Using Multivariate Cluster Analysis Approach in Quality Assessment of Shallow Groundwater underlying a large Irrigation Project at Al-Fadhli Eastern Province Saudi Arabia

Mahbub Husain[1], Syed Munaf Ahmed[2] and Walid Abderrahman[3]

Abstract: A multivariate statistical technique Cluster analysis was used to assess the shallow groundwater quality. The main idea behind using this technique was to accommodate all the water quality variables in quality assessment including trace elements and some other ions which are not considered to be either major or minor. Conventional techniques for water quality assessments like; Stiff and Piper diagrams can only consider major and minor ions and because of increasing anthropogenic effects on groundwater. Water quality is no more just a major and minor ions study. This quality assessment study was carried out by collecting 34 samples of groundwater (17 samples each in summer and winter) at SHADCO irrigation project located in eastern province of Saudi Arabia. The collected samples were analyzed for: pH, TDS, Conductivity and alkalinity, sulphate, chloride, bicarbonate, nitrate, phosphate, bromide, fluoride, calcium, magnesium, sodium, potassium, arsenic, boron, copper, cobalt, iron, lithium, manganese, molybdenum, nickel, selenium, mercury and zinc. Cluster analysis in both Q and R mode was used. Q-mode analysis resulted in three water types for both summer and winter season. Q-mode analysis also showed the spatial as well as temporal variation in water quality. While cluster analysis in R-mode lead to the conclusion that there are two major sources contaminating the shallow groundwater in the area.

1. Introduction

Since World War II, there has been a rapid rise in nonpoint sources of groundwater pollution i.e. pollution entering the system over a wide area. Agriculture is now the dominant source of nonpoint pollution of groundwater (Jones, 1997). The commonly found contaminants in groundwater due to agriculture are nitrate, chloride, sodium, calcium, magnesium, ammonia, phosphate etc. and trace elements (George *et al.* 1987; Burkhart and Koplin, 1992; Fetter, 1992; Spalding and Exner, 1993; Beke *et al.* 1993; Kelly, 1997).

In the case described here, the quality of shallow groundwater aquifer at SHADCO irrigation project is studied. This area is located in the eastern province of Saudi Arabia, in the area which is considered to be economically important in eastern Saudi Arabia (Rasheeduddin 1999). Irrigated agriculture in the study area started in 1985 with deep wells drilled in Umm-Er-Radhuma aquifer to extract water for irrigation. Several crops including wheat, barley, Alfa-Alfa and rhodus are cultivated in these fields. Soil type at SHADCO is representative of the soil in the Saudi Arabia. Sand is 90% of the soil constituents. Since such soil has very low water-retaining property frequent irrigation is needed to keep the soil moist. Excessive irrigation at the study area has lead to the formation of shallow groundwater aquifer and subsequent water logging in the low lying areas. Another major cause for formation of this shallow aquifer is impermeable marl layer underlying the entire project area (personal communication with SHADCO hydrogeologists, Al-Assar 1992) and most of the eastern province (ElSiddig, 1999). Shallow aquifer constitutes 90% sand (SHADCO 1989, Al-Assar 1992) and spread beneath whole of the irrigation project. Most hydrogeological investigations in the area have been carried out on the relatively deep and very deep aquifers (Neogene, Dammam, Umm-er-

[1] Earth Sciences Department, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.
[2] Water Resources Division, Abunayyan Group, Al-Khobar, Saudi Arabia.
[3] Water Resources Division, Center for Environment and Water, Research Institute, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.

Radhuma and Wasia) by various workers and consultants. The well known among them are: ITALCONSULT (1969), BRGM (1976), GDC (1980) and Abderrahman and Rasheeduddin (1994). Whereas the shallow aquifers have been scarcely studied. As a consequence both the hydrogeochemical character and hydrogeological behavior of shallow aquifer is not well known. This aquifer being highly permeable makes them highly risky for the pollution of relatively deep aquifers (Munaf, 1999). It should also be noted that high concentrations of trace elements in shallow groundwater pose threat to agricultural production and health of humans and animals (Deveral and Fujii 1989).

Many different sources and processes can be attributed for the contaminants resulting in groundwater. Detailed hydrochemical research is needed to evaluate the different processes and mechanisms involved (Helena *et al.* 1999). Conventional techniques like Stiff and Piper can only consider major and minor ions to investigate chemical quality of the aquifer. In view of limitations of these methods and increasing number of chemical parameters now being measured in groundwater studies. Wider ranging statistical analysis of data is needed (Ashley and Lloyd, 1978). Dalton and Upchurch, 1978; Usunoff and Guzman, 1989, Olmez *et al*, 1994 and Helena *et al*, 1999, have shown Multivariate statistical techniques can be a very useful tool for interpretation of water quality data.

The aim of this paper is to assess the use of cluster analysis in classifying waters and also to identify the main sources (natural and anthropogenic) responsible for the hydrochemistry of the shallow groundwater aquifer so that corrective and preventive methods can be implemented. Whereas with water classification, water re-use at the study area can be investigated.

