

# Assessment of Groundwater Quality Using Multivariate Statistical Techniques in Terengganu

Usman Nasiru Usman\*, Mohd Ekhwan Toriman, Hafizan Juahir,  
Musa Garba Abdullahi, Ali Auwalu Rabi, Hamza Isiyaka

East Coast Environmental Research Institute, University Sultan Zainal Abidin, Malaysia

**Abstract** Despite the importance of groundwater in Terengganu, Malaysia, quality assessment has received little attention, and effort to use hydrochemistry data to solve particular problems are even fewer or non-existent. This paper, reports results from large hydrochemistry data analysed using multivariate statistical techniques such as Cluster Analysis (CA), Discriminant Analysis (DA) and Principal Component Analysis (PCA) with the objectives of determining the spatial variability of groundwater and to identify the sources of pollution that presently affects the groundwater. The water quality data was monitored at ten different wells, over the period of six years (2006-2011) using 24 water quality parameters. The CA allowed the formation of three clusters between the sampling wells reflecting differences on water quality at different locations. DA as a data reduction techniques was used to evaluate spatial variability in water quality, as it uses only 3 parameters ( $\text{Ca}^+$ ,  $\text{NO}_2$ , and PH) affording 73.33% correct assignment to discriminate between the clusters using forward stepwise mode from the original 24 parameters, while backward stepwise mode yielded 83.33% correct assignment to discriminate nine parameters ( $\text{Ca}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{Fe}^{2+}$ ,  $\text{SO}_4^-$ ,  $\text{Cl}^-$ , AS, Mn,  $\text{NO}_2$ , and conductivity). PCA was used to examine the root of each water quality parameter due to nature and anthropogenic activities based on the three cluster regions. It identified eight PC's, responsible for 76.45% of the total variance in the data set. The main factors obtained indicate that parameters influencing groundwater quality of the clusters are mainly related to natural (dissolution of soil and rocks), pointsource (municipal wastewater and industries) and non-point source pollution (agriculture) in the region. The results of this study clearly demonstrate the usefulness of multivariate statistical techniques in Geochemistry.

**Keywords** Multivariate statistical techniques, Cluster analysis, Discriminate analysis, Principal component analysis

## 1. Introduction

Without water life could not be possible on the earth for all living organisms, plants, lower animals and human being. It is next to the air in importance for human existence. Groundwater is a vital renewable resource that has been taken for granted and giving little protection. Despite the enormous important of these resources, the importance of water to life on earth, particularly human being, becomes clearer when one consider its role in various aspect of human endeavour; they include domestic, industrial, agricultural, laboratory and recreational uses.

Groundwater is considered a limited resources in most locations: increasing evidence on groundwater contamination in recent years, coupled with concern about human health and ecological effects of contaminants such as nitrate, pesticides has heightened pressure on public agencies to better groundwater (National Research Council, 2000).

However, contaminated water can caused diseases, especially in areas where people throng. In 1981, WHO estimated that 80% of all sickness and disease is caused by polluted water and estimated that in developing countries three out of five people have no access to safe drinking water and only one out of four has sanitary facilities.

Global efforts to improve access and quality of drinking water have been hindered by rapid urbanisation. According to the data presented at world water in Stockholm, the percentage of people in urban areas with access is declining. According to Anders Berntell, Executive Director of Stockholm International Water Institutes (SIWI) "bad water kills more people than HIV, Malaria and War together, affecting the lives of families and economic development of many countries of the world " He said at the world water week conference being held in Swedish capital September, 2010.

Water quality is mainly influence by anthropogenic effects; unplanned disposal of waste water generated from municipal, industrial and agricultural sources with little or no treatment prior to discharge is a common practice in many developing countries including Malaysia (Juahir, 2009) and this lead to the subsequent leaching of the pollutant into the ground and

\* Corresponding author:  
crcproduction@gmail.com (Usman Nasiru Usman)  
Published online at <http://journal.sapub.org/scit>  
Copyright © 2014 Scientific & Academic Publishing. All Rights Reserved

cause significant degradation of the groundwater. However, due to low hydraulic gradient of the area, as the area is flat, the potential of sea water intrusion threat is likely to be higher particularly during low recharge (drier month), (Shamsuddeen et al, 2014).

