Groundwater Hydrochemistry Assessment of North Dhi-Qar Province, South of Iraq Using Multivariate Statistical Techniques

Salam Hussein Ewaid ¹, Kareem Ghafel Mhajej ², Salwan Ali Abed ³, Nadhir Al-Ansari ⁴

¹Technical College of Shatra, Southern Technical University, Basra 61001, Iraq
²³College of Science, University of Al-Qadisiyah, P.O.Box.1895, Iraq
⁴Luleå University of Technology, 97187 Luleå, Sweden
salwan.abed@qu.edu.iq

Abstract

Multivariate statistical techniques including correlation, principal component analysis (PCA), and cluster analysis (CA) were applied in this study to assess the groundwater hydrochemistry of the North Dhi-Qar Province, South of Iraq. The water samples were taken from 16 water wells in the period from January to October 2020 and ten water variables were analyzed, pH, electrical conductivity (EC), total dissolved solids (TDS), Ca⁺², Mg⁺², Na⁺, K⁺, Cl⁻, SO₄⁻², and HCO₃⁻. The results obtained from Spearman’s correlation showed that the positive and negative correlation of P < 0.05 between water variables is different at two-tailed grades. Results from the PCA have shown that approximately 85% of the overall variance has been clarified by the three PCs achieved. The main causes of variation in the hydro-chemical properties of water samples of the wells can therefore be determined. PC 1 represents about 36.75% of the variance and holds a high loading for EC, HCO₃⁻, Cl⁻, K⁺, and EC. PC2, which explains 35% of the total variance, has high loadings for EC, Na⁺, TDS, Ca⁺², and SO₄⁻². PC 3 shows high loadings for pH, which accounts for 13.235% of the variation in the water hydrochemistry. The hierarchical cluster analysis (CA) grouped the 16 sampling wells into three clusters of similar water quality characteristics. In the analysis of space changes in water quality, this research demonstrates the use of multivariate statistical methods for the interpretation of complex data sets. This will thus improve future studies preparation.

Keywords: cluster analysis; groundwater; principal component analysis; Dhi-Qar

1. Introduction

Water is one of the most essential resources in the world required for the nature of life in its pure form. With its accompanying reinforcement of industrial activities, the growing population expansion has a negative effect on the quality of soil and surface water worldwide. Indiscriminate discharge of urban and industrial waste, combined with agricultural runoffs, substantially contributes to water pollution (Ewaid and Abed, 2017).
In deciding the quality of water supplies in this region, the geological nature of a certain area play a significant part. For example, the dissolution of minerals from dominant geological forms usually decides whether the surface water or the groundwater of that particular region has a chemical properties. In discussing the overall quality of water supplies in any given area, both geological and anthropogenic factors are relevant (Ewaid, 2018).

Within Dhi-Qar Province, South of Iraq, