na^{+}–K^{+}–Cl^{-} type. In summer, the major facies are: (i) Mg^{2+}–Ca^{2+}–Na^{+}–K^{+}–HCO_3^{-}–Cl^{-} type, followed by (ii) Na^{+}–K^{+}–HCO_3^{-}–Cl^{-} type, (iii) Mg^{2+}–Ca^{2+}–Na^{+}–K^{+}–HCO_3^{-} type and (iv) Mg^{2+}–Ca^{2+}–Na^{+}–K^{+}–Cl^{-} type.

In post-monsoon, the order of median concentration of cations is $Mg^{2+} > Ca^{2+} > Na^{+} > K^{+}$. In summer, the order of median concentration of cations is $Na^{+} > Mg^{2+} > Ca^{2+} > K^{+}$. In case of anions concentration, the order of median concentration is $HCO_3^{-} > Cl^{-} > SO_4^{2-}$ in both sampling data sets.

The cations experience major change in concentrations in two sampling seasons (post-monsoon and summer). The concentrations of Ca^{2+} and Mg^{2+} decreased in summer compared to their post-monsoon concentrations. The concentration of Na^{+} increased in summer. The median values of concentrations of major cations Ca^{2+}, Mg^{2+}, Na^{+} and K^{+} in post-monsoon and summer are shown in Figure 3.

The concentration of these parameters appears to be affected by several geochemical processes in the aquifer. The aquifer is recharged by monsoon rain each year. Freshening is a predominant process and during this period groundwater regains some of the previous year background conditions. Intrusion of Sea water can occur in summer under favourable hydraulic condition as the groundwater table lowers. Sea water intrusion and groundwater recharge from the land surface would be occurring in a cyclic manner due to change in seasons each year.

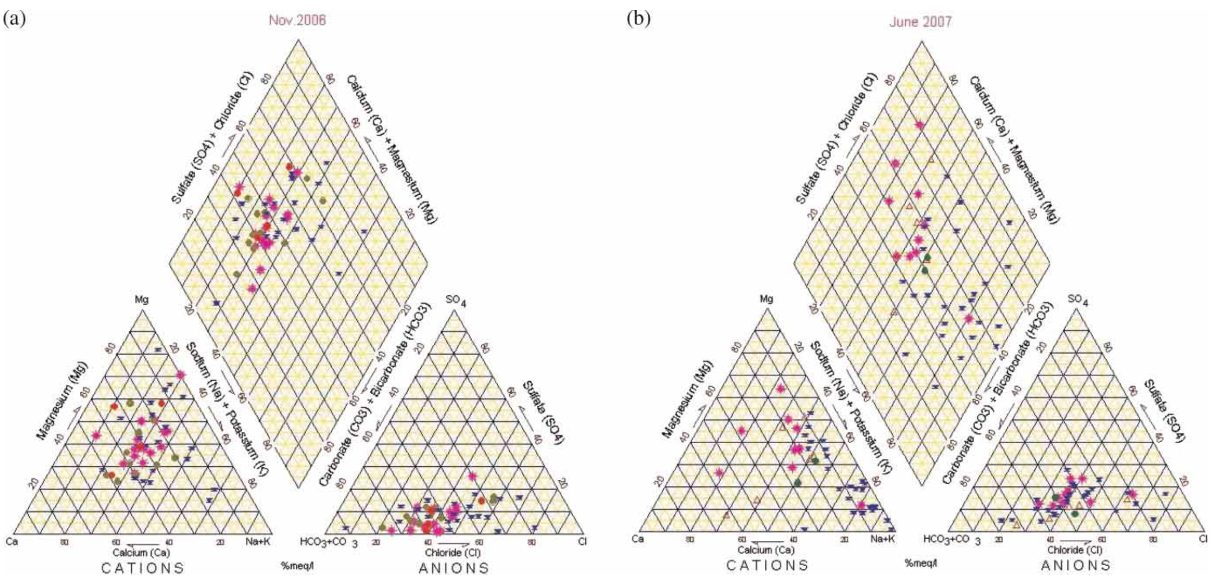

Figure 2 | Piper trilinear diagram of major ions in (a) Post-monsoon and (b) summer.