Different multivariate statistical techniques such as Cluster Analysis (CA), Discriminant Analysis (DA), and Principal Component Analysis (PCA) were employed in this research work. These techniques can be used to obtain relationship between parameters and sampling site, 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 (Nosrati, 2011). CA was employed to examine the spatial groupings of the sampling wells. It is a common method to classify variables into cluster (Massart and Kaufmann, 1983). CA and PCA are commonly supported by DA as a confirmation for CA and PCA and are usually referred to as pattern recognition techniques (Adams, 1998). The application of different pattern recognition techniques to reduce the complexity of large data set has proven to give a better interpretation and understanding of water quality data (Brown et al, 1980).

Therefore the main objectives of this research work are to determine the spatial variability of groundwater and to identify the root of the pollution that presently affects the groundwater.

## 2. Materials and Methods

### 2.1. Study Area

Terengganu is situated in the North-Eastern Peninsular of Malaysia and it is bordered to the North-West by Kelantan and to the South-West by Pahang and to the East by South China Sea; with a total area of land of 13035km<sup>2</sup> and the maximum elevation of the state is 1507m.

Terengganu has a population of 1,015,776 people as of 2006, Malay make up 94.7% of the population and Chinese 2.6% while Indians 0.2%. Other ethnic group raises the remainder 2.4%. The state population was only 48.7% urban; the majority lived in the rural areas of the state.

The study area has a strong tropical monsoon climate, relatively uniform temperature within 21°C and 32°C range, January till April; the weather is dry and warm with humidity in the lowland consistently high between 82%-86% annually. The annual average rainfall is 2,032mm-2540mm with the most it, falling between Novembers till January.

The geological system of the area ranges from the Cambrian to the quaternary (570 million years to 10,000 years ago). Almost half of the peninsular Malaysia notably in the main range is occupied by granites. These granitic emplacements coincide with the culmination of the late Triassic orogenic event during which older strata were folded and deformed.

