International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

Interpretation of Water Quality parameters for Karak Springs in South Area of Jordan Using Principal Components Analysis

Suleiman M. Tashtoush*

Mutah University, Department of Mathematics & Statistics, Jordan
Email: suleiman25@yahoo.com

Abstract

Groundwater resources are considered to be significant and economical water resources. The groundwater is the main source of water in the Karak area. The importance of springs in the area is that it is the main source of drinking water and agriculture activities. The chemical composition of groundwater is controlled by many factors that include the composition of precipitation, mineralogy of the watershed and aquifers, climate and topography. Since the data obtained in this study is multivariate and correlative in nature. Principle components analysis (PCA) method is used for the interpretation of the data. Which applied to 13 physical and chemical water quality parameter from 16 springs in Karak in southern area of Jordan, 192 samples were collected during the study year (2004) to give simpler and more easily interpretation results for the evaluation of these parameters. This method allows also finding interdependences between variables. The PCA produced four significant main components explain more than 87% of the variance, Namely, seasonal effects, aquifer effects, agricultural activates, rainfall effects, and geological effect; that represent 55.828%, 14.043%, 10.005%, and 7.86% respectively of the total variance of water quality in the study area. Finally the results of PCA reflected a good look on the water quality monitoring and interpretation of the ground water in the specified area.

***Keywords**:* Principle components analysis; water quality; groundwater; Karak; Jordan.

----

* Corresponding author. Tel: 00962795481209
E-mail address: suleiman25@yahoo.com

1. Introduction

Surface and groundwater are the primary sources of water for human consumption. Izadi [1] mentioned that "Groundwater resources are considered to be significant and economical water resources". Groundwater is an important source of drinking water from wells or springs in the study area. According to [2], the quality of water is identify by its physical and chemical properties, the particulate problem in case of water quality monitoring is the complexity associated with analysis the large number of measured variables, the data sets contain reach information about the behavior of the water body. The classification modeling and interpretations of monitoring data are the most important step in the assessment of water quality. Water quality parameters interact with each other to define the resource water quality parameters, many researchers treated water quality parameters individually by describing the seasonal variability and their causes. The author in [3] used factor analysis technique is very useful in the analysis of data corresponding to large number of variables, analysis via this technique produce easily interpretable results, and this method have been used successfully in hydrochemistry for many years, water quality assessment and environmental research employing multi component techniques are will described in the literature. Multivariate statistical approaches allow driving hidden information from the data set about their possible influences of the environment on water quality. In recent years many studies have been done using Principal Components Analysis (PCA) in the interpretation of water quality parameters, Lohoni and Todino in [4] utilized PCA to provide a quick analytical method for the water quality of Chao Phraya River in Thailand. Mazlum in [5] studied the factors that caused variations in water quality at the monitoring station on the Porsuk Tributary in the Sakarya river basin by using PCA; researchers referred that PCA is more reliable than factor analysis and it a pure mathematical technique without any assumption. Also PCA has been successfully applied to sort out hydrogeological and hydrogeochemical processes from commonly collected groundwater quality data as demonstrated in [6, 7, 3, 8, 9]. The author in [10] constructed a statistical model based on the PCA for coastal water quality data from the Cochin coast in south west India, which explain the relationships between the various physicochemical variables that have been monitored and environmental conditions effect on the coastal water quality, authors in [11-14] used a multivariate statistical techniques in the assessment of groundwater. Authors in [15-16] applied the multivariate statistics (cluster analysis and principal components to reveal information about the natural and anthropogenic origin of contaminants in ground water). Batayneh in [17] used PCA to study the groundwater hydrogeochemistry of the sedimentary rock shallow aquifer system in the Yarmouk Basin of north Jordan. Reference [18] used PCA to detect the temporal and spatial variations of groundwater composition in the Nile Delta aquifer. Amadi in [19] used PCA to identify factors contributed to the changes in the groundwater chemistry in Benin Formation of Southern Nigeria.

PCA is used in this study in order to interpret the variation and relationship between parameters, to identify the factors and sources influencing groundwater quality and to suggest useful tools for both management of water resources and monitoring of groundwater quality of springs in Karak area.

2. Methodology and methods

2.1 Study area

According to [20], the investigated area in general was considered as a very arid to semiarid area, while it is marked by sharp seasonal variation in both temperature and precipitation. It is about 80km far from the sea and the annual rainfall is about 321 mm/year. The winter season in Jordan is the main season of rainfall; the season starts in late October to April. The water resources in Jordan mainly depend on rainfall, which is subject to great variability. Meanwhile, Salameh in [21] noted that "There is an observed population growth, so there is high demand for water".