2. The Study Area and Sampling and Analysis

The study area is located at Al-Fahdley in the Eastern Province of Saudi Arabia. The study area is a part of a larger area owned by Al-Sharqiya Agricultural and Development Company (SHADCO). The whole area owned by SHADCO is bounded on the north by latitude 26^{o}50'31.15" N, on the south by latitude 26^{o}34'34.91" N, on the east by longitude 49^{o}16'15.39" E, and on the west by longitude 48^{o}59'58.23" E (Figure1). The study area ranges around 10 kilometers from west to east and 9 kilometers from north to south.

To detect the maximum possible chemical variations. The study was carried out by collecting two batches of samples. One during the summer dry period, the other during the winter period was collected. The sampling was conducted at the end of the dry season (Sept-Oct 1998), when maximum concentrations were expected, and at the end of the rainy season (Feb-March 1999), when maximum dilution was anticipated. Seventeen samples were collected for both summer and winter. Figure 2 shows the location of sampling points. Methods describe in Standards Methods for the Examination of Water and Wastewater, 1985, 16th, edition, American Public Health Association (APHA), American Water Works Association (AWWA), Water Pollution Control Federation (WPCF) were followed. Stoppered Polyethylene bottles were used. The collected water samples were refrigerated at 4^{O}C and were analyzed as soon as possible. Conductivity, temperature and pH were measured in-situ using a portable water checker.

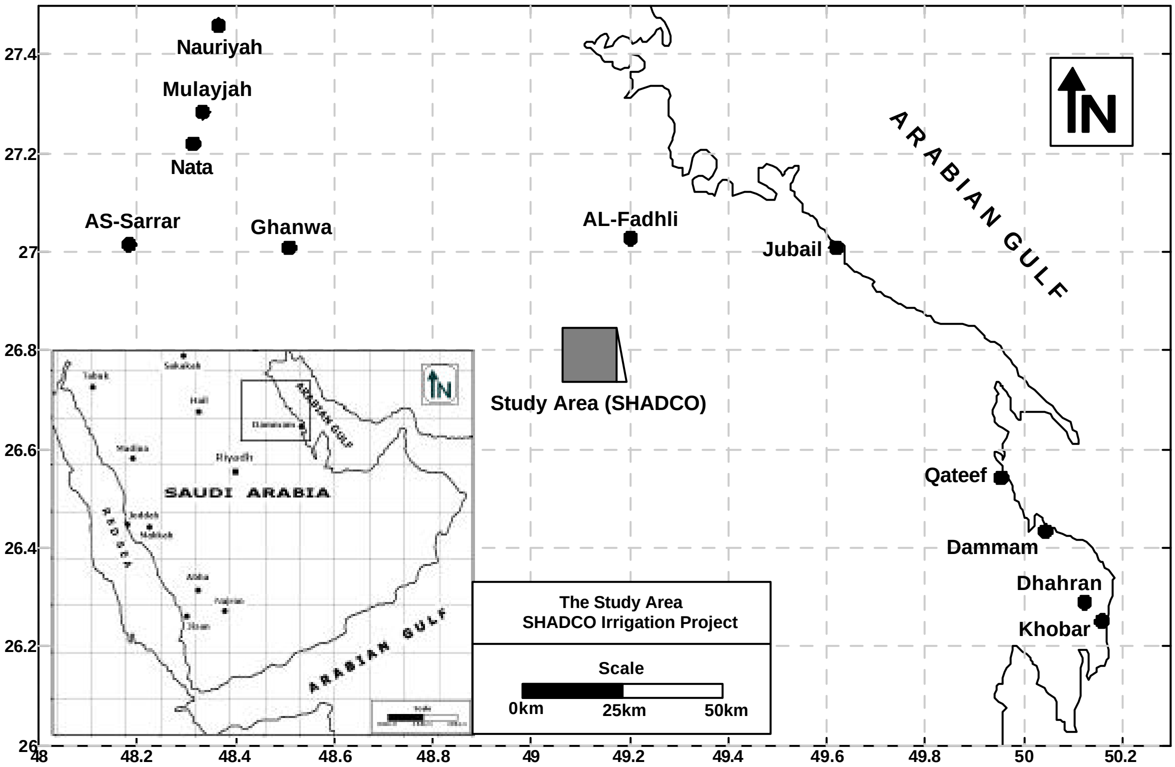

Figure 1: Location Map of the Study Area

The water samples were analyzed for major and some of the minor ions, trace elements. Anions analyzed include sulphate, chloride, bicarbonate, nitrate, phosphate, bromide and fluoride; cations include calcium, magnesium, sodium, potassium; trace elements include arsenic, boron, copper, cobalt, iron, lithium, manganese, molybdenum, nickel, selenium, mercury, and zinc. Anions were quantified using Ion Chromatography. Bicarbonates were determined by titration. Mercury was quantified using cold vapor AAS, cations were determined by ICP-AES.