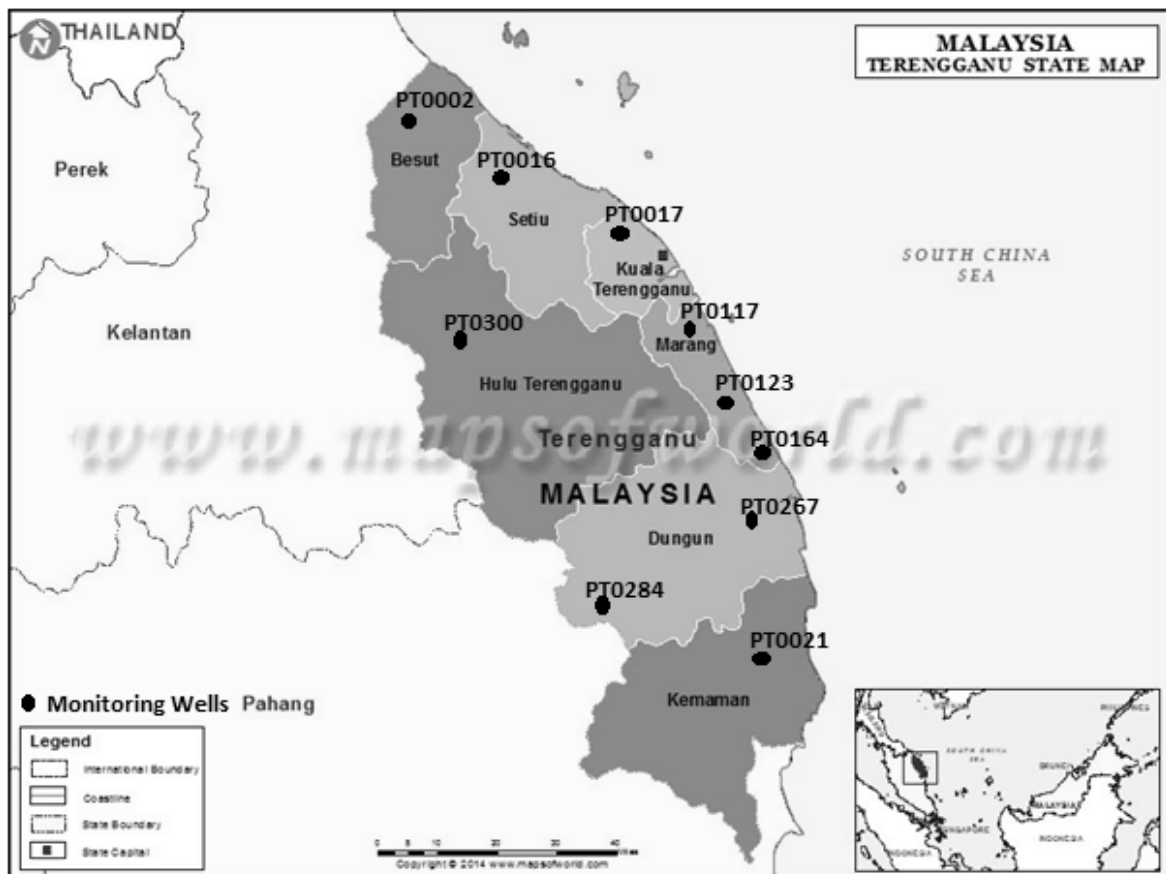


Figure 1. Map of the study area showing the monitoring wells

## 2.2. Data Collection and Treatment

Secondary data were used in this research work. The water quality data in this study were obtained from ten monitoring wells by the department of mineral and geosciences, Terengganu. Each of the ten monitoring wells were observed and identified based on the availability of recorded data from the period of 2006-2011. The ten wells are: PT002, PT017, PT021, PT116, PT117, PT123, PT164, PT267, PT284, and PT300. Even though there are 50 water quality parameters but only 24 consistently sampled parameters were selected and a total of 60 samples and 1440 observation were used for the analysis.

The water quality data obtained from the department of mineral and geosciences is in note pad format, it was then later converted into Microsoft Excel 2007 for all groundwater quality parameters. The monitoring wells were also sorted out (A-Z) in the normalised data set, while non-numerical variables were also transformed into numerical variables for convenient analysis. All the statistical analyses were performed using Microsoft excel 2007 and XLSTAT 2014 versions.

## 2.3. Analytical Methods

Environmetric method is deemed to be the best approach to avoid misinterpretation of large complex environmental monitoring data (Simeonov et al, 2002). The most common environmetric methods used to determine the spatial variability and to identify the pollution sources are Cluster Analysis (CA), Discriminant Analysis (DA), and Principal Component Analysis (PCA).

### 2.3.1. Cluster analysis

This is a group of multivariate techniques which primarily classify (Massart and Kaufmann, 1983) variables or cases (observation or samples) into cluster with high homogeneity level within the class and high heterogeneity level between classes. The spatial variability of groundwater was determine by CA. CA was first performed to group all sample site in order to classify them into cluster to minimized their number. We use CA to link sample site in the configuration of a tree with different branches (Dendogram) which provide visual summary of the clustering process, presenting a picture of the group and their proximity. Branches that have linkage closer to each other indicate a stronger relationship between sample/variables or cluster of sampling site/variable.

In this present study, CA was applied for the grouping of ten different wells using ward's linkage method (Ward 1963). A classification scheme using Euclidean distance (straight line distance between two point in C-dimensional space define by C variable) for similarity measurement together with Ward method for linkage produces the most distinctive groups where each member within groups is more similar to

its fellow member than to any member outside the group (Guler et al, 2002).

### 2.3.2. Discriminant Analysis

The main objective of DA is to discriminate between two or more groups in term of the discriminating variables. It was performed on the data set based on three different modes, i.e. Standard mode, forward stepwise and backward stepwise modes to construct the best discriminant functions (DFs) to confirm the three clusters determined by means of CA and to evaluate spatial variation in portable water quality in Terengganu, Malaysia. In forward stepwise mode, variables are included step-by-step beginning with the more significant until no changes are obtained, whereas, in backward stepwise mode, variables are removed step-by-step beginning with less significant until significant changes are obtained. The membership of a well ina cluster1, 2 and 3 was the dependent variables whereas all the measured parameters constituted the independent variables.

### 2.3.3. Principal Component Analysis

PCA analysis was used as a method of factor extraction, for this study it requires a preceding estimate of the amount of variation in each groundwater quality parameter explained by the factors. Eigenvalues are the amount of variance explained by each factor; each parameter had a variance of 1 with a total variance of 24 for the entire data set. Factor with eigenvalue >1 explained more total variation in the data than individual groundwater quality parameters, and factor with eigenvalue <1 explain less total variation than individual variable, Therefore only factor with eigenvalue >1 were retained for the interpretation, retained factors were subjected to varimax rotation (Kaiser; 1960 and Vega et al; 1998).

Varimax rotation is an orthogonal rotation method that minimized the number of variables that have high loading on each factor. The VF coefficient having correlation greater than 0.75 are considered as strong and indicate high proportion of its variance explained by the factor, between 0.50 and 0.75 is considered as moderate loading while 0.30-0.50 as weak significant factor loading, indicating much of that attribute's variance remains unexplained and it is less important (Reghunath, et al 2002).

## 3. Results and Discussion

### 3.1. Descriptive Statistics

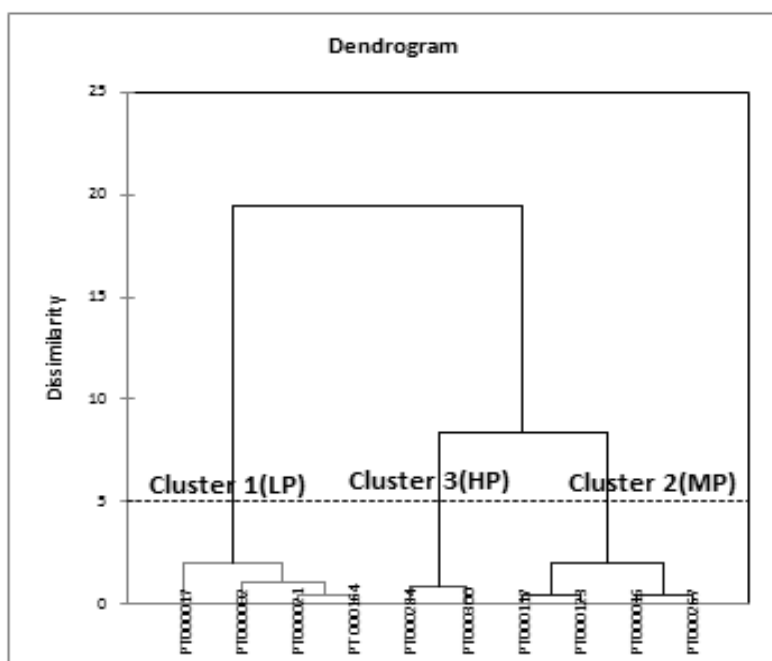
Basic statistics were carried out in order to give initial information about the water quality data. The table below shows the details of descriptive statistics on the water quality variables measured in six years.

**Table 1.** Mean and Standard Deviation of Groundwater Quality Parameters

Variable	Minimum	Maximum	Mean	SD
Turbidity (NTU)	0.050	834.000	36.143	109.574
Colour (HU)	5.000	15.000	5.250	1.434
Na <sup>+</sup> (ppm)	1.200	53.000	11.707	13.595
K <sup>+</sup> (ppm)	0.500	5.900	1.700	1.514
Ca <sup>2+</sup> (ppm)	0.500	101.000	19.423	25.468
Mg <sup>2+</sup> (ppm)	0.250	26.000	3.952	4.440
Fe <sup>2+</sup> (ppm)	0.050	24.000	5.668	5.779
SO <sub>4</sub> (ppm)	2.500	90.000	7.383	11.956
CO <sub>3</sub> (ppm)	0.500	18.000	1.175	3.036
F (ppm)	0.250	1.000	0.323	0.165
P (ppm)	0.010	0.570	0.061	0.105
HCO <sub>3</sub> (ppm)	0.500	291.000	77.308	77.860
Cl <sup>-</sup> (ppm)	0.500	131.000	12.575	19.165
NO <sub>3</sub> (ppm)	0.025	18.000	1.191	2.872
As (ppm)	0.003	0.140	0.007	0.019
NH <sub>4</sub> (ppm)	0.250	1.100	0.293	0.155
Mn0 (ppm)	0.050	0.400	0.116	0.092
Zn (ppm)	0.005	1.200	0.112	0.190
SiO <sub>2</sub> (ppm)	0.050	56.000	13.964	14.437
total solid (ppm)	28.000	6032.000	236.767	767.471
dissolve solid (ppm)	16.000	504.000	106.867	95.392
NO <sub>2</sub> (ppm)	0.003	0.028	0.009	0.008
PH	5.100	8.500	7.075	0.991
Conductivity (μs/cm)	2.900	973.000	163.465	172.208

### 3.2. Cluster Analysis

CA was carried out on the water quality data set to evaluate the spatial variability among the monitoring wells. This analysis resulted in the grouping of monitoring wells into three groups as shown in figure 2.

**Figure 2.** A Dendrogram showing the three regions of the cluster

Cluster1 includes four wells (PT002, PT017, and PT021 and PT164) classified as less polluted (LP) wells.