The pollution and scarcity of water resources are the most important challenges facing the country, while the pollution of water is caused by the high population growth (the average population growth is around 3.5% per year), arid to semiarid climate and the lack of sewer systems, which results to infiltrate of wastewater into spring and groundwater resources (see authors in [21-22]). In the towns and villages in Jordan wastewater from septic tanks and cesspools infiltrate into the ground water accelerates the pollution problems. It increases the pressure on the water resources by increasing the metals, salinity, and ionic composition. Springs water in the investigated area is the major source of drinking water and irrigational activities. The average discharge in the springs in Karak area ranges from 1.79m3/h to 553.54 m^3/h. The surface water in the study area is feeding the spring during winter. Approximately 89% of the total amount of the rainfall water precipitated over area evaporates into the atmosphere and the rest feeds the groundwater. The ground water resources in the investigated area consist of 16 springs emerging from Upper Cretaceous. The purpose of the present study is to investigate the chemical and physical characteristics of spring water, and monitoring the changes in water quality during the studying period from January 2004 to January 2005 and studying the influence the chemical parameters and characterize the suitability of springs for drinking and agricultural purposes.

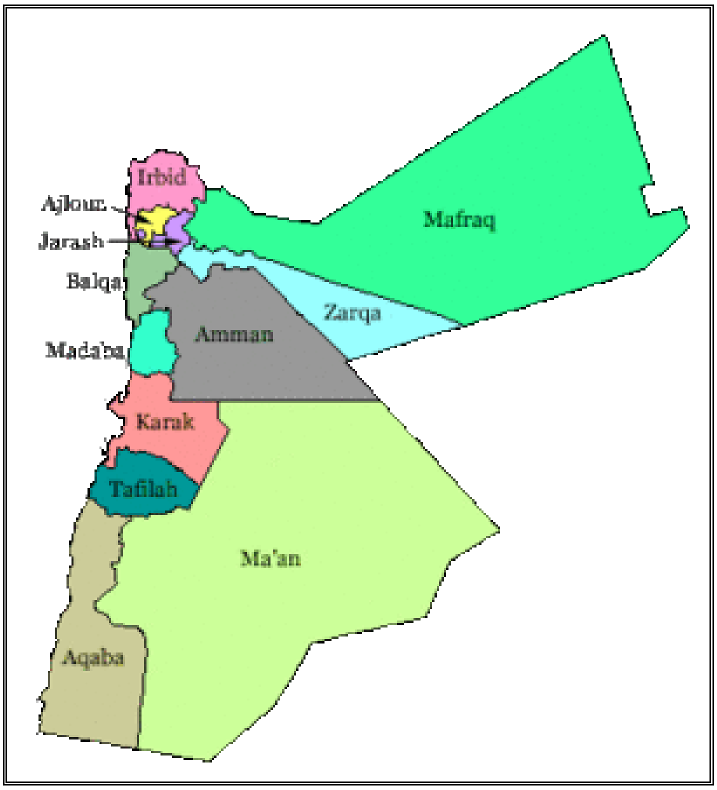

Figure 1: Jordan Map

2.2 Data collection and processing

Samples were collected from 16 springs and wells in Karak area; monthly samples were collected from these positions during the studying period from January 2004 to January 2005; So 192 samples were taken through this period, for each sample 13 parameters (variable) had been measured including: Ph, EC, Na^{+1}, Ca^{+2}, Mg^{+2}, K^{+1}, CL^{-1},Temperature (Temp), Total Hardness (TH), Total Dissolved Soled (TDS), $HCO3^{-1}$, $NO3^{-1}$, and $SO4^{-2}$.

Table 1: Descriptive statistics of the water parameters

Variable	N	Mean	Median	Min	Max	Variance	Standard Deviation	WHO standards
Ph	192	7.6557	7.656	6.34	8.32	.137	0.3705	6.5-8.5
EC	192	1155.938	1155.938	620.00	3675	512369.242	715.7998	400
Na+1	192	59.422	59.422	4.14	365.93	6321.647	79.5088	200
Ca+2	192	96.273	96.273	62.12	158.92	619.904	24.8979	100
Mg+2	192	32.591	59.422	21.03	64.55	165.826	12.8774	50
K+1	192	15.039	15.039	1.96	50.05	133.06	11.5352	20
CL-1	192	111.83	111.830	2.84	626.4	18057.761	134.3792	250
HCO3-1	192	290.241	290.241	37.82	740.92	5183.721	71.9981	125-350
NO3-1	192	64.215	64.215	9.3	115.94	387.091	19.6746	50
SO4-2	192	132.04	132.040	36.48	474.72	14133.786	118.8856	250
Temp(C^0)	192	17.211	17.211	13.11	20.55	3.426	1.8509	12-25
TH	192	374.308	374.3089	252.03	651.47	12247.553	110.6687	500
TDS	192	656.531	656.531	389.01	1886.72	134612.524	366.8958	500-1000

Descriptive statistics of the data set are presented in table (1). Which show that the concentrations of the parameters in groundwater in the study area are all within the WHO [23] allowable limit for drinking water. EC, Mg, and NO_3 higher than the limits in some springs. In total hardness (TH) classification; the water in the area is very hard.

2.3 Principal Components Analysis

The parameters obtained from the laboratory analysis were used as variables inputs for Principal Components Analysis (PCA). This type of ordination reduces the dimensionality of the data set and minimizes the loss of information caused by reduction. From the standardized covariance or correlation matrix of the data the initial factor solution were extracted by the multivariate principal components extraction, then a number of PC were selected from the initial according to their eigenvalues and scree diagram. Authors in [24] noted that " Orthogonal rotation of the selected initial components to terminal factor solutions was done by Kaiser's Varimax schemes that attempt to achieve simple structure with respect to both the rows and columns, this method maximizes the variance of the loadings on the factors and hence adjusts them to be either (±1) or near zero".

Principal Components Analysis was performed on correlation matrix of the raw data in which a water samples is described by 13 physical, chemical and biological parameters. This technique aims to transform the observed variables to a new set of variables (PC) which are uncorrelated and arranged in decreasing order of importance so that to simplify the problem, (tables 2a) represented the determined initial PC and its eigenvalues and percent of variance contributed in each PC, Fig. 2 Show the scree plot of the eigenvalue for each component. eigenvalues accounts and scree plot showed that the first four PC is the most significant components which represent more than 87% of the variance in water quality of ground water in Karak area 55.828% by PC1, 14.043% by PC2, 10.005% by PC3, and 7.860% by PC4; in addition it have eigenvalues >1. Components loading and communalities for each variable in four selected components before varimax rotation were explained in table 2; and after varimax rotation in table 2b, communalities provide an index to the efficiency of the reduced set of components and degree of contribution of each variable in the selected four components; in this case communalities showed that all the variables have been described to an acceptable levels in the selected components. Consequently, the conclusive result was the selected components can be considered significant in the analysis. Table 4 shows the correlation components matrix (components Score covariance matrix) of varimax rotated four PC's; it note that there are no correlation between components, each components represent a discrete unit from others.

Authors in [25] stated that "Components loading (correlation coefficients), which measure the degree of closeness between the variables and the PC, the largest loading either positive or negative, suggests the meaning of the dimensions; positive loading indicates that the contribution of the variables increases with the increasing loading in dimension; and negative loading indicates a decrease". In general, component's loadings larger than 0.6 may be taken into consideration in the interpretation, in other words, the most significant variables in the components represented by high loadings have been taken into consideration in evaluation the components , as authors in [5] mentioned.

