Science and Technology 2013, 3(4): 99-104
DOI: 10.5923/j.scit.20130304.01

Multivariate Data Analysis to Identify the Groundwater Pollution Sources in Tulkarm Area / Palestine

Nawaf Abu-Khalaf, Saed Khayat[*], Basel Natsheh

Technical and Applied Research Centre (TARC), Palestine Technical University (PTUK) -Tulkarm, Palestine

Abstract The multivariate data analysis is used to analyse groundwater samples from 13 different wells along a period of 13 years. The results show that the most significant pollutants for the groundwater resources in the region are Cl and Na which are come mainly from the Wadi Zomer stream as the main source. HCO_3, show relatively low concentration and restricted to those wells which have low abstraction rates and, slightly effected by the pollutants, and received a good replenishment of. Ward's method was used for cluster analysis. It managed to classify the wells into three groups, according their geochemical and locations characteristics. Some wells were clustered near each other, since they share the same effects of the surrounded environment. The geological formations and the layers that water taped from play also a role in the water quality distribution. The results emphasize that there is a pollution-dilution process that the groundwater undergoes between a freshwater end-member from the upper Cenomanian Turonian replenished water that prevailed more HCO_3 content, and the polluted water end member that contains high Na/Cl.

Keywords Multivariate Data Analysis, Principal Component Analysis, Hierarchical Cluster Analysis, Ward's Method, Ground Water, Pollution, Quality, Palestine

1. Introduction

Water is considered as one of the most important natural resources in the Middle East region. There is a scarcity of sufficient water, due to population growth, economic and agricultural development, and an arid climate. This is a great challenge for this region. Moreover, this scarcity of water resources in this region is unfortunately combined with rapid fresh water resources quality deterioration, due to salinity and contamination processes. Understanding of the origin and mechanisms of the contamination process is a crucial to solve and manage this problem[1].

One of the methods in water quality assessment that has become widely applied over the last ten years is multivariate data analysis (MVDA). The former was used for applications related to water, e.g. evaluation and interpretation of ground water quality[2-4], providing insight into the hydrochemical processes in coastal aquifers [5,6], possible sources of pollution/polluting processes and identifying critical water quality issues[7-10], and interaction of river and water/groundwater and groundwater mixing[11].

Almost in all mentioned references above, two MVDA techniques were applied, i.e. principal component analysis (PCA) and hierarchical cluster analysis (HCA). These techniques were generally used to identify and quantify quality parameters and pollutants sources.

PCA is one of the unsupervised linear techniques in MVDA. It is performed to reduce a large data set of variables into few new linear uncorrelated (i.e. orthogonal) variables factors called principal components (PCs). PCs are also called latent variables. These factors can be interpreted to reveal underlying data structure. The first principal component (PC1) accounts for the maximum possible proportion of the total variance in the data set and the second component (PC2) accounts the maximum of the remaining variance and so on. The maximum number of PCs is equal to the number of variables. The total variance accounted by all the PCs will be equal to the number of variables. For interpretation, only a few numbers of PCs are retained in the analysis. PCA can be explained by scores and loading plots. Scores plot explain the relations between samples and loading plot explain the relations between variables[12].

HCA is an unsupervised pattern recognition technique, and its algorithms produce a sequence of nested partitions including similar groups. Clusters in HCA are formed sequentially, starting with the most similar pair of variables and forming higher clusters step by step. Cluster process formation is repeated until a single cluster containing all the variables are obtained. The result of the clustering can be displayed in a tree-like structure, called a dendrogram. The dendrogram can be broken at different levels to yield

[*] Corresponding author:
saed.khayat@gmail.com (Saed Khayat)

Published online at http://journal.sapub.org/scit

different clusters of the data set. However, it should be noted that the decision of the final cluster is rather arbitrary. The hierarchical agglomerative clustering methods differ in the way they calculate the similarity between two clusters, i.e. single link, complete link, group average and Ward's method. The former methods depend on calculating the similarity between two patterns using a distance measure. The most popular distance method is the Euclidean distance. The Ward's method is distinct from other methods, because it uses an analysis of variance approach to evaluate the distances between clusters. Cluster membership in this method is assessed by calculating the total sum of squared deviations from the mean of a cluster. The criterion for fusion is that it should produce the smallest possible increase in the error sum of squares. In general, this method is very efficient and less sensitive to outliers. The Ward's method with squared Euclidean distance used as a dissimilarity measure has been found to provide meaningful dendrogram of clusters with the proximity or similarity of clusters measured with a rescaled distance[13, 14].

PCA and HCA have been successfully used in earlier studies for evaluation and interpretation of groundwater quality data set[2, 7, 8, 15-17].

The objective of this paper was to identify pollution sources in the groundwater in Tulkarm area / Palestine using MVDA, and the significant variables that cause the variability in the groundwater quality, mainly the geological formations and the rate of abstraction. PCA loading plot was used to interpret the most important quality parameters. Ward's method and squared Euclidean distance were used to classify wells according to their geochemical and location characteristics. However, it should be emphasized that this paper is related to a study that was made by Khayat et al.[1].

2. Materials and Methods

2.1. Samples

Groundwater samples were collected once a year during the last 13 years, in the period between 2000-2012. The samples were collected from 13 different wells in Tulkarm area in Palestine (Figure1). The samples were taken from wells which dogged in different geological formations. These are: Cenomanian-Turonian, Eocene and alluvial formation which cover the area of the Zomar stream path (Figure1). Most of the samples were collected in spring time each year. Standard methods were thoroughly followed for the collection, preservation and analyses of samples[18]. Samples were gathered and chemically analyzed yearly in an environmental research laboratory. Eight different water chemical parameters were analyzed i.e. Cl, NO_3, Ca, Mg, K, SO_4, Na and HCO_3[18].

2.2. Statistical Analysis

Data were analysed using MVDA. The Unscrambler software package (version 10.2, CAMO Software AS, Oslo, Norway) was used for performing PCA and HCA on averaged, standardized data set, to eliminate the effect of scale of measurement of data. The data set matrix comprises 8 water quality parameters and 169 water samples (i.e. 169 is the data of 13 wells in 13 years).

3. Results and Discussion

The PCA and HCA were used for interpretation the process of pollution in ground water.

PCA was used to identify the most important geochemical parameters, and the relations between different parameters. The first and second principal components (PC1 and PC2) explained 44% and 18% of the variation of the data, respectively. Figure 2 shows PCA loading plot along the first principal component. It can be seen the Na and Cl have the highest loading, which indicates their high contribution among the quality parameters.

Ward's method was used for HCA (Figure 3). Its goal was to cluster the geochemical parameters and to find cluster of similarity between geochemical parameters and wells.

Figures 2 suggested that the most significant elements that control the pollution are Na and Cl. In general, two significant end-members are control the water quality in the whole wells, they are HCO_3 that give the indication of fresh water quality, which are bearing the calcium carbonate rocks in one hand, and Na/Cl that indicate the pollution from the Wadi stream in the other hand and a third mixed group indicate the effect of other sources that come from other pollution sources from agricultural activities or leached from soil through direct infiltration, especially in the west part of the study area[1]. Those elements are classified in Figure 3. The whole area of Tulkarm has two geological end-members, the first one is the Karstic Hebron-Jerusalem formation which is composed mainly of limestones and dolomitic lime stones that are rich in carbonate, calcium, and magnesium. This water type is not highly susceptible for the deterioration in the eastern part of Tulkarm as it's more closed to recharge area. However, the HCA figure shows that there is a clear influence for Na/Cl that mimic with fresh water and deteriorate the water quality.

Due to the fact that the only source of NO_3 in the groundwater is come basically from the anthropogenic pollutants, the presence of the NO_3 with Na in one narrow cluster suggest strongly that most of the excess amount of Na is come mainly from the anthropogenic pollution. This figure support the finding of Khayat et al.[1] about the lateral flow of pollutants from the wastewater drained in Wadi Zomar stream. This is also applied on the Cl which has less relation with NO_3. This small deviation in this cluster is caused mainly from an additional amount of chloride that is come from the alluvial sediments that is present in the west of Tulkarm area. This finding can emphasize the presence of two different chloride sources that deteriorate the water quality in the study area, one from Wadi Zomar and the other from the alluvial sediments especially in the west.