3. Q-mode Cluster Analysis

To see the major water types and any temporal change in water quality. The data was analyzed in Q-mode in order that similarities between specimens or samples could be discovered (Ashley and Lloyd, 1978,) rather than similarities between variables. Sanchez *et al*, 1987 have used Q-mode technique to identify the sources of airborne particles. They used *K*-means algorithm since it is considered to be most effective in larger data sets. Here in our case we have used Q-mode to assess water types or different areas with respect to parameters in water (Chandrasekhram and Berner, 2001; Grande *et al.* 2003). Cluster analysis was performed on samples collected in summer and winter separately by using the statistical package "STATISTICA". Weighted pair-group method was applied and Euclidean distance was chosen as a measure of similarity. Cluster analysis revealed three distinct groups or clusters in summer as well as in winter samples as shown in tree diagrams or dendrograms of Figure 3 and Figure 4. Table 2 also shows the cluster number and its members for both the seasons.

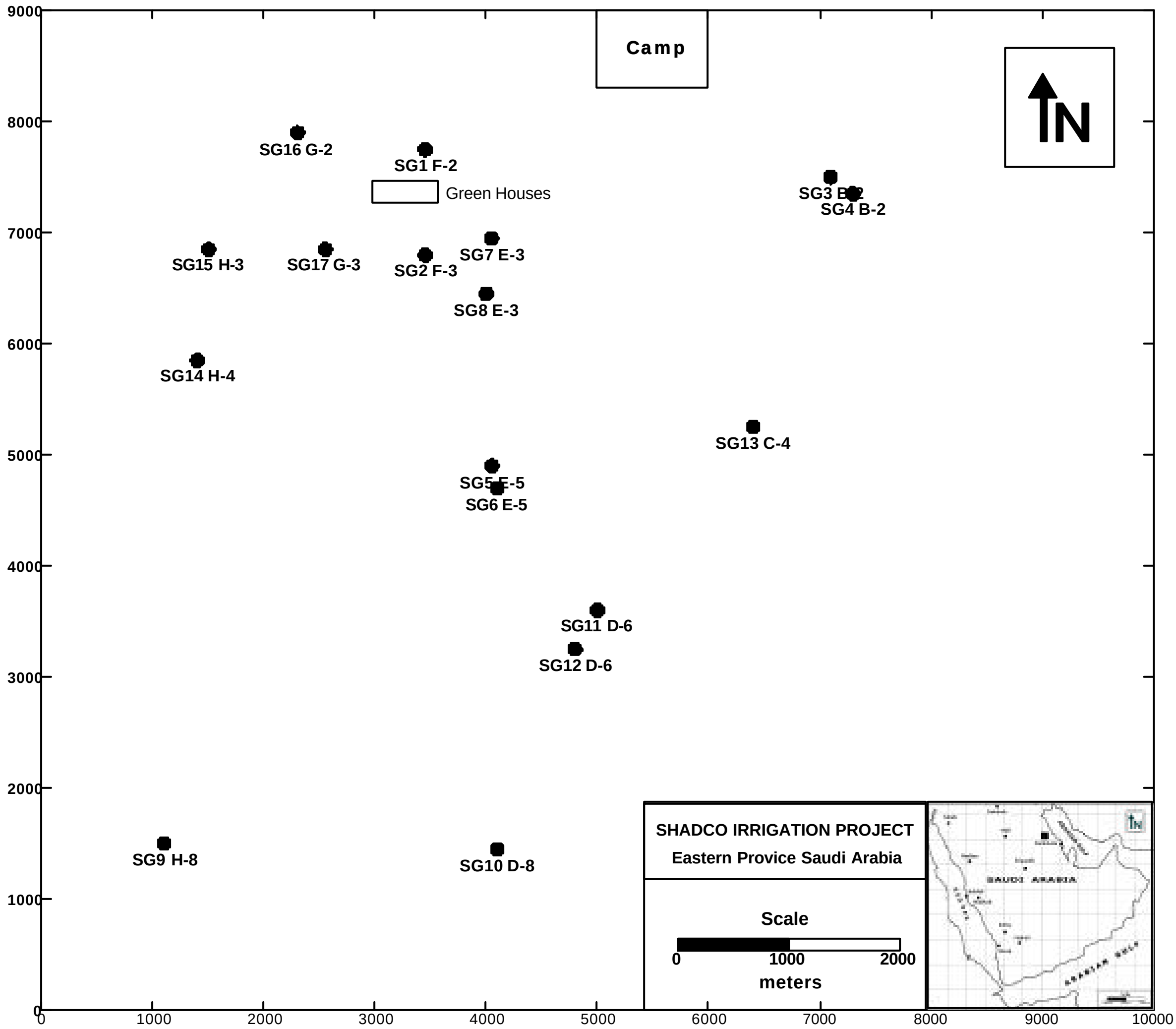

Figure 2: Map showing the location of sampling points at the study area.