Cluster 2 includes four wells (PT116, PT117, PT123 and PT267) as moderately polluted (MP) and cluster 3 contained 2 wells (PT284 and PT300) as highly polluted (HP). The clustering of wells indicates that water quality of groundwater is varied smoothly and such variation is likely due to the natural hydrogeological environment and the multipurpose nature of the study area. Omo-Irabor *et al* (2008) also suggest that the multipurpose nature of land use and their effects on groundwater quality hamper the precise spatial classification of monitoring sampling wells. The outcome indicates that for rapid evaluation of groundwater quality, only one well in each cluster is needed to represent a logical, accurate spatial assessment of the water quality for the whole network. The CA techniques shorten the need for numerous sampling stations, monitoring from three monitoring wells that represent three different regions is sufficient. Figure 2; shows the three regions given by CA and its possible pollution sources within the study area.

### 3.3. Discriminant Analysis

In order to determine the spatial variation of groundwater quality among different wells, DA was employed and it was performed using original data of 24 parameters after classification into three major clusters obtained from the CA. Cluster groups (LP, MP and HP) were run as dependent variables, while water quality parameters were treated as independent variables. DA was carried out via standard mode, forward stepwise and backward stepwise modes, the accuracy of spatial classification using standard, forward stepwise, backward stepwise modes discriminate functions were 90.00%, 73.33% and 83.33% respectively.

Three parameters were found to be the most significant variable that best discriminate the clusters ( $\text{Ca}^{+}$ ,  $\text{NO}_2^{-}$  and PH), which means that these three parameters account for the most expected spatial variation in the groundwater quality. Backward stepwise mode on the other hand resulted several parameters ( $\text{Ca}^{+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Fe}^{2+}$ ,  $\text{SO}_4^{-}$ ,  $\text{Cl}^{-}$ , As, Mn,  $\text{NO}_2$  and Conductivity) to discriminate the three clusters. Forward stepwise mode DA was proven to be a useful tool in recognising the discriminant parameters in spatial variation of portable water quality; this is because in forward stepwise mode, variables are included step by step beginning with the more significant variables until no changes are obtained. The spatial DA suggest that calcium, nitrite, and PH were the most significant parameters for discriminating among the cluster yielded by CA, accounted for most of the expected spatial variation in portable water quality. Thus, DA is a method that can determine the classification into predetermined group.

### 3.4. Principal Component Analysis

PCA is performed on the normalized data set (24 parameters) to identify the major variables affecting groundwater quality. Factor with eigenvalue of 1.0 or greater are considered significant and factor with highest Eigenvalues are the most significant (Kim and Mueller, 1987) and are retained in order to understand the underlying data structure (Jackson, 1991) which has expressed that the selected PCs are able to carry more information than a single original variables.

Eight major PCs were extracted which accounted 76.45% variance of the original data structure. The result of the PCs is given in the table 3.

**Table 2.** Classification matrix for DA of spatial variation of the groundwater in Terengganu

		Region assigned by DA			
Sampling regions	%Correct	HCL	LCL	MCL	Total
Standard mode					
HCL	91.67%	11	0	1	12
LCL	91.67%	0	22	2	24
MCL	87.50%	0	3	21	24
Total	90.00%	11	25	24	60
Forward stepwise					
HCL	58.33%	7	0	5	12
LCL	83.33%	0	20	4	24
MCL	70.83%	0	7	17	24
Total	73.33%	7	27	26	60
Backward stepwise					
HCL	66.67%	8	0	4	12
LCL	87.50%	0	21	3	24
MCL	87.50%	0	3	21	24
Total	83.33%	8	24	28	60