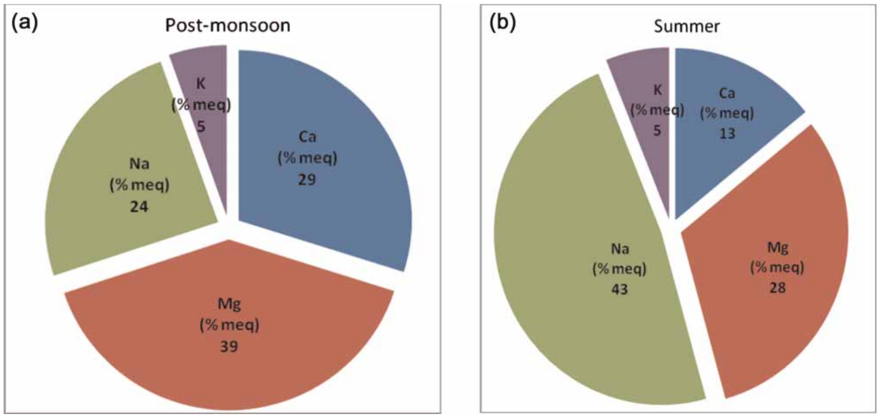

Figure 3 | Median concentration of cations (% milli-equivalent) in (a) groundwater samples in post-monsoon and (b) groundwater samples in summer.

Principal component analysis

In the post-monsoon data set, the KMO measure of sample adequacy is 0.727 and the significance level of Bartlett's test of sphericity is <0.001. In the summer data set, the KMO measure of sampling adequacy is 0.7 and the significance level of Bartlett's test of sphericity is <0.001. The results indicate that PCA can be performed on the two data sets. The PCA was performed on the standardized data set of eleven water quality parameters. In the post-monsoon data set, four factors were retained having Eigen values greater than unity and together they account for 73% of the variability of the data set. In the summer data set, four factors were retained having Eigen values greater than unity and together they account for 81% of the variability of the data set. The factor loadings obtained after varimax orthogonal rotation from the data set of post-monsoon and summer seasons are given in Tables 1 and 2, respectively.

The factor loadings include both positive and negative loadings. Loadings close to ± 1 indicate a strong correlation between a variable and the factor. Loadings higher than ± 0.75 are considered strong correlation, loadings between

Table 1 | Varimax orthogonal rotated factor loadings from PCA of standardized water quality data set (post-monsoon)

Variable	Factor 1	Factor 2	Factor 3	Factor 4
TDS	**0.841**	0.441	0.08	0.028
SO_4^{2-}	**0.845**	0.284	−0.104	0.039
Cl^-	**0.928**	0.146	0.060	0.066
Alkalinity	0.306	**0.855**	−0.089	0.132
K^+	0.302	**0.794**	0.093	0.145
Mg^{2+}	0.156	**0.765**	0.148	−0.039
Na^+	0.295	**−0.526**	0.581	0.099
Ca^{2+}	0.183	0.361	**0.534**	0.059
NO_3^-	0.250	−0.029	**− 0.733**	0.117
Fe^{3+}	−0.269	−0.095	0.156	**− 0.809**
pH	−0.464	0.040	0.185	**0.606**
Eigen value	2.963	2.662	1.278	1.095
% Variance	26.94	24.2	11.62	9.95
Cumulative % variance	26.94	51.14	62.76	72.71
Interpretation of process	Dilution of groundwater	Mineral dissolution	Weathering and anthropogenic pollution	Organic matter degradation and iron reduction

Note: significant factor loadings shown in bold.