Cluster analysis can be used to group the commonly collected water quality data, where each cluster indicates the water of particular quality. Clusters of samples mentioned in the Table 2 indicate that each cluster has a water quality of its own which is different from other clusters. If the example of TDS is taken for summer, cluster 1 includes: SG9-H8, SG10-D8, SG13-C4, SG16-G2, where the TDS concentration ranged from 3000-6000 mg/L which is the characteristic brackish water. Cluster 2 includes the samples: SG1-F2, SG2-F3, SG3-B2, SG4-B2, SG5-E5, SG6-E5, SG7-E3, SG8-E3, SG11-D6, SG12-D6, SG15-H3, where the TDS mostly ranged from 20000 mg/L-30000 mg/L which is the characteristics of moderate saline water and cluster 3 includes samples: SG14-H4, SG17-G3, where the TDS ranged from 35000 mg/L-50000 mg/L which is the characteristic of highly saline water. Similarly, by considering the other water quality variables (major ions. Minor ions and trace elements) water quality or chemistry associated with each cluster can be assessed in detail (section 4.1.1 and 4.1.2).

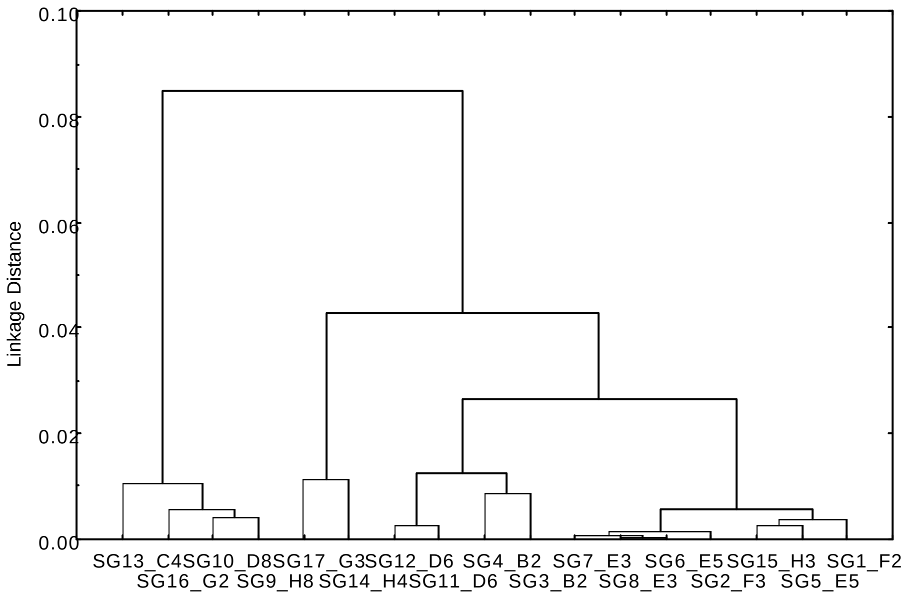

Figure 3: Tree diagram for water samples collected in summer

After dividing the samples into three different clusters and by considering the samples and their location for each cluster, zones were developed (I, II, and III) where each zone corresponds to its respective cluster, this implies that each zone has a water quality which is different from other zones. Figure 5 and Figure 6 shows the distribution of clusters or zones or types of water quality at the study area.

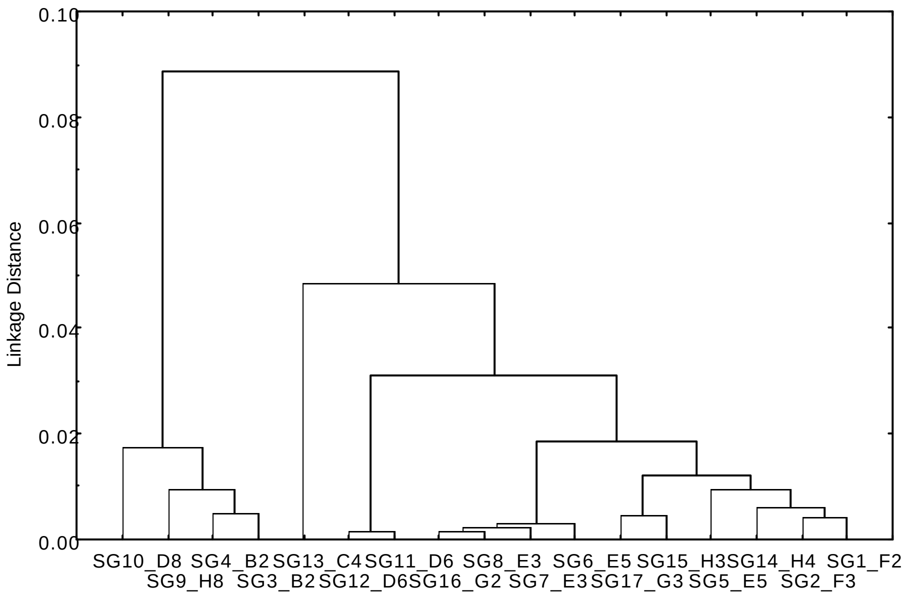

Figure 4: Tree diagram for water samples collected in winter

Table 1: Clusters and their members for summer and winter samples in Q-mode

Cluster Number	Members
Summer	
1	SG9-H8, SG10-D8, SG13-C4, SG16-G2,
2	SG1-F2, SG2-F3, SG3-B2, SG4-B2, SG5-E5, SG6-E5
	SG7-E3, SG8-E3, SG11-D6, SG12-D6, SG15-H3
3	SG14-H4, SG17-G3
Winter	
1	SG3-B2, SG4-B2, SG9-H8, SG10-D8
	SG1-F2, SG2-F3, SG5-E5, SG6-E5, SG7-E3, SG8-E3
2	SG14-H4, SG15-H3, SG16-G2, SG17-G3
3	SG11-D6, SG12-D6, SG13-C4