**Table 3.** Factor loading and eigenvalues of principal components

variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Turbidity (NTU)	-0.108	-0.015	-0.024	-0.012	-0.040	0.058	-0.015	-0.025
Colour (HU)	-0.011	-0.034	-0.005	-0.031	0.103	0.028	<b>0.987</b>	-0.015
Na (ppm)	<b>0.573</b>	0.073	0.000	0.209	<b>0.640</b>	0.018	-0.151	-0.035
K (ppm)	0.213	0.088	-0.109	-0.009	<b>0.902</b>	-0.016	0.198	0.032
Ca (ppm)	<b>0.795</b>	-0.064	0.098	0.018	-0.082	0.063	0.050	0.063
Mg (ppm)	<b>0.876</b>	-0.076	0.096	-0.043	-0.154	0.034	0.017	0.146
Fe (ppm)	-0.142	-0.042	0.201	0.040	0.073	0.133	-0.070	0.056
SO <sub>4</sub> (ppm)	0.202	0.218	0.104	<b>0.894</b>	0.032	-0.039	-0.032	-0.006
CO <sub>3</sub> (ppm)	0.150	-0.016	-0.071	-0.025	0.015	0.027	-0.015	<b>0.971</b>
F (ppm)	<b>0.325</b>	-0.050	-0.043	-0.025	-0.006	-0.048	-0.037	0.072
P (ppm)	-0.098	<b>0.346</b>	-0.048	0.174	-0.012	0.015	0.007	0.083
HCO <sub>3</sub> (ppm)	<b>0.864</b>	0.130	-0.063	-0.024	0.080	0.121	0.037	0.054
Cl (ppm)	<b>0.877</b>	0.013	-0.156	-0.008	0.192	-0.005	-0.080	-0.036
NO <sub>3</sub> (ppm)	-0.147	-0.023	-0.047	0.077	0.008	<b>-0.953</b>	-0.031	-0.028
As (ppm)	-0.061	0.018	0.084	<b>0.969</b>	0.026	-0.049	-0.011	-0.023
NH <sub>4</sub> (ppm)	-0.014	<b>0.928</b>	0.004	0.185	0.100	0.024	-0.040	-0.023
Mn (ppm)	-0.037	0.005	<b>0.883</b>	0.210	-0.109	0.058	-0.007	-0.095
Zn (ppm)	0.033	-0.031	0.190	-0.073	-0.141	0.038	0.013	-0.035
SiO <sub>2</sub> (ppm)	0.110	0.040	0.114	-0.029	0.221	0.119	0.077	-0.067
total solid (ppm)	0.019	-0.034	-0.027	-0.023	-0.019	0.033	-0.014	-0.005
dissolve solid (ppm)	<b>0.918</b>	-0.055	0.042	0.093	0.227	0.047	-0.001	0.081
NO <sub>2</sub> (ppm)	-0.120	0.163	0.005	0.080	0.017	0.094	0.061	-0.068
PH()	<b>0.438</b>	-0.050	0.164	-0.018	0.098	0.217	-0.028	0.160
Conductivity(us/cm)	<b>0.960</b>	0.010	-0.047	0.076	0.168	0.034	0.003	0.025
Eigenvalue	<b>6.310</b>	2.774	2.311	1.868	1.536	1.406	1.102	1.043
Variability (%)	<b>26.293</b>	11.560	9.630	7.784	6.399	5.857	4.591	4.345
Cumulative %	<b>26.293</b>	37.854	47.483	55.267	61.666	67.523	72.114	76.459

PC1 accounts for 26.29% of the total variance, showing strong positive loading on  $\text{Ca}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{HCO}_3^-$ ,  $\text{Cl}^-$ , Dissolve Solid and conductivity, while moderate positive loading on sodium and a weak positive loading on fluorine and PH. The high loading factor of conductivity is due to the active participation of dissolve ions in the groundwater quality. The major variables constituting PC1 ( $\text{Ca}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{HCO}_3^-$ ,  $\text{Na}^+$ ) is related to the hydro chemical variables originating from mineralization of groundwater. The presence of  $\text{Cl}^-$  may also be an indicator of point source pollution by urban waste water discharge, while PH is related to municipal waste.

PC2 accounted 11.56% of the total variance and it is mainly participated by  $\text{NH}_4$  with strong positive loading and weak positive loading on Phosphorous.  $\text{NH}_4$  is closely related to the organic matter contents of the sediment and this high amount of nutrients might also result from the application of manure in agricultural activities (Terceiro et al, 2008).

Out of the total variance, 9.63% is explained by PC3 and is mainly carried by Mn. The dissolution and weathering process of the mineral is mainly responsible for the release of

Mn, however, the activities is also controlled by the redox level of groundwater.

Additionally, 7.78% of the total variance of water quality is exhibited by  $\text{SO}_4$  and As with a strong positive loading under PC4. Dissolution of gypsum and sodium sulphate mineral could increase  $\text{SO}_4$  concentration in groundwater. In general  $\text{SO}_4^{2-}$  content is low in groundwater and reveals the higher level of groundwater reducing condition. Moreover, the released of As is reported from the natural source under the reducing groundwater environment (Chapagain et al, 2009).

PC5 explained 6.39% of the total variance of water quality in groundwater, with a strong positive loading on  $\text{K}^+$  and moderate positive loading on  $\text{Na}^+$ . Association between  $\text{K}^+$  and  $\text{Na}^+$  suggest the dissolution of calcite and dolomite affected by erosion and deposition from upland area.  $\text{K}^+$  can be enriched in natural water due to the weathering of igneous rock and magmatic rocks. In the weathering of igneous rock, potassium feldspars are usually the main source of  $\text{K}^+$  ion.

