IOSR Journal Of Environmental Science, Toxicology And Food Technology (IOSR-JESTFT)
e-ISSN: 2319-2402,p- ISSN: 2319-2399.Volume 6, Issue 1 (Sep. - Oct. 2013), PP 36-41
www.Iosrjournals.Org

Assessment of groundwater quality in Karur block of Tamil Nadu using multivariate techniques: A case study

S. Krishnaraj[1], Sanjiv Kumar[2] and K. P. Elango[3]
[1]*Department of Chemistry, R.V.S College of Engineering & Technology, Coimbatore, India*
[2]*School of Sciences, IGNOU, New Delhi, India*
[3]*Department of Chemistry, Gandhigram Rural Institute-Deemed University, Gandhigram, India*

Abstract: *In this case study, multivariate statistical techniques, such as principal component analysis, factor analysis and cluster analyses were applied for evaluation of temporal/spatial variations in the groundwater quality. These techniques were employed for the better interpretation of large complex water quality data set monitored in the four seasons from twenty five groundwater locations of Karur block, Tamil Nadu during the year 2012. The water samples were characterized for the physico-chemical parameters such as temperature, pH, total alkalinity, electrical conductivity, total hardness, calcium ions, magnesium ions, total dissolved solids, fluorides, chlorides and sulphates. The data obtained were subjected to principal component analysis (PCA) for simplifying its interpretation and to define the parameters responsible for the main variability in water quality variance. The results of principal component analysis evinced that all the parameters equally and significantly contributed to water quality variations in the study area in all the seasons. Hierarchical cluster analysis grouped twenty five sampling stations into three clusters (i.e.) relatively less polluted (LP), moderately polluted (MP) and highly polluted (HP) sites based on the similarity of water quality characteristics. The water quality index (WQI) of these groundwater samples ranged from 47 to 107, 53 to 96 and 45 to 94 in post-monsoon, summer and pre-monsoon seasons, respectively. This investigation revealed that the groundwater of the study area needs some degree of water treatment before consumption. Thus, this study demonstrates the usefulness of multivariate statistical techniques for effective groundwater quality management.*
Key words: *water quality index, principal component analysis, cluster analysis, factor analysis.*

I. Introduction

In recent years, the increasing threat to ground water quality due to human activities has become a matter of great concern. Hence continuous monitoring on groundwater becomes mandatory in order to minimize the groundwater pollution and to have control on the pollution causing agents. The application of different multivariate statistical techniques helps in the interpretation of complex data matrices to understand the water quality and offers a valuable tool for reliable management of water resources as well as rapid solution to problems [1,2]. Thus, the main objective of this study is to assess the groundwater quality and its suitability for drinking and domestic purposes using multivariate techniques as a large population of Karur town depends on groundwater for drinking purpose. Different chemometric analysis such as factor analysis (FA), cluster analysis (CA) and principal component analysis (PCA) are all used for better assessment of the water quality and to identify the pollution source apportionment [3,4]. They are designed to reduce the number of variables to a small number of indices while attempting to preserve the relationships present in the original data. With this background, an attempt was made to implement principal component analysis method in order to identify practical pollution indicators in the study area. Furthermore, this study also intends to provide a basis on developing realistic tools that could help local decision-makers on the suitable management of the groundwater in the area.

II. Materials and Methods

2.1. Study area

The Karur block of Tamil Nadu, India is located at 10.95º N and 78.08 ºE with a mean elevation 122 m. The average annual rainfall is about 855 mm. The city gets most of its seasonal rainfall from the north east monsoon between late September to mid November. Vast mineral deposits, availability of water and good infrastructure are conducive for industrialization in the Amravati river basin of Karur has resulted in heavy textile based industrialization. Many small, medium and large scale textile industries are situated in the region and these establishments have adversely affected the ground water quality. Besides, the increased population and improper drainage system in the study area have immense potential to influence the ground water quality.