3.1. Water Types in summer

3.1.1. Water type-I

This water is basically chloride and sulphate dominated, calcium and sodium also exists, but relatively in low concentration. Bicarbonate is the lowest. As far as trace elements are concerned, this water has the lowest concentrations of most of the trace elements when compared with other waters of the study area. Boron being the highest while copper was the lowest and it also has lithium and manganese but relatively in low concentration. This water type has highest manganese concentration. Areally, this water type is found in north-western, central-east and southern part of study area (please see Figure 5).

3.1.2. Water type-II

Water type-II is sodium and chloride dominated, it also has sulfate, calcium but in relatively low concentration. Again, bicarbonate is the lowest concentration. Among trace elements, boron concentration is highest and manganese is lowest. This water is found in the central and north-eastern part of the study area (please see Figure 5).

3.1.3. Water type-III

This water is sodium and chloride dominated, it also has sulfate, calcium and magnesium but in relatively lower concentrations. Bicarbonate being the lowest. Based on the major ion concentrations this water is similar to water type-II. However, because of high concentration of nitrate and some trace elements, this water was considered as different type. In addition this water has highest nitrate and arsenic concentrations when compared with other water types of the study area. Water of this type is found in central and north-western part of the study area (please see Figure 5).

3.2. Water Types in winter

3.2.1. Water type-I

This water is chloride and sulphate dominated. It also has sodium and calcium but in relatively lower concentrations, and bicarbonate was lowest. Among trace elements boron was highest and manganese was the lowest. Most of the trace elements were in the lowest concentrations in this water type when compared with other types of the study area. This water was found in southern and north-eastern part of the study area (please see Figure 6).

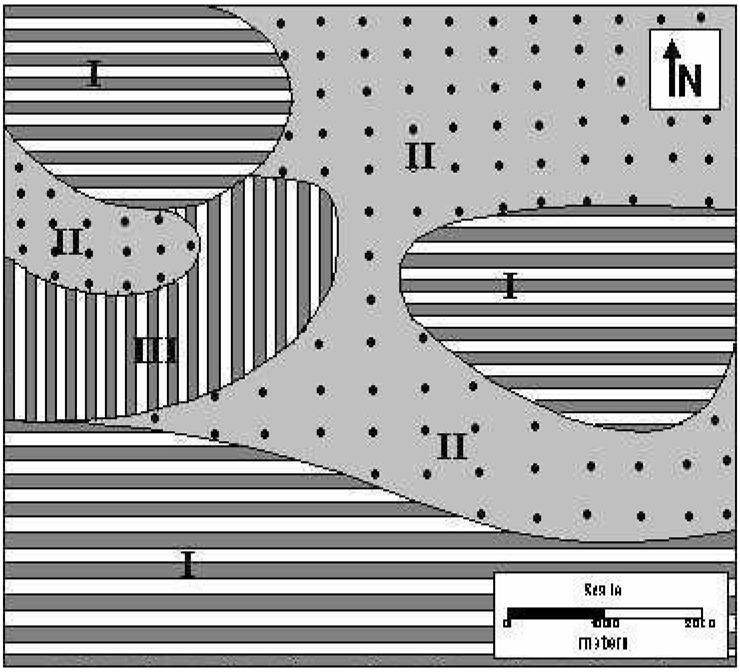

Figure 5: Distribution of clusters from tree diagram of Figure 3

3.2.2. Water type-II

Water of this type is sodium and chloride dominated, it also has relatively lower concentrations of sulfate and calcium. Bicarbonate was the lowest. For trace elements, this water type has the highest concentrations of most of the trace elements when compared with other water types. Boron was the highest and zinc was the lowest. This water type is found in central and north-western part of the study area (please see Figure 6).

3.2.3. Water type-III

This water is chloride and sulphate dominated, it also has sodium and calcium but in relatively lower concentrations. Bicarbonate was the lowest. This water type based on major ions would be of type-I, because of trace elements and nitrate concentrations it was treated as different type. Among the trace elements, boron was the highest and manganese was the lowest. This water has the lowest concentration of nitrate. Water of this type is found in central and south-eastern part of the study area (please see Figure 6).

3.3. Assessing Temporal Changes in Water Quality

Water in north-western part of the study area in summer was basically a chloride and sulfate dominated but in winter it was found to be sodium and chloride dominated. The reasons for chloride replacing sulfate in this area could be to due to huge discharge of water from greenhouses located in the Field F2 may have caused the existing salts in soil to dissolve thus increasing the chloride concentration. Similarly in north-eastern part, in summer it was chloride and sulfate dominated water but in winter it changed to sodium and chloride dominated. The reasons for this could be the same dissolution process as mentioned earlier. In the rest of the places their is no significant temporal change in water quality. Please refer figures 5 and 6.