Table 2 | Varimax orthogonal rotated factor loadings from PCA of standardized water quality data set (summer)

Variable	Factor 1	Factor 2	Factor 3	Factor 4
TDS	**0.833**	0.369	0.298	0.150
Cl^-	**0.850**	0.120	0.330	0.013
$SO_4{}^{2-}$	**0.858**	0.250	0.322	0.013
Mg^{2+}	**0.841**	0.267	0.092	0.194
$NO_3{}^-$	**0.741**	0.130	−0.316	0.049
Na^+	0.258	**0.848**	0.035	−0.049
K^+	0.239	**0.878**	0.117	0.011
Ca^{2+}	0.562	0.127	**0.685**	−0.043
pH	−0.073	−0.066	**− 0.866**	−0.039
Fe^{3+}	0.040	−0.197	−0.007	**0.868**
Alkalinity	0.193	0.482	0.045	**0.682**
Eigen value	3.890	2.082	1.655	1.287
% Variance	35.36	18.93	15.05	11.7
Cumulative % variance	35.36	54.29	69.34	81.04
Interpretation of process	Mixing of saline and fresh groundwater and anthropogenic pollution	Cation exchange	Mineral precipitation	Organic matter degradation and iron reduction

Note: significant factor loadings shown in bold.

±0.5 and ±0.74 are considered moderately correlated and loadings approaching 0 indicate weak correlations (Liu *et al.* 2003). Based on the significant factor loadings (greater than ±0.5), each factor is assigned a process which the significant variables are likely to be associated within the factor. The processes which have been interpreted from the factor loadings are also provided in Tables 1 and 2.

During post-monsoon season, Factor 1 is strongly loaded with TDS, SO_4^{2-} and Cl^- which explains 26.9% of the variability in the data set. The process assigned to this factor is dilution of groundwater since the concentrations of chloride and sulphate are very much reduced due to recharge effect of rain water as compared to the concentration of these parameters in summer samples. Factor 2 explains 24.2% of the variability and is highly correlated with alkalinity, K^+, Na^+ and Mg^{2+}. The loading of Na^+ is moderate and negative reflecting the reduction of Na^+ concentration due to the recharge process. This factor is interpreted as mineral dissolution factor. K^+ and Mg^{2+} are released to groundwater due to dissolution of minerals bearing these ions during recharge of aquifer by rainfall. Factor 3 accounts for about 11.6% of the variance of the data set and includes Ca^{2+} and NO_3^- which are moderately correlated to the factor. The sign for NO_3^- is negative suggesting that its concentration is reduced. Two distinct processes are suggested, weathering and anthropogenic pollution. The Ca^{2+} concentration is increased due to weathering or dissolution of calcite minerals such as carbonate shell which is abundant in the study area. The other contributing process is suggested as anthropogenic pollution due to presence of NO_3^- in this factor. The source of nitrate may be from on-site sanitation, livestock waste and municipal landfill sites in the study area. Due to dilution, concentration of NO_3^- is reduced. Factor 4 accounts for 9.9% of variability and has loadings of Fe^{3+} and pH. The loading of Fe is strong and negative and the loading of pH is moderate and positive. The negative association of Fe^{3+} and pH reveals that pH value is decreased due to dissolution of Fe^{3+} by microbial degradation aided by the presence of organic matter originated from waste water. This process is very complicated and needs detailed study.

In summer, Factor 1 accounts for 35.4% of variability and includes TDS, SO_4^{2-} and Cl^-, Mg^{2+} and NO_3^-. The loadings of TDS, SO_4^{2+}, Cl^- and Mg^{2+} are very strong and that of NO_3^- is moderate. Two processes are proposed, mixing of saline Sea water with fresh groundwater and anthropogenic pollution. During the summer season, without recharge, lowering of groundwater table occurs and the saline waterfront moves landward due to lower ground water head. In the mixing zone, mixing of saline water with fresh groundwater increases the concentrations of Mg^{2+}, SO_4^{2-}, Cl^- and TDS. Loading of NO_3^- in this factor indicates pollution of groundwater by domestic waste water infiltrating from the on-site sanitation systems. Factor 2 accounts for 18.9% of variability and includes parameters Na^+ and K^+ which are strongly correlated to this factor. The process attributed to this factor is cation exchange. Factor 3 accounts for 15.1% of variability and rendered two parameters Ca^{2+} and pH of which Ca^{2+} has moderate loading and pH has strong negative loading. The negative association of pH with Ca^{2+} means that pH increases as the calcium concentration in water drops. The process suggested is mineral precipitation, i.e., precipitation of calcium in groundwater onto the aquifer material, i.e., sand. Factor 4 accounts for about 11.7% of data variability and rendered two parameters: strongly correlated Fe^{3+}, and moderately correlated alkalinity. This factor is identified as organic matter degradation/iron reduction process. The microbial degradation of organic matter in the aquifer is associated with iron reduction.

