XXXVIII IAH Congress

Groundwater Quality Sustainability
Krakow, 12–17 September 2010

Extended Abstracts

Editors:
Andrzej Zuber
Jarosław Kania
Ewa Kmiecik

University of Silesia Press 2010

abstract id: **236**

topic: **1**
Groundwater quality sustainability

1.5
Groundwater quality and mining

title: **Determination of processes affecting groundwater quality in coastal aquifer of Puri City using multivariate statistical analysis**

author(s): **Prasanta K. Mohapatra**
Orissa Water Supply and Sewerage Board, India,
prasant_mohapatra@hotmail.com

Ritesh Vijay
National Environmental Engineering Research Institute, India,
r_vijay@neeri.res.in

Paras R. Pujari
National Environmental Engineering Research Institute, India,
prpujari123@gmail.com

keywords: principal component analysis, hierarchical cluster analysis, groundwater quality, Puri city

INTRODUCTION

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: organic matter degradation/ aerobic respiration, iron reduction, cation exchange, mixing of salt water with fresh water etc. In order to characterize water quality of a place, it becomes necessary to collect large dataset of water quality parameters. Due to effects of multiple processes on a water quality parameter as well as the collection of large dataset of parameters, it becomes difficult to interpret relationships among the parameters and it is not possible to interpret the governing biogeochemical processes. Multivariate statistical analysis has been used in recent years to provide a quantitative measure of relatedness of water quality parameters and to suggest the underlying 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., 2009), (ii) providing insight into the processes (Machado et al., 2008; Báez-Cazull et al., 2008), (iii) identifying critical water quality issues and possible sources of pollution/ polluting processes (Sargaonkar et al., 2008; Kumar, Riyazuddin, 2008) and (iv) interaction of river water/groundwater and groundwater mixing (Reghunath et al., 2002).

In this paper, the multivariate statistical analysis such as principal component analysis (PCA) and hierarchial cluster analysis (HCA) were applied to the data sets of chemical analysis of groundwater samples collected in two seasons (one set in November 2006, after monsoon rain and another set in June 2007, before monsoon), to elicit hidden processes affecting groundwater quality in the coastal aquifer of Puri city in Orissa.

Study area

The study area is a 16 km^2 urban area of which the Puri built up area occupies about 10 Km^2 and the remaining land surface is almost vacant. Most of the vacant land surface constitutes two groundwater well fields which supply domestic water to Puri city. The city is located on the coast of Bay of Bengal in Orissa. The aquifer beneath is sandy and unconfined, having depth of about 40 m and lies on a thick clay layer. Thin clay layers are also found in isolated lens and patches at shallow depths within the unconfined aquifer. The area receives rainfall during monsoon season (2nd week of June to October).

METHODOLOGY

Water Quality Parameters and Data Analysis

The water quality data of Puri city were obtained from two sampling seasons. In the first sampling season in November 2006, after monsoon rain, groundwater samples were collected from 51 numbers of tube wells and production wells scattered all over the city area and two groundwater well field areas. In the second sampling season in the first week of June, before rain, water samples were collected from 43 numbers of tube wells and production wells from the city area as well as two groundwater well field areas. The water samples were analysed for chemical parameters at NEERI laboratory following Standard Methods. The eleven water quality parameters considered for multivariate statistical analysis are pH, total dissolved solid (TDS), alkalinity, Na, K, Ca, Mg, Fe, SO_4, NO_3 and Cl. Before performing PCA and HCA, the parameters in the data sets were tested for normality. The distribution of all parameters in the two data sets was found non-normal except for pH, which fits normal distribution. In order to avoid the problem of difference in scale, i.e., range of values and unit, among the water quality parameters, the parameters were standardized using z-score: $z = (y_i - \hat{y})/s$, where $\hat{y}$ is the average value of a parameter in the sampling season and s is its standard deviation. The software SPSS Statistics 17.0 was used for data standardization and multivariate statistical analysis.

Principal Component Analysis

The PCA is performed to reduce the large data set of variables, i.e. water quality parameters, into few factors called the principal components which can be interpreted to reveal underlying data structure. The maximum number of principal components is equal to the number of variables. But, only few factors based on Kaiser criterion, i.e., principal components having latent root or Eigen value greater than one, are retained in the analysis. 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. 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 and indicates the proportion of variance which might be caused by underlying factors. A high value (close to 1) indicates usefulness of PCA. The Bartlett's test of sphericity indicates whether correlation matrix is an identity matrix, which indicates that variables are unrelated.