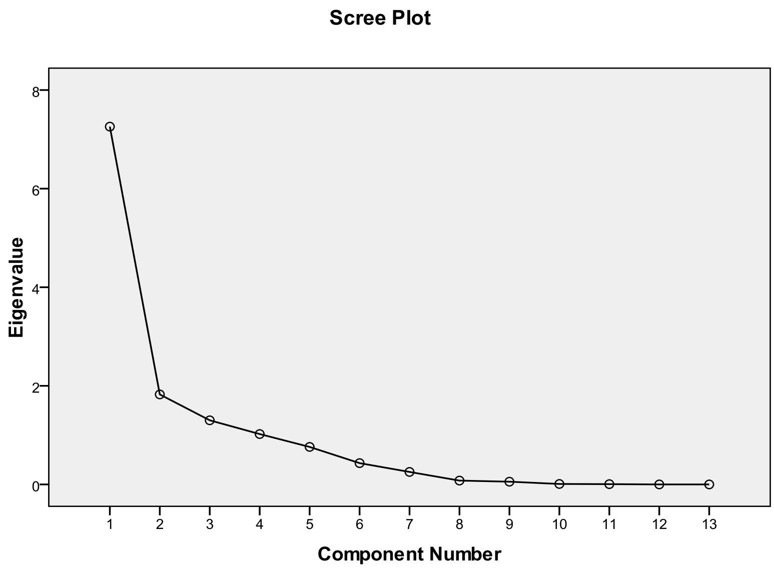

Figure 2: Scree plot

Table 2: Explains initial components

Components	Initial Eigenvalues		
	Total	% of Variance	Cumulative%
1	7.258	55.828	55.828
2	1.826	14.043	69.871
3	1.301	10.005	79.876
4	1.022	7.860	87.736
5	0.761	5.852	93.588
6	0.432	3.324	96.912
7	0.253	1.949	98.861
8	0.078	0.598	99.459
9	0.055	0.426	99.885
10	0.009	0.071	99.957
11	0.006	0.043	100.00
12	9.187E-16	7.067E-15	100.00
13	-1.669E-17	-1.284E-16	100.00

Table 2a: Unrotated component matrix.

	Component Number			
	1	2	3	4
Ph	.337	-.236	.764	.070
EC	.940	-.275	.029	.089
Na+1	.882	-.394	.062	.029
K+1	-.019	-.021	-.391	.903
Ca+2	.784	.457	-.128	-.145
Mg+2	.964	.162	-.012	-.017
CL-1	.853	-.440	.038	.073
NO3-1	.224	.696	.114	.009
SO4-2	.955	-.101	-.048	-.020
HCO3-1	.690	.496	-.223	.018
Temp(C^0)	-.004	.483	.685	.395
TH	.901	.334	.035	.081
TDS	.979	-.157	-.026	.047

Table 2b: Rotated component matrix with its communalities.

	Component Number			
	1	2	3	4
Ph	.456	-.183	.673	-.251
Ec	.957	.221	.038	.029
Na^{+1}	.964	.092	.013	-.031
Ca^{+2}	.439	.811	-.039	-.090
Mg^{+2}	.755	.619	.056	-.031
K^{+1}	.028	-.037	-.016	.984
CL^{-1}	.963	.038	-.003	.019
HCO_3^{-1}	.343	.800	-.057	.101
NO_3^{-1}	-.140	.675	.268	-.024
SO_4^{-2}	.874	.399	-.036	-.026
Temp(C^0)	-.177	.239	.871	.108
TH	.607	.751	.005	-.065
TDS	.928	.354	-.002	.024

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.
Rotation converged in 8 iterations

Table 2c: Component Transformation Matrix

Component	1	2	3	4
1	.867	.496	.039	-.025
2	-.491	.842	.221	.028
3	.057	-.189	.898	-.392
4	.063	-.092	.378	.919

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

2.4 Correlation

Pearson correlation matrix was used to determine the relationship between variables.

Table 3: Karl Pearson Correlation Matrix of the Chemical Variables Karak Area

	Ph	EC	Na	Ca	Mg	K	Cl	HCO_3	NO_3	SO_4	Temp	TH	TDS
Ph	1	.366	.361	.109	.295	-.162	.336	.054	-.085	.295	.286	.202	.310
EC		1	**.969**	.546	**.857**	.030	**.961**	.492	.107	**.905**	-.074	**.716**	**.977**
Na+1			1	.445	**.753**	-.034	**.989**	.385	.052	**.850**	-.122	.610	**.951**
Ca+2				1	**.850**	-.068	.401	**.737**	.310	**.733**	.082	**.968**	.665
Mg+2					1	-.023	**.724**	**.732**	.272	**.912**	.060	**.955**	**.902**
K+1						1	.016	.079	-.079	-.017	.029	-.049	.022
CL-1							1	.326	.019	**.818**	-.139	.571	**.932**
HCO3-1								1	.398	.603	.045	**.764**	.596
NO3-1									1	.060	.326	.304	.159
SO4-2										1	-.057	**.847**	**.951**
Temp(C^0)											1	.075	-.065
TH												1	**.805**
TDS													1

According to [26], Samples showing correlation coefficients of r>0.7 are considered to be strongly correlated, where as r>0.5-0.7 shows moderate correlation. The strong to perfect correlation between the chemical parameters is an indication of common source. The correlation matrix (Table 3) describes the interrelationship between variables and the results for 13 hydrochemical parameters show that a strong correlation between; (EC and Na, Mg, Cl, SO_4, TH, TDS), (Na and Mg, Cl, SO_4, TDS), (Ca and Mg, HCO_3, SO_4,TH), (Mg and Cl, HCO_3, SO_4,TH, TDS), (Cl and SO_4, TDS), (HCO_3 and TH), (SO_4 and TH) and (TH and TDS). Moderate correlation between; (EC and Ca), (Na and TH), (Ca and TDS), (Cl and TH), and (HCO_3 and SO_4, TDS).