PC6, explaining 5.85% of the total variance has a strong

negative loading on  $\text{NO}_3^-$  and is difficult to interpret. There are two possible explanations for this negative relation. First, the negative correlation with  $\text{NO}_3^-$  indicates that concentration of  $\text{NO}_3^-$  is the result of different pollution process involving industrial and municipal waste water (Kennel et al, 2008), Fertilizer and the application of agricultural pesticides (Koh et al, 2010, Shrestha and Kazama, 2007). Kaown et al (2009) also showed that mineralisation of organic N fertilizer was dominant source for nitrate in groundwater. Second, this factor can be interpreted as denitrification and nitrate reduction combined with other geochemical process (Levins and Gosk, 2008).

PC7 and PC8 explained 4.59 and 4.34% of the total variance of water quality in groundwater respectively. PC7 is mainly showed strong variation by colour, whereas,  $\text{CO}_3$  has carried major variation of water quality under PC8. Perhaps the most cause of groundwater colour is the presence of minerals and organic matter. Red and brown colour is due to iron; black to manganese or organic matter and yellow to dissolved organic matter such as tannins. Natural processes such as dissolution of carbonate mineral and dissolution of atmospheric, and soil  $\text{CO}_2$  gas could be a mechanism supplying  $\text{CO}_3^{2-}$  to the groundwater. It can also be related to atmospheric pollution from gaseous emanation into the atmospheric from petroleum related industrial and vehicular exhausts (Omo-Irabor et al, 2008).

#### 4. Conclusions

The study has examined water quality of groundwater in Terengganu, Malaysia. The groundwater is classified as HP, MP, and LP which was analysed using multivariate statistical techniques to determine the spatial variability of groundwater and to identify major variables affecting the water quality of groundwater.

CA resulted in three main cluster of sampling site with different characteristics. Continuously, DA determined only three parameters i.e.  $\text{Ca}^+$ ,  $\text{NO}_3^-$  and PH affording 73.33% correct assignation to discriminate between the clusters using forward stepwise mode from the original 24 parameters. Therefore, forward stepwise mode was proven to be useful in recognising the discriminate parameters in spatial variation of portable water quality as it begins with more significant variables than backward stepwise mode.

PCA was used to examine the root of each water quality parameter due to nature and anthropogenic activities based on three cluster regions. Eight varimax factors (VFs) accounted for 76.45% of the total variance in the data set were found. The largest source of variation (26.29%) appears to be from water quality parameters associated natural process (dissolution of rocks), point source pollution (industrial and municipal waste water) and non-point source pollution (mostly from agricultural activities). It is noteworthy that PCA confirm exactly the result of CA and determine the pollution source.

Therefore, the result of this study clearly demonstrates the

usefulness of multivariate statistical analysis in geochemistry. Additionally this result may be used to reduce the number of samples analysed both in space and time without much loss of information. This will assist the decision makers to identify priorities to improve water quality that has deteriorated due to pollution from various anthropogenic activities.

#### ACKNOWLEDGEMENTS

Authors are thankful to the Department of Mineral and Geoscience Terengganu for providing the hydrochemical data of groundwater. They are also thankful to the post graduate school, Universiti Sultan ZainalAbidin for supplying vehicle for site visit and data collection.

#### REFERENCES

- [1] Adam M. J (1998). The principle of multivariate data analysis in P. R Ashurst & M. J Dennis (Eds). Analytical methods of food authentication (p. 350) London: Blackie Academic professional.
- [2] Anders Berntell, World Water Week 2010, Stockholm International Water Institutes (SIWI), Stockholm, Sweden.
- [3] Brown, S. D, Skogerboe, R. K, & Kowalski, B. R. (1980). Pattern recognition assessment of water quality data: Coal strip mine drainage. *Chemosphere*, 9, 265-276. Doi: 10.1016/0045-6535(80)90003-X.
- [4] Chapagain, S. K, Shrestha, S., Nakamura, T., Pandey, V. P., & Kazama, F. (2009). Arsenic occurrence in Groundwater of Kathmandu Valley, Nepal. *Desalination and Water Treatment*. 4, 248-254.
- [5] Guler, C, Thyne, G. D, McCray, J. E, Turner, A. K (2002). Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeology J* 10; 455-474
- [6] Jackson, J. E. (1991). A user's guide to principal components. New York: Wiley
- [7] Juahir, H., Ekhwan, T. M, Zain, S. M, Mokhtar, M, Zaihan, J, & Ijankhushaida, M. J (2008). The use of chemometrics analysis as a cost-effective tool in sustainable utilization of water resources in the Langat River Catchment. *American-Eurasian journal of Agricultural & Environmental Sciences*, 4(1), 258-265.
- [8] Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151. Doi: 10.1177/001316446002000116.
- [9] Kannel PR, Lee S, Lee YS (2008) Assessment of spatial-temporal patterns of surface and ground water qualities and factors influencing management strategy of groundwater system in an urban river corridor of Nepal. *J Environ Manage* 86:595-604
- [10] Kaown D, Koh D-C, Mayer B, Lee K-K (2009) Identification