Hierarchial 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 seen visually as a dendrogram. The HCA of the standardized data set using Ward's method based on Euclidean distance was performed to classify the parameters into clusters based on their similarities. This method was found to provide best dendrogram of meaningful clusters with the proximity of clusters measured in a rescaled distance cluster combine.

RESULTS AND DISCUSSION

Hydrochemistry 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 water quality parameter concentrations vary across space and time in the two sampling seasons (NEERI, 2008; Vijay et al., 2009).

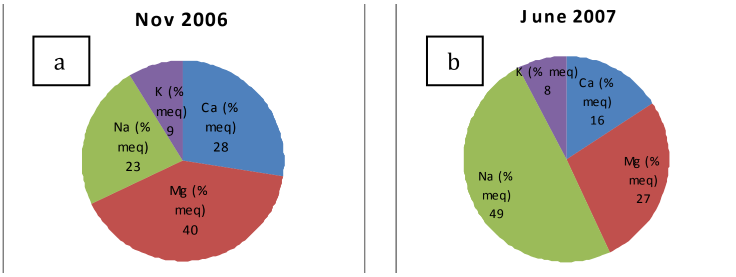

Figure 1. Changes in chemical facies (average value of cations in % milli-equivalent) in groundwater samples in a) post-monsoon (Nov 2006) and b) pre-monsoon (June 2007).

The average concentrations of major cations in two sampling periods are shown in Figure 1. The concentration of Na^{+} ions are observed greatly increased in pre-monsoon along with reduction in concentration of Mg^{2+} and Ca^{2+} ions. The concentration of parameters appears to be affected by several geochemical processes occurring in the aquifer. The aquifer is recharged by monsoon rain each year during second week of June to end of October. Dilution and freshening is a predominant process during this period. The intrusion of sea water occurs in summer due to favourable hydraulic condition. Sea water intrusion and freshening of groundwater occur in a cyclic manner and have influence on the chemistry of groundwater quality parameters.

Principal Component Analysis

In the post-monsoon data set, four factors were found significant having Eigen values greater than unity and together they account for 73% of the variability of the data set. In the pre-monsoon data set, four factors were found significant 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 two data sets are given in Table 1.

Table 1. Varimax orthogonal rotated factor loadings from principal component analysis of water quality dataset of post-monsoon (Nov 2006) and pre-monsoon (June 2007).

Parameter	Factor 1 (27%) Post-monsoon	Factor 1 (35%) Pre-monsoon	Factor 2 (24%) Post-monsoon	Factor 2 (19%) Pre-monsoon	Factor 3 (12%) Post-monsoon	Factor 3 (15%) Pre-monsoon	Factor 4 (10%) Post-monsoon	Factor 4 (12%) Pre-monsoon
Fe	-0.269	0.039	-0.095	-0.197	0.156	-0.004	**-0.809**	**0.870**
pH	-0.464	-0.075	0.040	-0.068	0.185	**-0.861**	**0.606**	-0.039
TDS	**0.841**	**0.832**	0.441	0.367	0.080	0.302	0.028	0.150
Alkalinity	0.306	0.195	**0.855**	0.483	-0.089	0.040	0.132	**0.681**
Na	0.295	0.255	**-0.526**	**0.848**	**0.581**	0.038	0.099	-0.050
K	0.302	0.236	**0.794**	**0.878**	0.093	0.121	0.145	0.011
Ca	0.183	**0.564**	0.361	0.134	**0.534**	**0.685**	0.059	-0.043
Mg	0.156	**0.843**	**0.765**	0.268	0.148	0.088	-0.039	0.192
SO4	**0.845**	**0.856**	0.284	0.245	-0.104	0.333	0.039	0.014
NO3	0.250	**0.742**	-0.029	0.131	**-0.733**	-0.319	0.117	0.047
Cl	**0.928**	**0.848**	0.146	0.116	0.060	0.338	0.066	0.014
Interpretation of Processes	Dilution of saline and fresh water	Mixing of saline and fresh water/ Anthropogenic pollution	Mineral dissolution	Cation exchange	Weathering/ Anthropogenic pollution	Mineral precipitation	Organic matter degradation/ Iron reduction	Organic matter degradation/ Iron reduction

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 ±0.5 to ±0.74 are considered moderately correlated and loadings approaching 0 indicate weak correlations (Liu et al., 2003). Based on the significant factor loadings, each factor has been 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 of each factor are given in Table 1.