In the summer season when Sea water intrusion takes place due to prevailing hydraulic conditions, ion exchange process takes place. Concentrations of Ca^{2+} and Mg^{2+} in groundwater are supposed to increase by ion exchange. In this process, Na^+ present abundantly in Sea water is adsorbed on the exchanger surface, i.e., the clay layer. Ca^{2+} previously present in the exchanger surface is exchanged with Na^+ and Ca^{2+} is released to the groundwater. Groundwater becomes enriched with calcium chloride or magnesium chloride (Appelo & Postma 2005). But, in summer sampling, the concentration of Na^+ has increased and that of Ca^{2+} and Mg^{2+} have reduced compared to their respective concentrations during post-monsoon (Figure 3). The increase of Na^+ concentration in summer cannot be explained by cation exchange process. The decreased concentrations of Mg^{2+} and Ca^{2+} observed in summer water samples can be explained by precipitation process. Due to Sea water mixing and cation exchange process, calcium and magnesium ions reach super saturation state causing their precipitation (Chapelle 1983). In the study area, near the beach, sandy cemented formation by calcium precipitation was observed during excavation work for sewerage construction up to 7 m depth. This observation supports the hypothesis of Ca^{2+} and Mg^{2+} precipitation. The major changes in concentrations of Ca^{2+}, Mg^{2+} and Na^+ in summer may be explained jointly by cation exchange and mineral precipitation processes.

During recharge of aquifer in the monsoon season, i.e., by freshening of aquifer, groundwater becomes rich in Ca^{2+}, Mg^{2+} and HCO_3^- due to dissolution of calcium and magnesium bearing minerals and carbonate shell present in the aquifer. Because of flushing and reverse ion exchange process during this period, Na^+ present in exchanger surface is replaced by Ca^{2+}. This causes groundwater to be rich in Na^+ and calcium bicarbonate type water converts to sodium bicarbonate type water. With more flushing, sodium bicarbonate type water converts to calcium bicarbonate type water.

Hierarchical cluster analysis

HCA is a powerful data mining technique. In this study, after careful consideration of available combinations of similarity/dissimilarity measurements, Ward's method with squared Euclidean distance as similarity measure provided visually meaningful dendrograms and distinct groups. The HCA of data sets in post-monsoon and summer sampling produced two dendrograms shown in Figures 4(a) and (b) respectively.

HCA was performed on the variables to obtain grouping of variables into clusters. The dendrograms consist of several clusters and each cluster contains one or several variables. For facilitating interpretation, clusters are selected based on visual examination of the dendrogram Figures 4(a) and (b).

In post-monsoon and summer data sets, each dendrogram has classified the eleven variables into four clusters.

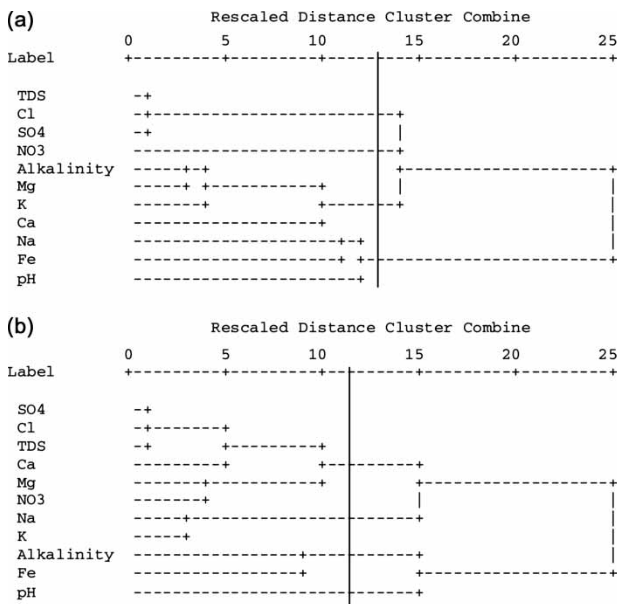

Figure 4 | (a) Dendrogram showing cluster of variables (water quality parameters) in post-monsoon and (b) dendrogram showing cluster of variables (water quality parameters) in summer.