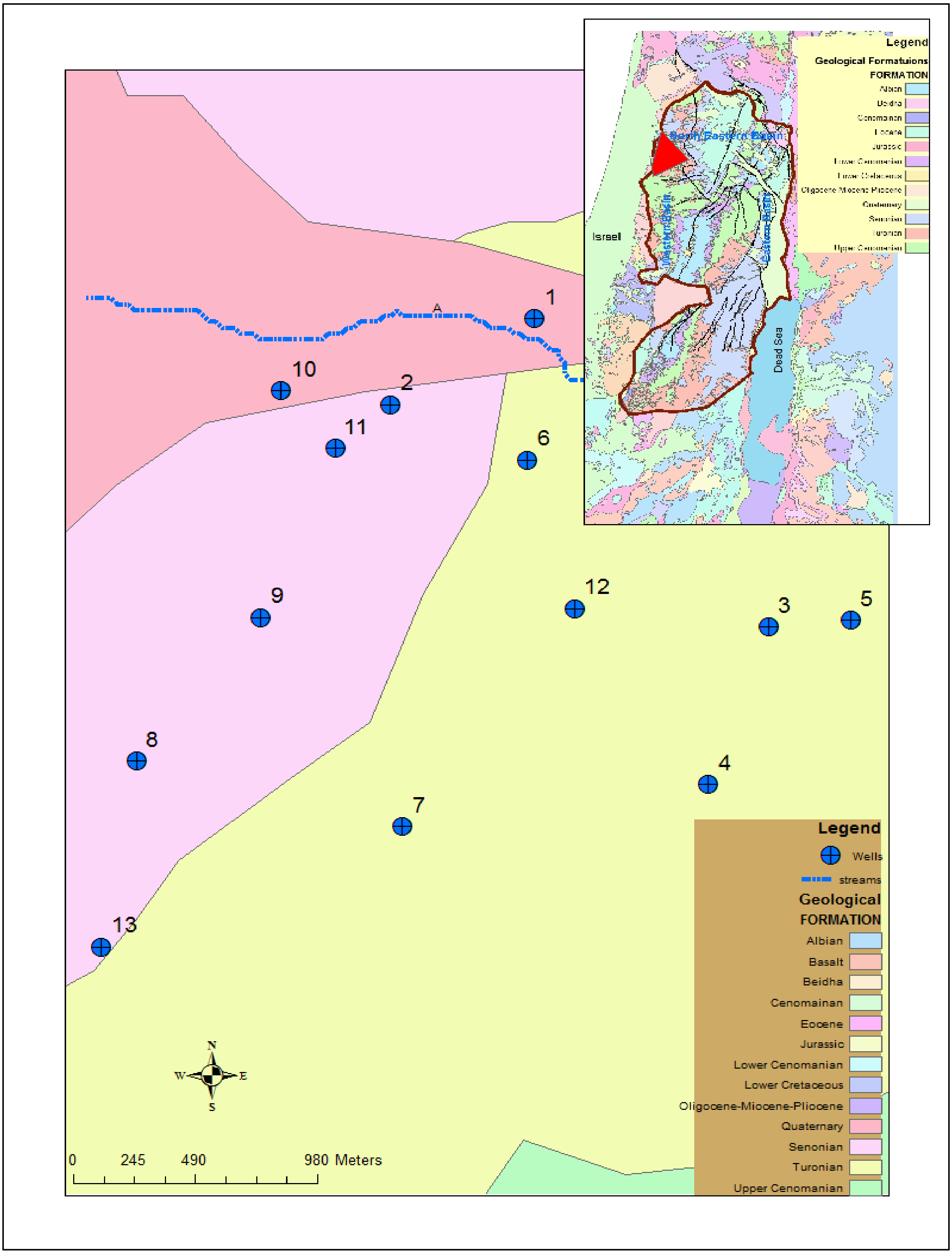

Figure 1. Study area showing the major geological formations and the sampled wells distributed in the studied area at Tulkarm / Palestine