2.2. Sample collection and monitoring parameters

A total of 25 water quality monitoring stations were identified and water samples were collected in the middle month of four seasons namely post-monsoon (January – March), summer (April –June), pre-monsoon (July – September) and monsoon (October – December) of the year 2012. The groundwater samples were analyzed for parameters which include pH, electrical conductivity, total dissolved solids, total alkalinity, total hardness, Ca (II), Mg (II), Na, K, fluorides, sulphates and chlorides using standard protocols [5] and the quality of the data was ensured through careful standardization.

2.3. Water Quality Index (WQI)

The water quality index (WQI) of groundwater was calculated using weighted Arithmetic Index method and quality rating/ sub index (Qi) corresponding to the i^{th} parameter P_i is a number reflecting the relative value of this parameter.

Step I: Unit weight (Wi) was calculated by a value inversely proportional to the recommended standard (S_i) of the corresponding parameter.

Step II: Quality rating (Qi) is calculated by using the following expression

$$\text{Qi} = \sum_{i=1}^{n}[(Mi - Ii)/(Si - Ii)] * 100$$

Where Mi=Estimated values of the i^{th} parameter in the Lab

Ii= Ideal values of the i^{th} parameter

Si= Standard values of the i^{th} parameter

All the Ideal values (Ii) are taken as zero except for pH=7 and F=1.0

Step III: The overall WQI was calculated by aggregating the quality rating (Qi) with unit weight (Wi) linearly

$$\text{WQI} = (\sum_{i=1}^{n} QiWi)/(\sum_{i=1}^{n} Wi)$$

TDS, pH, chlorides, fluorides, total hardness, total alkalinity, calcium, magnesium were recognized as preliminary indication of quality and it is used in calculating WQI for public water supply.

2.4. Multivariate statistical techniques

2.4.1. Principal Component Analysis

The data obtained from the laboratory analysis were used as inputs for principal components analysis (PCA). PCA was performed using the XLSTAT 2013 software. As a multivariate data analytic technique, PCA reduces a large number of variables to a small number of variables without sacrificing too much of the information [6]. More concisely, PCA combines two or more correlated variables in to one variable. This approach has been used to extract variables and infer the processes that control water chemistry [7,8]. The principal component is expressed as

$$Z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + \ldots\ldots\ldots\ldots + a_{im}x_{mj}$$

Where *z* is the component score, *a* is the component loading, *x* is the measured value of variable, i is the component number and m the total number of variables.

2.4.2. Cluster Analysis

Cluster analysis is a group of multivariate technique whose primary purpose is to assemble objects based on the characteristics they possess. Cluster analysis classifies objects, so that each object is similar to the others in the cluster with respect to a pre determined selection criterion. The resulting clusters of objects should then exhibit internal (within cluster) homogeneity and external (between clusters) heterogeneity. Hierarchical Agglomerative Clustering is the most common approach, which provides intuitive similarity relationships between any one sample and the entire data set and is typically illustrated by a dendrogram (tree diagram) [9]. The dendrogram provides a visual summary of the clustering processes, presenting a picture of the groups and their proximity, with a dramatic reduction in dimensionality of the original data. The Euclidean distance usually gives the similarity between two samples and a distance can be represented by the differences between analytical valves from the sample [10].

2.4.3. Factor Analysis

The main purpose of factor analysis (FA) is to reduce the contribution of these significant variables to simplify even more of the data structure coming from PCA. This purpose can be achieved by rotating the axis defined by PCA, according to well establised rules and contributing new variables, also called varifactors (VF). PC is a linear combination of observable water quality variables, whereas VF can include unobservable, hypothetical, latent variables [11,7]. The FA is expressed as

$$Z_{ji} = a_{f1}f_{1i} + a_{f2}f_{2i} + \ldots\ldots\ldots\ldots + a_{fm}f_{mi} + e_{fi}$$

Where *z* is the measured variable, *a* is the factor loading, *f* is the factor score, *e* is the residual term accounting for errors or other source of variation, *i* is the sample number and *m* the total number of factors.