Isa et al in [27] stated that “It can also be an indication of weathering of calcite mineral, as illustrated in Equations (1) (carbonate acid formation) and (2) (calcite weathering equation):

$$H_2O + CO_2 \rightarrow H_2CO_3 \quad (1)$$

$$CaCO_3 + CO_2 + H_2O \rightarrow Ca^{2+} + 2HCO_3^- \quad (2)$$

First, H_2O in the atmosphere reacts with CO_2 to form carbonic acid (Equation 1), then the Rain water falls on the land surface and dissolves part of the aquifer’s parent material, CaCO3 (calcite), as shown in (Equation 2). These processes give rise to surplus Ca and HCO_3 ions”, thus during the recharge process a lot of Ca and HCO_3 ions are released into the groundwater. Which is composed to mainly of limestone and dolomite limestone’s that are rich in carbonate, calcium, and magnesium. The presence of the NO3 with Na suggests strongly that most of the excess amount of Na is come mainly from the anthropogenic pollution.

3. Principal Components interpretations

Principal Component one; has a high loading of TH, TDS, EC, Mg, HCO3, SO4, Ca, Na, and Cl and explain 55.828 % of the total variance (Table 2a). This component can be described to the intrusion of ground water system which increases the concentrations of these ions in the springs in other words the effect of geological information of the springs. Since this component reflect the wastewater from the domestic and industrial and its organic load disposed to the springs. Principal Components two explain 14.043 % of the total variance has high loading of NO3, this may indicate to the effect of drainage water from the agriculture and storm water. Principal Component three have 10.005 % of the total variance has high loading of Ph and Temp, This factor can be ascribed to the variation of natural atmospheric conditions, from temperature and rainfall that effect, so this component represent seasonal effects upon water quality. In summary the four extracted Principal Components representing four different processing.

4. Result and conclusion

From the 13 components in table 2 the first four components are sufficient to explain the monitoring area. These components explain more than 87% of the total variance of the original data set in Karak area. Moreover, the first four selected principle components explained more than 50% of the variance of each quality variable (see communalities in table 2a and 2b). Principal components analysis of water quality data for Karak area showed that seasonal effects, aquifer effects, agricultural wastes and storm water, geological effects, and domestic are caused the main variation in water quality of the Karak area.

Acknowledgment

The author would like to thank Prof. Omar Al-Khashman for making this data available for analysis.

References

[1] A. Izadi, K. Davary, A. Alizadeh, B. Ghahraman, A. Sadeghi, and A. Moghaddamnia, "Application of 'panel data' modeling to predict groundwater levels in the Neishaboor Plain, Iran", Hydrogeology Journal,Vol. 20, 435 – 447, 2012.

[2] H. Bayacioglu, "Surface water quality assessment using factor analysis", Water SA, Vol.32, No. 3, 389-394,2006.

[3] P. Praus, "Water Quality Assessment Using SVD-Based Principal Component Analysis of HydrologicalData", Water SA., Vol. 31, No. 4, 417-422, October 2005.

[4] N. Lohani and G. Todino , "Water Quality Index for Chao Pharya River", Journal of ASCE, EnvironmentalDivision, 110 (6), 1163-1176, 1984.

[5] N. Mazlum, A. Ozer, and S. Mazlum, "Interpretation of Water Quality Data by Principal Components

Analysis", J. of Engineering and Environmental Science, 23, 19-26, 1999.

[6] R. Jayakumar and L. Siraz, "Factor analysis in hydrogeochemistry of Coastal aquifers- preliminary study", Environmental Geology, 31(3/4), 174-177, 1997.

[7] S.R. Salman and Y.H. AbuRuka'h, "Multivariate and principle component Statistical Analysis of contamination in urban and agricultural Soils from north Jordan, Environmental Soils from north Jordan", Environmental Geology, 38(3) September, 265-270, 1998.

[8] S.B. Olobaniyi and F.B. Owoyemi, "Chemical Facies of groundwater in the Deltaic Plain Sands Aquifer of Warri, Western Niger Delta, Nigeria", AJST, Vol. 7, No. 1, 73-81, June 2006.