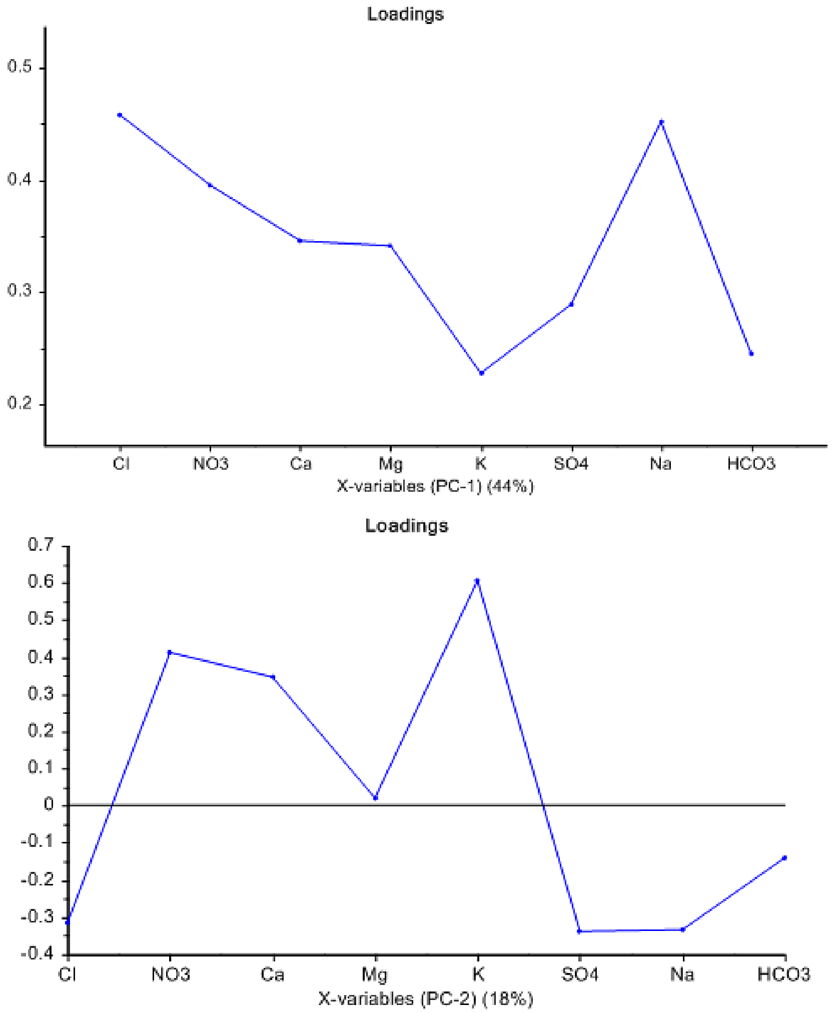

Figure 2. Principal component analysis loadings plot of geochemical parameters of water quality, along first and second principal components (PC1 and PC2), which explained 44% and 18% of variation of data, respectively