III. Results and Discussion

The groundwater samples collected during the four seasons were analyzed and their descriptive statistical data is presented in Table 1 and 2. Water quality index (WQI) was calculated by considering individual values of analysed water quality parameters (Table 3). The observed water quality was lower in post-monsoon and summer season as compared to pre-monsoon season. Nearly 64 % of groundwater samples each in post-monsoon and summer season were found in poor status and 32 % of the samples were in the very poor status. About 52 % of groundwater samples were found to be good for drinking purpose in pre-monsoon season where as only 4% were found to be good in both summer and post-monsoon season. This increment of poor category in post-monsoon and summer season as compared to pre-monsoon season indicates the groundwater quality in the study area is slowly getting to degradation. Waste water intrusion and the effect of anthropogenic actions may be the reasons for this high level contamination.

The water quality data collected were subjected to PCA to understand the influence of various parameters on the quality of the groundwater in the study area. PCA technique extracts the eigen value and eigen vectors from the covariance matrix of original variables. The principal components (PC) are the uncorrelated (orthogonal) variables obtained by multiflying the original correlated variables with the eigen vector, which is a list of coefficients (loadings or weighings). Thus, the PCs are the weighed linear combinations of the original variables. PC provides information on the most meaningful parameters, which describe the whole dataset while affording data reduction with a minimum loss of original information. The eigen values, percentage of variance and cumulative percentage for all the water samples in the study resulted from the PCA are presented in the Table 4. From the water quality parameters data, three principal components, explaining 76.9 - 86.5 % of the total variance was established on the basis of Kaiser criterion of the eigen values greater than or equal to 1.0 and from cattel Scree plots. A Scree plot shows the eigen values sorted from large to small as a function of principal component number (Fig 1). These plots indicated that after the third PC, starting the elbow in the downward curve, other components can be omitted.

The components loadings of the first five retained PCs are summerised in Table 4. The component loading was classified as per Liu [12], who categorised the component loadings as strong, moderate and weak corresponding to absolute loading values of greater than 0.75 as strong, 0.75 - 0.50 as moderate and 0.50 - 0.40 as weak. The higher the loading of component variable, the more that variable contributes to the variation accounted for by the principal component. Further if a given variable has a meaningful loading (if its absolute value exceeds 0.4) on more than one component, that variable may be scratched out and also be ignored from interpretation because such variables are not pure measures of any one construct.

The PCA of water quality parameters from the 25 sampling stations indicated that in groundwater, altogether three factors (PCs) explain 76.9 - 86.5 % of the total variance. During post-monsoon season, principal component 1 (PC1) accounts for 61.1 % of the total variance. The PC1 is associated with strong positive loadings of EC, TDS, TA, Ca (II) and Mg (II) and with moderate positive loadings of pH (Table 4). In these samples, the EC and TDS levels were found to be slightly higher than that recommended by the WHO. EC varied from 420 to 3182 with an average of 1599 µS and TDS varied from 280 to 2205 mg/L with an average of 1179 mg/L. Total alkalinity concentration ranged from 80 to 948 with an average of 401 mg/L, Ca (II) varied from 32 to 325 mg/L with an average of 126 mg/L and Mg (II) varies from 14 to 198 mg/L with an average of 81mg/L. In this component EC, TDS, TA, Ca (II) and Mg (II) correlate significantly ($0.979< r >0.781$) with each other, suggesting a common source. As it is obvious PC1 can be called as salt component because it is mainly saturated with EC, TDS and TA including Ca (II) and Mg (II) ions. All these parameters are typical indicators of natural diffuse inputs. PC 2 accounts for 16.6 % of the total variance with moderate loading on pH. This principal component can there be termed as acidity - alkalinity component. In this component pH ranged from 7.1 to 8.6 with an average of 7.48. During this season PC 3 explains 8.8 % of the total variance with weak loading on total hardness (TH). TH varied from 137 to 940 with an average of 399 mg/L. During post monsoon season, other water quality parameters *viz* Na, K, Cl, SO_4 and F were ignored from the interpretations as these variables have showed a meaningful loading (absolute values exceeds 0.4) on more than one component.