In the post-monsoon sampling period, in Factor 1, the variables with strong loadings are TDS, SO_4 and Cl. This factor explains 27% of the variability in the data set. The process assigned to this factor is dilution of groundwater since the concentration of the parameters chloride and sulphate is very much reduced due to recharge effect of rain water as compared to the concentration of these parameters in pre-monsoon, i.e., summer samples. Factor 2 explains 24% of the variability and is highly correlated with parameters alkalinity, K and Mg. This factor is interpreted as mineral dissolution factor. The K and Mg ions are released to groundwater due to dissolution of minerals bearing these ions during recharge of aquifer by rainfall. Factor 3 accounts for about 12% of the variance of the data set and includes Na, Ca and NO_3 parameters which are moderately correlated to the factor. The concentration Na is reduced in recharge process. The sign of NO_3 is negative. Two distinct processes are interpreted, weathering and anthropogenic pollution. The Ca concentration is increased due to weathering or dissolution of calcite minerals such as lime stone which is abundant in the study area. The other process is interpreted as anthropogenic pollution due to presence of NO_3 in this factor, which has its origin from on-site sanitation in the study area. Due to dilution, concentration of NO_3 is reduced. Factor 4 accounts for 10% of variability and has loadings of Fe and pH. The loading of Fe is strong and negative and the loading of pH is moderate and positive. The opposite sign of pH indicates that its value is decreased due to dissolution of Fe by microbial degradation aided by the presence of organic matter from sewage in the aquifer. This process is very complicated and needs detailed study.

In the pre-monsoon sampling, Factor 1 accounts for 35% of variability and includes TDS, SO_4 and Cl, Mg and NO_3. The loadings of TDS, SO_4, Cl and Mg are very strong and that of NO_3 is moderate. Two processes are interpreted, mixing of saline sea water with fresh groundwater and anthropogenic pollution. During summer season, without recharge, lowering of groundwater table occurs and the saline water front moves landward due to lower ground water head. In the mixing zone, mixing of saline water with fresh groundwater increases concentration of SO_4 and Cl parameters. The parameter NO_3 indicates pollution of groundwater by domestic waste water infiltrating from the on-site sanitation systems. The Factor 2 accounts for 19% of variability and includes parameters Na and K which are strongly correlated to this factor. The process attributed to this factor is called cation exchange. Factor 3 accounts for 15% of variability and rendered two parameters Ca and pH of which Ca has moderate loading and pH has strong negative loading. The process interpreted is called mineral precipitation, i.e. precipitation of calcium in groundwater onto the aquifer material, i.e. sand. Factor 4 accounts for about 12% of data variability and rendered two parameters: Fe, strongly correlated and alkalinity, moderately correlated with Factor 4. 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.

When sea water intrusion occurs, ion exchange process is said to take place. In this process, Na^+ ions present abundantly in sea water are adsorbed on the exchanger surface, i.e. the clay layer.

The Ca^{2+} ions previously present in the exchanger surface are exchanged with Na^{+} ions and Ca^{2+} ions are released to the groundwater (Appelo, Postma, 2005). The groundwater becomes calcium chloride or magnesium chloride type. During recharge of aquifer by rain, i.e. freshening of aquifer, groundwater becomes rich in Ca^{2+} and Mg^{2+} cations and HCO_3^- anion due to dissolution of calcium and magnesium bearing minerals present in the aquifer. Flushing takes place during this period and by the process of reverse ion exchange process, Na^{+} ions present in exchanger surface are replaced by Ca^{2+} ions. So, groundwater becomes rich in Na^{+} ions and calcium bicarbonate type water converts to sodium bicarbonate type water. With more flushing, sodium bicarbonate type water converts to calcium bicarbonate type water. But, in the study area, in pre-monsoon samples, concentrations of Na^{+} and K^{+} were found increased and concentrations of Ca^{2+} and Mg^{2+} reduced compared to their concentrations in post-monsoon samples. The increase of Na^{+} and K^{+} concentration in pre-monsoon samples can not be explained by cation exchange process since Ca^{2+} and Mg^{2+} concentrations in water samples are supposed to increase when ion exchange process takes place. The reduced concentrations of Mg^{2+} and Ca^{2+} in pre-monsoon can be explained by precipitation process. Due to mixing of sea water and cation exchange process, calcium and magnesium ions reach super saturation state causing their precipitation (Chapelle, 1983). In the study area, cemented sandy strata probably due to calcium precipitation have been observed at 5 to 7 m depth near the beach. This observation supports the hypothesis of precipitation.