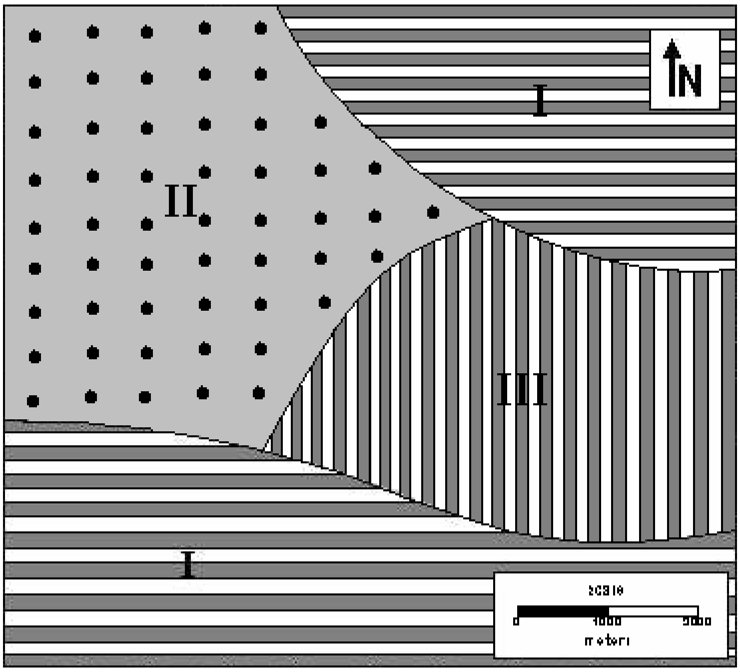

Figure 6: Distribution of clusters from tree diagram of Figure 4

4. R-mode Cluster Analysis

To investigate the group associations or to see the affinity between the various water quality variables, data was analyzed on R-mode so that their sources and the processes with which they are associated could be determined (Blifford and Meeker, 1967, Hopke *et al*, 1976, Gaarenstroom *et al*, 1977, Rapin, 1980). Cluster analysis was performed on pH, TDS, alkalinity, conductivity, calcium, magnesium, sodium, potassium, chloride, sulphate, bicarbonate, nitrate, bromide, fluoride, phosphate, boron, copper, iron, lithium, manganese, mercury, molybdenum, nickel, selenium, zinc, cobalt, arsenic. Weighted pair-group method was applied and Euclidean distance was used as the measure of similarity. Before the analysis could be done, the data was normalized by generating a correlation coefficient matrix. Cluster analysis revealed two distinct groups or clusters for summer as well as for winter data (Figure 7 and Figure 8). From above figures, different clusters and their members were extracted as follows: cluster 1 includes: bicarbonate, alkalinity, nitrate, bromide, fluoride, calcium, potassium, lithium, sulphate, sodium, boron, chloride, TDS and conductivity. Cluster 2 includes: arsenic, copper, phosphate, nickel, molybdenum, manganese, iron. Whereas for the winter data cluster 1 comprised of the following variables: phosphate, bromide, magnesium, potassium, calcium, sulphate, boron, lithium, sodium, chloride, TDS, conductivity, nitrate, fluoride, and cluster 2 was found to contains:

cobalt, copper, alkalinity, bicarbonate, manganese, molybdenum, zinc. (Please see Table 3)