During summer season, principal component 1 (PC1) accounts for 56.7 % of the total variance. The PC1 is associated with strong positive loadings of EC, TDS, TA and Ca (II) and with moderate positive loadings of pH and Mg (II) (Table 4). It was observed that the EC and TDS levels in the samples collected during summer season were found to be slightly higher than that recommended by the WHO. In the case of monsoon season, principal component 1 (PC1) accounts for 55.6 % of the total variance. The PC1 is associated with strong positive loadings of EC, TDS, TA and Ca (II) and with moderate positive loadings of pH and Mg (II). The EC and TDS levels were also found to be slightly higher during monsoon season than the critical levels recommended by the WHO. Overall, the results of PCA indicated that all the parameters equally and significantly contributed to the water quality variations in the study area in all the seasons. Our findings corroborates with the results of several researchers [13,14,4] who demonstrated the usefulness of PCA to identify the important component that influenced the water quality.

3.1. Spatial similarity and site grouping

Cluster analysis was used to detect the similarity groups between the sampling sites. It yielded a dendrogram (Fig 2), grouping all 25 sampling sites of Karur block into three statistically significant clusters. The Cluster 1, composed of the sampling stations assigned by numbers 1,2,3,4,5,8,9,13,14,15 and 21 and concerns 44 % of the total water samples, corresponds to low polluted (LP) sites. Cluster 2 represented by sample numbers 6,10,11,16,23 and 24 corresponds to moderately polluted (MP) sites. It occupies 24 % of the total water samples. It is observed that in all seasons, the cluster 1 and 2 are the same. Sample numbers 7, 18 and 25 are found to be in Cluster 1 in all the season except in monsoon season. Cluster 3 includes sample numbers 17 and 19 corresponds to highly polluted (HP) sites. Similar interpretations were given by several researchers [2,3] for classifying the water quality of sampling sites using the cluster analysis. Our results indicated that the cluster analysis technique is useful in getting reliable classification of groundwater in the selected area and will make it possible to design a future spatial sampling strategy in an optimal manner.

IV. Conclusion

This investigation employed different multivariate statistical techniques and evaluated spatial and temporal variations in the groundwater quality of Karur block, Tamil Nadu. This case study revealed that principal component analysis helped to extract and identify the factors responsible for variations in the groundwater quality at different sampling sites. Besides, the cluster analysis grouped 25 sampling sites into three clusters of similar water quality characteristics. Based on obtained information, it is possible to design a future optional sampling strategy, which could reduce the number of sampling stations and associated costs. This case study with Karur block as a model system demonstrated the scope of multivariate statistical techniques for analysis and interpretation of complex datasets to undertake meaningful decisions for effective management of groundwater quality. Such techniques need to be explored to save crucial response time to potential contamination risks.

References

[1] V. Simeonov, P. Simeonova, and R. Tsitouridou, Chemometric quality assessment of surface waters: two case studies, *Chem. Eng. Ecol. 11,* 2004, 449-469.

[2] S. Shrestha and F. Kazama, Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan, *Environmental Modelling and Software, 22,* 2007, 464-475.

[3] T.G. Kazi, M.B. Arain, M.K. Jamali, N. Jalbani, H.I. Afridi, R.A. Sarfraz, J.A. Baig, and A.Q. Shah, Assessment of water quality of polluted lake using multivariate statistical techniques: A case study, *Ecotoxicology and Environmental Safety, 72,* 2009, 301-309.

[4] L. Belkhiri, A. Boudoukha, L. Mouni, and T. Baouz, Multivariate statistical characterization of groundwater quality in Ain Azel plain, Algeria, *African Journal of Environmental Science and Technology, 4(8),* 2010, 526-534.

[5] APHA, *Standard methods for the examination of water and wastewater* (American Water Works Association, Washington DC, 1985).

[6] G. Qian, G. Gabor and R.P. Gupta, Principal component selection by the criterian of the minimum mean difference of complexity, *J. Multivariate Anal., 49(1),*1994, 55- 75.