[9] A. N. Amadi , P. I. Olasehinde, J. Yisa , E. A. Okosun, H. O. Nwankwoala, & Y. B. Alkali, "Geostatistical assessment of groundwater quality from coastal aquifers of Eastern Niger Delta, Nigeria", Journal of Geosciences, 2(3), 51–59, 2012.

[10] C.S. Iyer, M. Sindhu, S.G. Kulkarni, S.S. Tambe, and B.D. Kulkarni, "Statistical analysis of the physic-chemical data on the coastal water of Cochin", J. Environ. Monit., 5, 324-327, 2003.

[11] I. Gupta, S. Dhoge, and R. Kumar, " Study of variations in water quality of Mumbai coast through multivariate analysis techniques", Indian Journal of Morine Sciences, Vol. 38(2), 170-177, June 2009.

[12] M. Kumar and Y. Singh, "Interpretation of Water Quality Parameters for Villages of Sanganer Tehsil, by Using Multivariate Statistical Analysis", J. Water Resource and Protection, 2, 860-863, 2010.

[13] K. Nosrati and M. Van Den Eeckhaut, "Assessment of Groundwater Quality using Multivariate Statistical Techniques in Hashtgerd Plain, Iran", J. Environmental Earth Science 65 (1), 331-344, 2012.

[14] N. U. Usman, M. E. Toriman, H. Juahir, Abdullahi,. A. Rabiu, and H. Isiyaka, "Assessment of Groundwater Quality Using Multivariate Statistical Techniques in Terengganu", Science and Technology, 4(3), 42-49, 2014.

[15] N. HO, N. SA, U. FA, and A. AN, "Research Paper Statistical Characterization of Groundwater Qualityin Port Harcourt, Southern Nigeria", Sci-Afric Journal of Scientific Issues, Research and Essays Vol. 2 (8), 338-34, August 2014.

[16] S. J. Cobbina, F. A. Armah and S. Obiri, " Multivariate Statistical and Spatial Assessment of Groundwater Quality in the Tolon-Kumbungu District, Ghana", Research Journal of Environmental and Earth Sciences 4(1), 88-98, 2012.

[17] A. Batayneh and T. Zumlot, "Multivariate Statistical Approach to Geochemical Methods in Water Quality Factor Identification; Application to the Shallow Aquifer System of the Yarmouk Basin of North Jordan", Research Journal of Environmental and Earth Sciences 4(7), 756-768, 2012.

[18] N. El Arabi, Y. Idris, and A. Fekry, " Temporal and Spatial Change Detection of Variations in the Groundwater Composition by Multivariate Statistical Techniques", New York Science Journal, 6(11), 38-48, 2013.

[19] A. N. Amadi, P. I. Olasehinde, H. O. Nwankwoala, M. A. Dan-Hassan, and Mamodu Adegbe," Controlling Factors of Groundwater Chemistry in the Benin Formation of Southern Nigeria", International Journal of Engineering Science Invention, Vol. 3 (31), 11-16, 2014.

[20] Department of Meteorology, Internal report, 2004, Amman-Jordan.

[21] E. Salameh, Water quality degradation in Jordan. Royal Society for the Conservation of Nature and Freidrich Ebert Stiftung, Amman, Jordan, 1996, 179.

[22] A. Al-Kharabsheh and R. Ta'any, "Influence of urbanization on water quality deterioration during drought periods at south Jordan". Journal of Arid Environments, 53, 619-630, 2003.

[23] World Health Organisation . Guidelines for drinking water quality, 3rd edition, 2004, Geneva, Switzerland, WHO.

[24] D.G. Kleinbaum, L.L. Kupper, and K.E. Muller, Applied Regression Analysis and Other Multivariable Methods, PWS-KENT Publishing Comp., 2nd ed. 1988, USA, 718.

[25] F. W. Lawrence and S. B. Upchurch, "Identification of Recharge Area Using Geochemical Factor Analysis. Ground Water 20(6), 680-687, 1983.

[26] S. Adams, R. Titus, K. Pietersen and G. Harris, "Hydrochemical characteristics of aquifers near Sutherland in the Western Karoo, South Africa", J. Hydrol, 241(1-2), 91-103, 2001.

[27] N.M. Isa, A.Z. Aris, and W.N.A. Sulaiman, "Extent and severity of groundwater contamination based on hydrochemistry mechanism of sandy tropical coastal aquifer", Sci. Total Environ., 438, 414–425, 2012.