Hierarchial Cluster Analysis

HCA is a powerful data mining technique. In this study, the Ward's method with Euclidean distance provided visually meaningful dendrograms. The HCA of two data sets produced two dendrograms which are shown in Figure 2. The processes interpreted from factors and the clusters from two data sets are consistent.

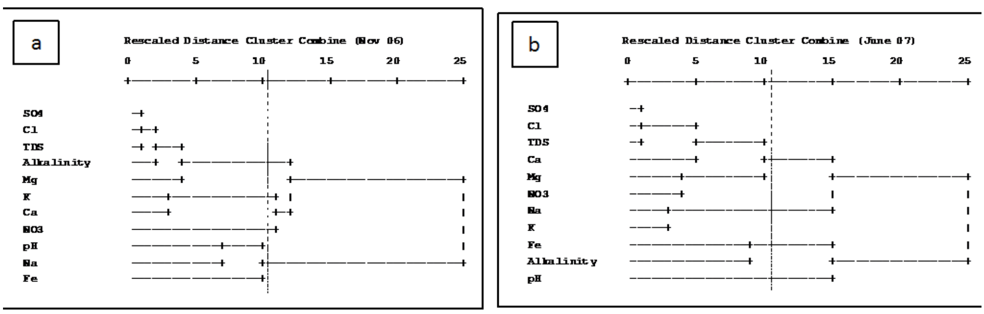

Figure 2. Dendrogram of a) post-monsoon (Nov 2006) data set and b) pre-monsoon (June 2007) data set

CONCLUSIONS

Groundwater quality data set of Puri city, Orissa in India collected during post-monsoon season of November 2006 and pre-monsoon season of June 2007 were analysed using two different multivariate statistical techniques such as PCA and HCA. In PCA of two standardized data sets, four factors were obtained from each data set with varimax rotation. Though exploratory in nature, PCA reduced the two large data sets into two small matrices. The factors aided in the interpretation of geochemical processes occurring in the coastal aquifer. From post-monsoon data set, the processes interpreted from four factors are dilution of groundwater, mineral dissolution, weathering with anthropogenic pollution and organic matter degradation with iron reduction. The processes interpreted from four factors in pre-monsoon data set are mixing of saline and fresh water with anthropogenic pollution, cation exchange, mineral precipitation and organic matter degradation with iron reduction. In each of the two data sets, HCA with Ward's method produced four grouping of variables. The interpretation of the processes by PCA and HCA are consistent. The present study validated usefulness of the PCA and HCA techniques to interpret complex geochemical processes in a coastal aquifer.

REFERENCES

Appelo C.A.J., Postma D., 2005: *Geochemistry, groundwater and pollution.* 2nd Edition, Rotterdam, A.A Balkema Publishers, 242 p.

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, Vol. 37, pp. 30–46.

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

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, Vol. 94(8), 25 April 2008, pp. 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, Vol. 313(1-3), pp. 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, Vol. 34(5), pp. 651–656.

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

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, Vol. 36, pp. 2437–2442.

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

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, Vol. 57(1), pp. 45–54.

Vijay R., Khobragade P, Mohapatra P.K., Kelkar P.S., Sohony R.A., 2009: *Suitability and assessment of groundwater quality due to anthropogenic sources: a case study of Puri city, India.* Proceedings of the International Symposium on Environmental Science and Technology, Shanghai, China, 2-5 June 2009, pp. 594–599.

International Association of Hydrogeologists

AGH University of Science and Technology

2-vol. set + CD
ISSN 0208-6336
ISBN 978-83-226-1979-0