[7] B. Helena, R. Pardo, M. Vega, E. Barrdo, J.M. Fernandez, L. Fernandez, Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, spain) by principal component analysis., *Water Res. 34*, 2000, 807- 816.

[8] M.C. Hidalgo and J. Cruz- Sanjulian, Ground water composition, hydrochemical evolution and mass transfer in a regional detrital aquifer (Baza Basin, southern spain), Appl. *Geochem., 16 (7-8),* 2001,745- 758.

[9] J.E. McKenna, An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis, *Environmental Modelling and Software, 18 (3),* 2003, 205-220.

[10] M. Otto, Multivariate methods, in: R. Kellner, J.M. Mermet, M. Otto and H.M. Widmer (Ed.), *Analytical Chemistry* (Wiley/VCH, Weinheim, Germany, 1998) 916.

[11] M. Vega, R. Pardo, E. Barrado and L. Deban, Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis, *Water Res.32,* 1998, 3581-3592.

[12] C.W. Liu, K.H. Lin and Y.M. Kuo, Y.M, Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan, *Sci. Total Environ. 313,* 2003, 77-89.

[13] K. Bengraine and T.F. Marhaba, Using principal component analysis to monitor spatial and temporal changes in water quality, *Journal of Hazardous Materials, B100,* 2003, 179-195.

[14] M. Karthikeyan, *Chemometric analysis of water quality parameters in Dindigul district and removal of fluoride using polymeric materials,* doctoral diss., Gandhigram Rural Institute – Demmed University, Gandhigram, 2010.

Table 1. Descriptive statistical data of groundwater samples during post-monsoon and summer

Variable	Post-monsoon				Summer			
	Minimum	Maximum	Mean	Std. Deviation	Minimum	Maximum	Mean	Std. Deviation
EC	420	3182	1599	728	375	2863	1438	654
TDS	280	2205	1180	540	322	1948	1067	482
TA	80	948	403	224	72	753	355	193
TH	137	940	400	254	168	855	423	211
pH	7.1	8.6	7.5	0.4	7.1	8.4	7.5	0.4
Ca (II)	32	315	126	80	39	310	109	60
Mg (II)	14	198	81	49	14	147	75	33

Variable	Post-monsoon				Summer			
	Minimum	Maximum	Mean	Std. Deviation	Minimum	Maximum	Mean	Std. Deviation
Na	56	470	184	99	50	261	137	53
K	14	158	48	46	12	142	43	42
F	0.3	0.8	0.6	0.1	0.4	0.8	0.7	0.1
SO_4	47	265	120	67	56	298	137	75
Cl	48	428	201	109	43	385	180	97

Table 2. Descriptive statistical data of groundwater samples during pre-monsoon and monsoon

Variable	Pre-monsoon				Monsoon			
	Minimum	Maximum	Mean	Std. deviation	Minimum	Maximum	Mean	Std. deviation
EC	142	2580	1102	645	372	2824	1426	645
TDS	289	1752	950	433	317	1927	1045	476
TA	62	692	324	178	173	876	455	205
TH	149	946	404	204	195	954	450	204
pH	6.9	8.2	7.3	0.3	7.1	7.9	7.3	0.2
Ca (II)	24	242	98	48	40	204	98	49
Mg (II)	14	147	76	33	17	164	73	44
Na	42	221	115	44	20	154	65	36
K	10	127	38	37	11	118	29	25
F	0.4	0.8	0.7	0.1	0.2	0.6	0.4	0.1
SO_4	50	274	121	65	25	168	76	41
Cl	46	328	178	101	32	264	108	62

Table. 3 Water Quality Index of groundwater samples

WQI	Status	Seasons					
		Post-monsoon		Summer		Pre-monsoon	
		No. of samples	%	No. of samples	%	No. of samples	%
0 - 24	Excellent	Nil	Nil	Nil	Nil	Nil	Nil
25 - 50	Good	1	4	1	4	13	52
51 - 75	Poor	16	64	16	64	6	24
Above 75	Very Poor	8	32	8	32	6	24

Table 4. Eigen value (EV), percentage of variance (V) and cumulative percent (C) of water quality parameters

PC	Post-monsoon			Summer			Pre-monsoon			Monsoon		
	EV	V (%)	C (%)	EV	V (%)	C (%)	EV	V (%)	C (%)	EV	V (%)	C (%)
1	7.33	61.1	61.1	6.81	56.8	56.8	6.21	51.8	51.8	6.67	55.6	55.6
2	2.00	16.6	77.7	1.94	16.2	73.0	1.98	16.5	68.3	1.57	13.1	68.7
3	1.06	8.9	86.6	1.24	10.3	83.3	1.19	9.9	78.2	1.00	8.3	77.0
4	0.61	5.1	91.6	0.63	5.3	92.7	0.69	5.7	84.0	0.93	7.7	84.7
5	0.43	3.6	95.2	0.49	4.1	95.7	0.53	4.4	88.4	0.76	6.3	91.0
6	0.23	1.9	97.1	0.37	3.1	95.8	0.50	4.2	92.6	0.41	3.4	94.4
7	0.15	1.2	98.4	0.18	1.5	97.3	0.35	2.9	95.5	0.30	2.5	96.9
8	0.07	0.6	98.9	0.15	1.2	98.6	0.29	2.4	97.9	0.16	1.3	98.2
9	0.06	0.5	99.5	0.08	0.7	99.3	0.14	1.2	99.1	0.12	1.0	99.2
10	0.04	0.4	99.8	0.04	0.4	99.6	0.07	0.6	99.7	0.06	0.5	99.6
11	0.01	0.1	100.0	0.03	0.3	99.6	0.03	0.2	99.9	0.03	0.3	99.9
12	0.01	0.04	100.0	0.01	0.1	100.0	0.01	0.1	100.0	0.01	0.1	100.0