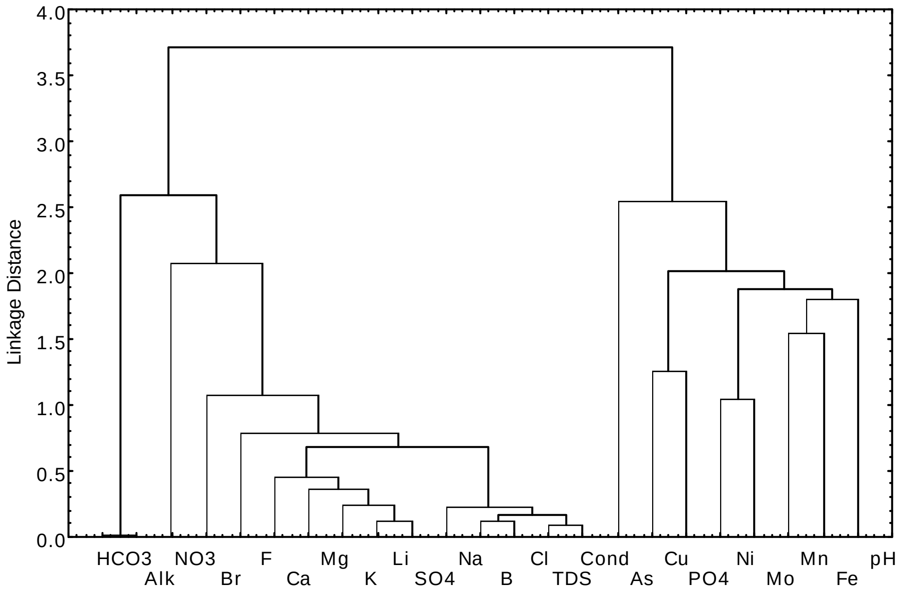

Figure 7: Tree diagram for 23 variables for the summer data

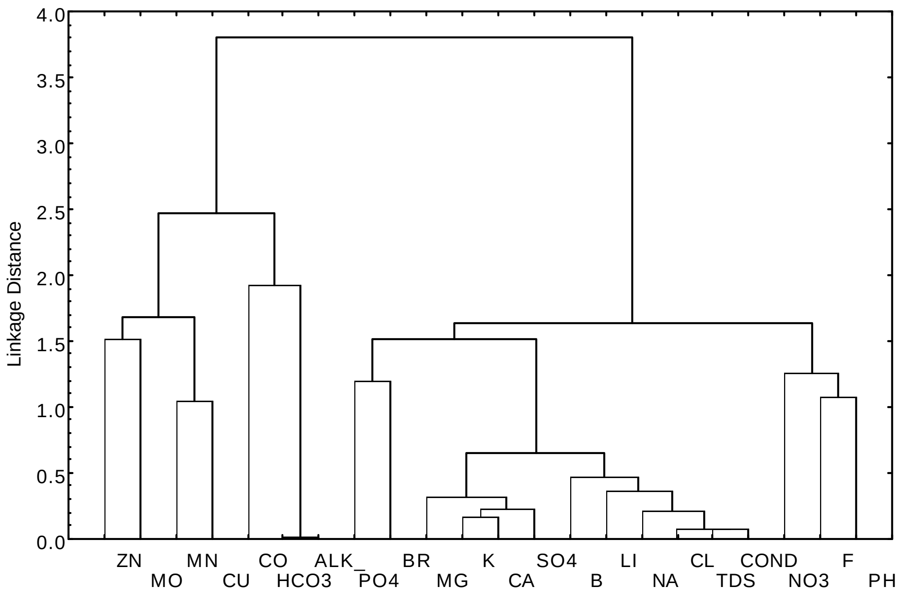

Figure 8: Tree diagram for 23 variables for the winter data

Data analysis in R-mode for summer and winter resulted in two clusters each as discussed earlier. Figure 7 and Figure 8 shows the clusters for both the seasons from chemical analysis of waters of the study area. It is evident that in both graphs their is a clustering of ions (Ca, K, Mg, Na, SO4, NO_3, Br, F, Cl, B and HCO_3) and other variables in one group and trace elements in another group (Table 3) The above said Figures also suggests that their are basically two pollutant poles: ions pole and trace elements pole. The term “pole” was used by Rapin, 1980 in his study to demonstrate the anthropogenic effects on sediment from Bay of Nice and Villefranche-sur-Mer on the French coast. He used cluster analysis and came up with two poles (clay pole and pollutant pole). Here in our case the source for trace elements most probably is due to the application of trace elements as micronutrients in the study area, but the other pole which consists of all ions simply cannot be attributed to a single source, their could be many natural sources and anthropogenic processes controlling this pole. To get a clear idea about the sources, a sequential extraction method called factor analysis has to be performed, which is out of scope of the present work.

Table 2: Clusters and their members for summer and winter samples in R-mode

Cluster Number	Members
Summer	
1	HCO3, Alkalinity, NO3, Br, F, Ca, K, Li, SO4, Na, B, Cl TDS and Conductivity
2	As, Cu, PO4, Ni, Mo, Mn, Fe
Winter	
1	Br, PO4, Mg, K, Ca, SO4, B, Li, Na, Cl, TDS, Conductivity, NO3 and F
2	Co, Cu, Mn, Mo, Zn, HCO3 and alkalinity

5. Conclusions

1. The agricultural activities have resulted in release of toxic metals like arsenic, copper, cobalt, molybdenum, and zinc in waters of the study area and TDS level was also found to be very high.

2. Agricultural impacts were found to severe in winter than summer as expected. As winter is irrigation period at the study area when fertilizers and pesticides are in huge quantities.

3. Cluster analysis in Q-mode resulted in three major water types for both the seasons.

4. In Q-mode cluster analysis, when overall chemistry of waters of summer and winter were compared, showed areal change in water quality.

5. Cluster analysis in R-mode resulted in two pollutant poles: ions pole and trace elements pole.

Acknowledgments

We gratefully acknowledge the support of King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia. Special thanks to Department of analytical chemistry, Research Institute, KFUPM for helping us in water analysis. SHADCO irrigation project management is to be thanked for their assistance in field work and also for allowing us to enter their property. Thanks are also due to water resources division, Abunayyan group, Al-Khobar, Saudi Arabia.