This classification is subjective and four groupings were obtained by moving the phenon line (vertical solid line in the dendrogram). The variables in four clusters correspond to significant variables in four factors in both post-monsoon and summer season data sets except small differences. The clusters are sensitive to extreme values in the data sets. The similarity of PCA factors and HCA clusters confirms the dominant processes suggested from PCA.

CONCLUSIONS

Groundwater quality parameters from the coastal aquifer beneath Puri city in Orissa State, India, analysed from samples collected during post-monsoon season (November 2006) and summer season (first week of June 2007) reveal variations in concentrations of major cations in two sampling seasons. The water quality data sets of parameters from post-monsoon and summer sampling were analysed using two different multivariate statistical techniques such as PCA and HCA to understand dominant processes affecting the water quality parameters. From PCA of the two standardized data sets, four factors were obtained from each data set with varimax rotation. The rotated factors allowed interpretation of different geochemical processes. Though exploratory in nature, PCA is found to be a powerful technique which reduced the two large data sets of (51×11) and (43×11) matrices into two matrices, each of (11×4), i.e., (variables $\times$ factors) size. The significant variables in the factors aided in the interpretation of geochemical processes occurring in the coastal aquifer. From post-monsoon data set, the processes inferred from four factors are dilution of groundwater, mineral dissolution, weathering with anthropogenic pollution, and organic matter degradation with Fe(III) reduction. The processes determined from four factors in summer data set are mixing of saline and fresh water with anthropogenic pollution, cation exchange, mineral precipitation, and organic matter degradation with Fe(III) reduction. From HCA with Ward's method, four grouping of variables (cluster) were classified from the dendrogram of each of the two season data sets. The variables in the clusters were identical to the variables from significant factor loadings of PCA factor groups. The processes suggested from PCA factors are confirmed by HCA clusters. The present study validated the usefulness of the PCA and HCA techniques to interpret complex seasonal geochemical processes in the coastal fresh water aquifer in two seasons.

The observed fluctuations in major cations such as increase in concentration of Na^{+} and decrease in concentrations of Mg^{2+} and Ca^{2+} in summer compared to their post-monsoon concentrations cannot be explained by cation exchange alone. The increase of Na^{+} concentration and decrease of Mg^{2+} and Ca^{2+} concentrations can be explained by cation exchange and precipitation process due to Sea water intrusion. The two processes namely organic matter degradation/Fe(III) reduction and anthropogenic pollution are suggested to be occurring in both post-monsoon and summer sampling seasons. The recharge of the aquifer during monsoon season and intrusion of Sea water during summer season occur on an annual basis. They also influence geochemical processes and groundwater quality parameters.

REFERENCES

Appelo, C. A. J. & Postma, D. 2005 *Geochemistry, Groundwater and Pollution*, 2nd edition. A.A Balkema Publishers, Rotterdam, The Netherlands.

Báez-Cazull, S. E., McGuire, J. T., Cozzarelli, I. M. & Voytek, M. A. 2008 Determination of dominant biogeochemical processes in a contaminated aquifer-wetland system using multivariate statistical analysis. *Journal of Environmental Quality* **37**, 30–46.

Chapelle, F. H. 1983 Groundwater geochemistry and calcite cementation of the Aquia aquifer in Southern Maryland. *Water Resources Research* **19** (2), 545–558.

CDP 2006 *Puri City Development Plan*. Puri Municipality, Puri, India.

Guler, C., Thyne, G. D., McCray, J. E. & Turner, A. K. 2002 Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeology Journal* **10**, 455–474.

IMD 2008 *Rainfall data of Puri from Jan 2003 to Dec 2008*. Meteorological Department, Government of India, Bhubaneswar, India

Jain, A. K., Murty, M. N. & Flynn, P. J. 1999 Data Clustering: A Review. *ACM Computing Surveys* **31** (3), 264–323.

Jayakumar, R. & Siraz, L. 1997 Factor analysis in hydrogeochemistry of coastal aquifers – a preliminary study. *Environmental Geology* **31** (3/4), 174–177.