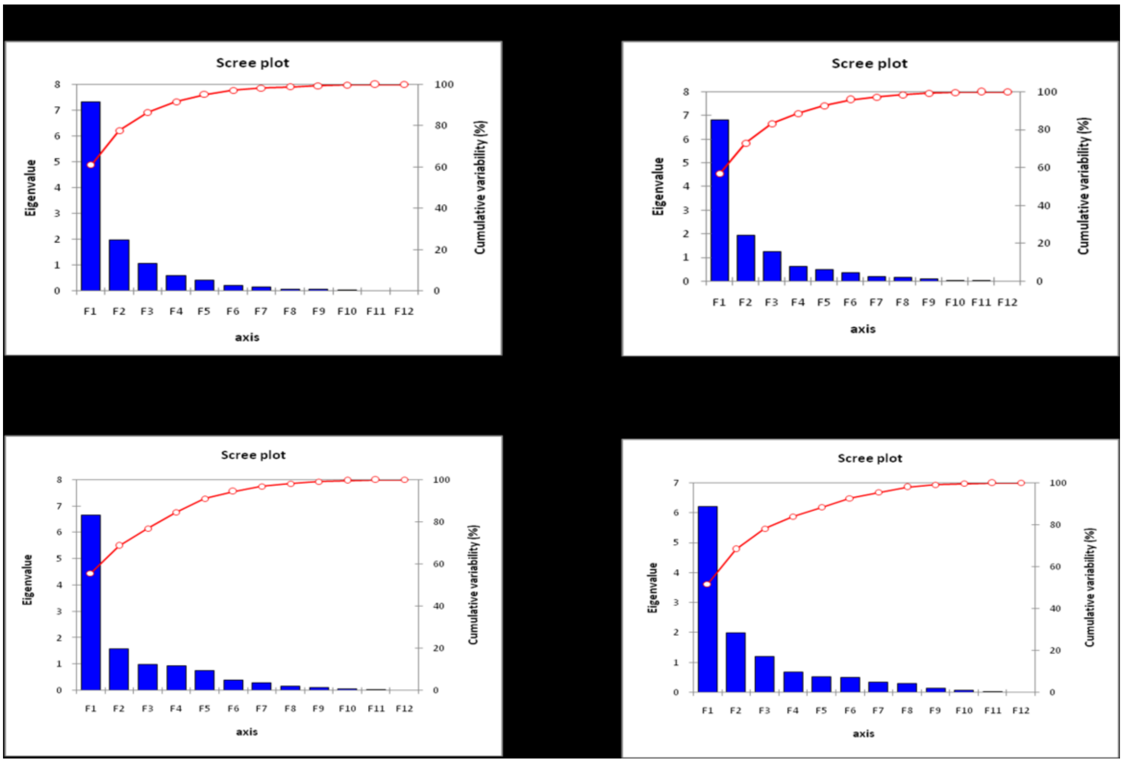

Fig 1. Scree plots of values of principal component analysis during different seasons of 2012

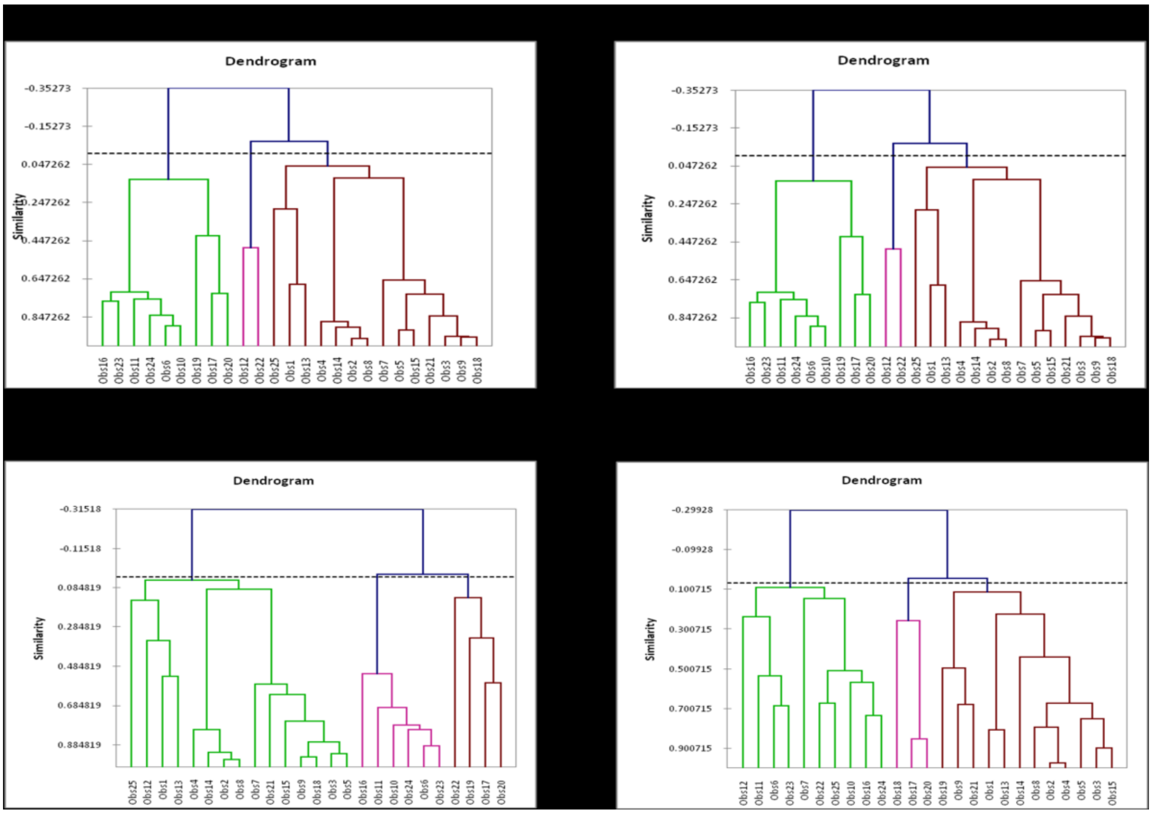

Fig 2. Dendrogram showing clustering of sampling sites