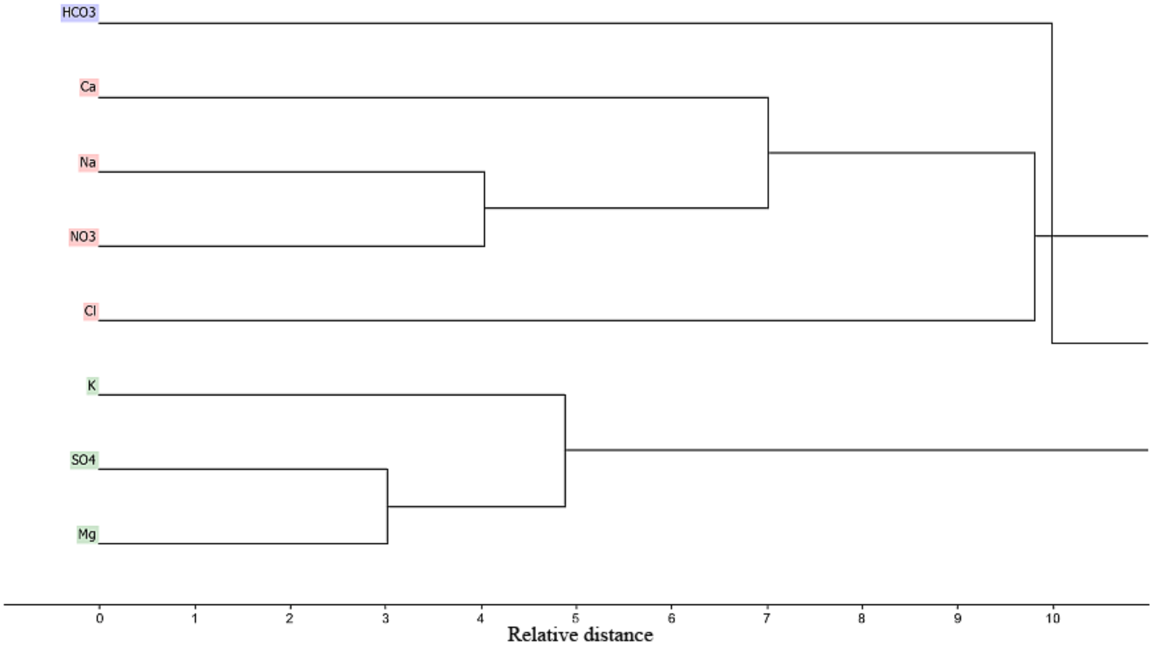

Figure 3. Dendrogram showing clustering of geochemical parameters of ground water. Three groups are presented, HCO_3 (in violet colour), Ca, Na, NO_3, and Cl (in pink colour), and K, SO_4, and Mg (in green colour)