Kumar, A. R. & Riyazuddin, P. 2008 Application of chemometric techniques in the assessment of groundwater pollution in a suburban area of Chennai city, India. *Current Science* **94** (8), 1012–1022.

Liu, C. W., Lin, K. H. & Kuo, Y. M. 2003 Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *The Science of The Total Environment* **313** (1–3), 77–89.

Machado, C. J. F., Santiago, M. M. F., Frischkorn, H. & Filho, J. M. 2008 Clustering of groundwaters by Q-mode factor analysis according to their hydrogeochemical origin: a case study of the Cariri Valley (Northern Brazil) wells. *Water SA* **34** (5), 651–656.

NEERI 2008 *Study of Groundwater Aquifer System at Jagannath Puri and its Protection from Contamination*. Final Report, NEERI, Nagpur, India. April 2008.

Panigrahy, P. K., Das, J., Das, S. N. & Sahoo, R. K. 1999 Evaluation of the influence of various physico-chemical parameters on coastal water quality, around Orissa, by factor analysis. *Indian Journal of Marine Sciences* **28**, 360–364.

Papatheodorou, G., Lambrakis, N. & Panagopoulos, G. 2007 Application of multivariate statistical procedures to the hydrochemical study of a coastal aquifer: an example from Crete, Greece. *Hydrological Processes* **21** (11), 1482–1495.

Rani, A. & Babu, D. S. S. 2008 A statistical evaluation of ground water chemistry from the west coast of Tamil Nadu, India. *Indian Journal of Marine Sciences* **37** (2), 186–192.

Rao, Y. R. S., Keshari, A. K. & Gosain, A. K. 2010 Evaluation of regional groundwater quality using PCA and geostitistics in the urban coastal aquifer, East Coast of India. *International Journal of Environment and Waste Management* **5** (1–2), 163–180.

Reghunath, R., Murthy, T. R. S. & Raghavan, B. R. 2002 The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. *Water Research* **36**, 2437–2442.

Ruiz, F., Gomis, V. & Blasco, P. 1990 Application of factor analysis to the hydrochemical study of a coastal aquifer. *Journal of Hydrology* **119**, 169–177.

Sahu, K. C., Panda, U. C. & Mohapatra, D. M. 1998 Geo-chemistry and minerology of sediments in Rushikulya estuary, East coast of India. *Chemical Environmental Research* **7** (1–2), 77–92.

Sargaonkar, A. P., Gupta, A. & Devotta, S. 2008 Multivariate analysis of groundwater resources in Ganga-Yamuna basin (India). *Journal of Environmental Science and Engineering* **50** (3), 215–222.

Singh, K. P., 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 Research* **38**, 3980–3992.

Singh, K. P., Malik, A., Singh, V. K., Mohan, D. & Sinha, S. 2005 Chemometric analysis of groundwater quality data of alluvial aquifer of Gangetic plain, North India. *Analytica Chimia Acta* **550** (1–2), 82–91.

Singh, S. K., Singh, C. K., Kumar, K. S., Gupta, R. & Mukherjee, S. 2009 Spatial-temporal monitoring of groundwater using multivariate statistical techniques in Bareilly District of Uttar Pradesh, India. *Journal of Hydrology and Hydromechanics* **57** (1), 45–54.

Sundaray, S. K. 2010 Application of multivariate statistical techniques in hydrogeochemical studies – a case study: Brahmani–Koel River (India). *Environmental Monitoring and Assessment* **164** (1–4), 297–310.

Sundaray, S. K., Panda, U. C., Nayak, B. B. & Bhatta, D. 2006 Multivariate statistical techniques for the evaluation of spatial and temporal variation in water quality of Mahanadi River – Estuarine System (India) – a case study. *Environmental Geochemistry and Health* **28**, 317–330.

Vijay, R., Khobragade, P. & Mohapatra, P. K. 2011 Assessment of groundwater quality in Puri City, India: an impact of anthropogenic activities. *Environmental Monitoring Assessment* **177**, 409–418.

First received 15 December 2010; accepted in revised form 15 March 2011