References

Abderrahman, W. A. and M. Rasheeduddin (1994). Future groundwater conditions under long-term water stresses in an arid urban area. *Water resources management. 8: 245-264.*

Al-Assar, R.S (1992). Numerical Simulation of Groundwater in the Umm-Er-Radhuma Aquifer at SHADCO Project, Eastern Province Saudi Arabia. *M.S Thesis, King Fahd University of Petroleum and Minerals.*

Ashley, R.P and Lloyd, J.W. (1978). An example of the use of factor analysis and cluster analysis in groundwater chemistry interpretation. *J. Hydrology 39: 355-364.*

Beke, G.J., Entz, T and Graham, D.P. (1993). Long Term quality of shallow groundwater at irrigated sites. *Journal of Irrigation and Drainage Engineering 119: 8-23.*

Blifford, I.H., Jr and Meeker G.O. (1967). A factor analysis model of large scale pollution. *Atmospheric Environment 1, 147-154.*

BRGM. (1976). Hydrogeological investigations of wasia aquifer in the eastern province of Saudi Arabia, regional study, final report and appendices. *Ministry of agricultural and water, Riyadh, Saudi Arabia.*

Burkhart, R.M and Kolpin, D.W. (1992). Hydrologic and land-use factors associated with Herbicides and Nitrates in near surface aquifers. *Journal of Environmental Quality 22: 646-656.*

Chandrasekhram. J.K.D and Burner. D.S.Z. (2001). Arsenic contamination in groundwater, Murshidabad district, West Bengal. *WR-10, proceed. Villasimus, Italy, June 2001.*

Dalton. M.G and Upschurch. SB (1978). Interpretation of hydrochemical facies by Factor Analysis. Groundwater 16: 228-233.

Deveral S.J and Fujii R. (1989). Chemistry of trace elements in soils and groundwater; *in Agricultural Salinity Assessment and Management ASCE. 64-90.*

ElSiddig. A.H.E. (1999). A study of Groundwater Management in a Confined Arid Zone Aquifer Al-Fadhli, Saudi Arabia. *Dissertation PhD. School of Earth Sciences, University of Birmingham.*

Fetter, C.W. (1992). Contaminant Hydrogeology. *Macmillan Publishing Co. New York.*

Gaarenstroom P.D., Perone S.P. and Moyers J.L. (1997). Application of pattern recognition and factor analysis for characterization of atmospheric particulate composition in southwest desert atmosphere. *Envir. Sci. Technol. 8, 795-800.*

Groundwater Development Consultants (GDC). (1980). Umm-er-Radhuma study: Bahrain assignment: Demeter house, station road, Cambridge, CBI 2RS, unpublished report to Ministry of agricultural and water, Riyadh, Saudi Arabia.

George, V.S., Bouwer, H, and Peter, J. W. (1987). Irrigation effects in Arizona and New Mexico. *Journal of Irrigation and Drainage Engineering 113: 30-48.*

Grande, J. A, Borrego. J, Torre, M.L. D and Sainz A. (2003). Application of cluster analysis to the geochemistry zonation of the estuary waters in the tinto and odiel rivers (Huelva, Spain). Environmental Geochemistry and Health 25: 233-246.

Helena, B.A, Vega. M, Barrado. E, Pardo. R and Fernandez. L. (1999). A case of hydrochemical characterization of an alluvial aquifer influenced by human activities. *Water Air and Soil Pollution 112: 365-387*

ITALCONSULT (1969). Water and agricultural development studies for area-IV, eastern province, Saudi Arabia: unpublished report to Ministry of agricultural and water, Riyadh, Saudi Arabia.

Jones, J.A.A. (1997). Global Hydrology. *Addison Wesley Longman Ltd., Essex.*

Kelly, W.R. (1997). Heterogeneities in groundwater geochemistry in a sand aquifer beneath an agricultural field. *Journal of Hydrology 22: 589-596.*

Munaf, S. A. (1999). Quality assessment of irrigated water at SHADCO irrigation project at Al-Fadhli, Eastern Province Saudi Arabia. *M.S. Thesis, King Fahd University of Petroleum and Minerals.*

Olmez, L, Jack W.B, and J.F. Villaume. (1994). A new approach to understanding multiple-source groundwater contamination: Factor analysis and chemical mass balance**.** *Water Research 28: 1095-1101.*

Rasheeduddin. M. (1999). A study of Groundwater resource development study of a regional arid zone multi-aquifer system: Eastern province, Saudi Arabia. *Dissertation PhD. School of Earth Sciences, University of Birmingham.*

Rapin. F. (1980). Dissertation, University of Geneva.

Sadiq.M and Alam. I. (1997). Metal Concentration in a Shallow Groundwater Aquifer underneath Petrochemical Complex. *Water Research 26: 3089-3096.*

SHADCO (1989). A Report on Analysis of Soil and Water Samples from SHADCO Irrigation Project. *Kingdom of Saudi Arabia, Ministry of Agriculture and Water.*

Spalding, R.F, and Exner, M.E. (1993). Occurrence of Nitrate in Groundwater - A Review. *Journal of Environmental Quality 22: 392-402.*

Standard Methods for Examination Water and Waste Water Analysis. (1985). *APHA, AWWA, WPCF. 16th edition.*
STATISTICA 5.0 for Windows, *StatSoft, Inc. Tulsa OK.*
SURFER (W32), version 8.0. *Surface Mapping Systems, Golden Software.*
Usunoff, E.J and Guzman, A. (1989). A multivariate analysis in hydrochemistry: An example of the use of factor and correspondence analyses. *Groundwater 27: 27-34*.