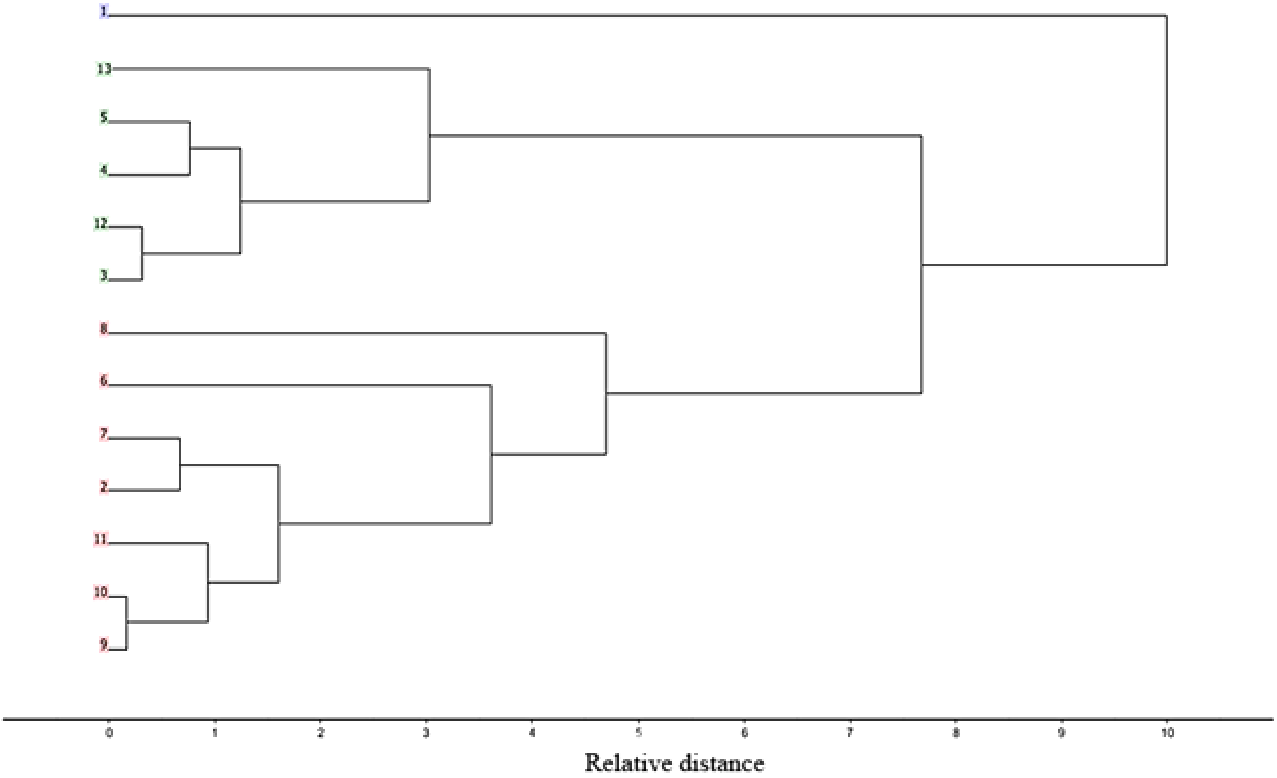

Figure 4. Dendrogram showing clustering of wells according to their locations. . Three groups are presented, well nr. 1 (in violet colour), wells nr. 3, 4, 5, 12, and 13 (in green colour), and wells nr. 2, 6, 7, 8, 9, 10, and 11 (in pink colour)

As shown in Figure 4, the Ward's diagram classified the wells into three groups according to type and degree of pollution. The first group is the group of well number 1 that has low electrical conductivity (EC) and high HCO_3 values.

This wells located to the north of Wadi stream and used only at low rate of abstraction. Moreover, the location of this well further to the upper cliff in the north promote more freshwater replenishment from the upper Cenomanian Turonian layer beneath the alluvial deposits. The other group is the group of wells number 3, 4, 5, 12, and 13. The NO_3 values for these wells are relatively higher. These additional amount of NO_3 are accompanied also with the addition of Na and Cl. This group is considered as an intermediate group that receive a continuous recharge from the recharge area. The location of those wells relatively far from the Wadi path, and the low abstraction rate, make the influence of the wastewater stream of Wadi Zomar somehow limited.

The third group of wells numbered 2, 6-11 are located either surrounding the Wadi stream, or mostly municipal wells that undergoes a continuous abstraction to meet the municipal demand in Tulkarm like wells number 7 and 8. Those wells have the highest EC, NO_3, Na and Cl values. The NO_3 values for these wells exceed in some years the amount of 100 mg/L. This increase is also shown in the values of other pollutants, and the concentration of HCO_3 are become low due to less contribution from the replenished freshwater. This behaviour is mostly noticed in the summer months where no rain occurs.

In general, the wells in Ward diagram are clustered first as a three groups according to their locations from the wastewater stream of Wadi Zomar and to the abstraction rate in one hand, and also clustered at small scale according to its location from each other, for example the wells 4 & 5, 12 & 3, 9 & 10, and 2 & 7. Those wells are relatively closed to each other and apparently tapping the water from the same layers that sharing the same effects of the surrounded environment. The geological formations and the layers that water taped from play also a role in the water quality distribution.

The results emphasize that there is a pollution-dilution process that the groundwater undergoes between a freshwater end-member from the upper Cenomanian Turonian replenished water that prevailed more HCO_3 content, and the polluted water end member that contains high Cl/Na.

4. Conclusions

The above mentioned results reflect the needs for more mitigation procedure in order to avoid more deterioration for the groundwater resources in the study area. The main causes of such deterioration were found to be as follow:

– The effect of groundwater heavy abstraction: this effect is more significant with the closeness to the Wadi stream. In this context, better management for the amount of daily abstraction most be taking into account especially in the summer time.

– The effect of different geological formations and the layers that water taped from, which also play a role in the water quality distribution. The results emphasize that there is a dilution process that the groundwater undergoes between a freshwater end-member from the upper Cenomanian Turonian replenished water that prevailed more HCO_3 content, and the polluted water end member that contains high Na/Cl. The quality of the wells that tapped freshwater from the Cenomanian Turonian layer can be easily remediate during winter time by direct replenishment. Therefore, any planning for dogging new well must take this consideration into account.