of nitrate and sulphate sources in groundwater using dual stable isotope approaches for an agricultural area with different land use (Chuncheon, mid-eastern Korea). *Agric Ecosystem Environ* 132 (3–4):223–231.

- [11] Kim J. O, Mueller C. W (1987). *Factor Analysis: Statistical Methods and Practical Issues*, Sage University Paper Series on Quantitative Applications in the Social Sciences, series no 07–014. Sage Publications, Beverly Hills.
- [12] Koh D-C, Mayer B, Lee K-S, Ko K-S (2010) Land-use controls on sources and fate of nitrate in shallow groundwater of an agricultural area revealed by multiple environmental tracers. *J Contam Hydrol* 118:62–78.
- [13] Levins, I, Gosk, G, (2008), Trace elements in groundwater as indicators of anthropogenic Impact. *Environ Geol* 55:285–290.
- [14] Massart, D. L., & Kaufman, L. (1983), the interpretation of chemical data by the use of cluster analysis. New York: Wiley.
- [15] National Research Council (2000). A Review of the Draft of the NCI-CDC Working Group to revise the “1985 Radio epidemiological Tables” National Academy Press 2101. Constitution Avenue, NW, Washington DC, 20418.
- [16] Nosrati K., M Van Den Eeckhaut (2012). Assessment of Groundwater Quality using Multivariate Statistical Techniques in Hashtgerd Plain, Iran. *J. Environmental Earth Science* 65 (1) 331-344.
- [17] Omo-Irabor OO, Olobaniyi SB, Oduyemi K, Akunna J (2008) Surface and groundwater water quality assessment using multivariate analytical methods: a case study of the Western Niger Delta, Nigeria. *Phys Chem Earth, Parts A/B/C* 33(8–13): 666–673.
- [18] Reghunath R, Murthy S. T. R & Raghavan B. R (2002). The utility of multivariate statistical techniques in hydrogeochemical studies. An example from Karnataka, India, *Water Research* 36. 2437-2442.
- [19] Shamsuddeen M. K, Sefie, A, Normi, A, Tawnie, I, Suratman, S, (2014). Impact of sea level rise to coastal groundwater at Kuala Terengganu, Terengganu. *Hydrogeology Research Centre. National Hydraulic Research institutes Malaysia*. Lot 5377, Jalan Putra, 43300, Selangor Malaysia.
- [20] Shrestha S, Kazama F (2007) Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. *Environ Model Softw* 22:464–475.
- [21] Simeonov, V., Einax, J. W., Stanimirova, I., & Kraft, J. (2002). Environmetric modelling and interpretation of river water monitoring data. *Anal Bio anal Chem*, 374,898–905.
- [22] Terceiro, P., Lobo-Ferreira, J. P., & Leitão, T. E. (2008). Análise da qualidade da água e questões de governança Albufeirão Alqueva. *Comunicação apresentada no 9º Congresso da Água-Água: Desafios de hoje, exigências de amanhã*. Cascais, Portugal. [http://www.aprh.pt/congressoagua2008/PDF/Lobo-Ferreira\\_Alqueva.pdf](http://www.aprh.pt/congressoagua2008/PDF/Lobo-Ferreira_Alqueva.pdf). Accessed 20 January 2009 (in Portuguese).
- [23] Vega, M., Pardo, R., Barrado, E., & Debán, L. (1998). Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research*, 32(12), 3581–3592. Doi: 10.1016/S0043-1354 (98) 00138-9.
- [24] Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *J Am Stat Assoc*, 69, 236-244.
- [25] WHO (1997), World Health Organisation. *Guidelines for drinking water quality*, Geneva. Second Edition, Volume 3.