– Multivariate data analysis (MVDA) techniques, e.g. PCA and HCA, are capable of interpretation the pollution resources of ground water, using loading and dendrogram plots.

ACKNOWLEDGEMENTS

The authors would like to thank the Palestine Technical University-Kadoorie (PTUK) for providing financial support to publish this work.

REFERENCES

[1] Khayat, S., Marei, A., Natsheh, B., Abu-Khalaf, N., 2012. Mechanisms of groundwater pollutants transport in Tulkarm area / Palestine. Resources and Environment 2 (6), 281-290.

[2] Mohapatra, P. K., Vijay, R., Pujari, P. R., Sundaray S. K., Mohanty B. P., 2011. Determination of processes affecting groundwater quality in the coastal aquifer beneath Puri city, India: a multivariate statistical approach. Water Science & Technology 64 (4), 809–817. doi:10.2166/wst.2011.605.

[3] Singh, K. P., Malik, A., Mohan, D., Sinha, S., 2004. Multiva riate 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.

[4] Singh, K. P., Malik, A., Singh, V. K., Mohan, D., Sinha, S., 2009. Chemometric analysis of groundwater quality data of alluvial aquifer of Gangetic plain, North India. Analytica Chimia Acta 550 (1–2), 82–91.

[5] 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.

[6] 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.

[7] Marco, C., Gonçalves, A. M., 2012. Combining statistical methodologies in water quality monitoring in a hydrological Basin-Space and time approaches. In: Water quality monitoring and assessment, Voudouris, K. and Voutsa, D. (Ed.), 121-142. ISBN: 978-953-51-0486-5. doi: 10.5772/33867.

[8] Osei, J., Nyame, F. K., Armah, T. K., Osae, S. K., Dampare, S. B., Fianko, J. R., Adomako, D., Bentil, N., 2012. Application of multivariate analysis for identification of pollution sources in the Densu Delta Wetland in the vicinity of a landfill site in Ghana. Journal of Water Resource and Protection 2 (12), 1020-1029. doi:10.4236/jwarp.2010.212122.

[9] Fathy, S. A. H., Abdel Hamid, F. F., Shreadah, M. A., Mohamed, L. A., El-Gazar M. G., 2012. Application of principal component analysis for developing water quality index for selected coastal areas of Alexandria Egypt. Resources and Environment 2 (6), 297-305. doi: 10.5923/j.re.20120206.08.

[10] Huang, L.-M., Deng, C.-B., Huang, N., Huang, X.-J., 2013. Multivariate statistical approach to identify heavy metal sources in agricultural soil around an abandoned Pb–Zn mine in Guangxi Zhuang Autonomous Region. China. Environmental Earth Sciences 68 (5), 1331-1348.

[11] Vasanthavigar, M., Srinivasamoorthy, K., Prasanna, M. V., 2013. Identification of groundwater contamination zones and its sources by using multivariate statistical approach in Thirumanimuthar sub-basin, Tamil Nadu, India. Environmental Earth Sciences 66 (6), 1-13.

[12] Xanthopoulos, P., Pardalos, P. M., Trafalis, T. B., 2013. Principal component analysis. In: Robust data mining. pp. 21-26. Springer New York. doi: 10.1007/978-1-4419-9878-1.

[13] Oikonomakou, N., Vazirgiannis, M., 2010. A review of web document clustering approaches. In: Data mining and knowledge discovery handbook, 2nd ed. Oded, M., Lior, R. (Eds.), Springer US., pp. 931-948.

[14] Embrechts, M. J., Gatti, C. J., Linton, J., Roysam, B, 2013. Hierarchical clustering for large data sets. In: Advances in intelligent signal processing and data mining. Georgieva, P., Mihaylova, L., Lakhmi C., Jain, L. C. (Eds.). Springer Berlin Heidelberg, pp. 197-233.

[15] 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.

[16] 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.

[17] 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.

[18] American Public Health Authority (APHA), 1995. Standard methods for the examination of water and waste water, 19th edition. Published by: American Public Health Association, Washington, D.C.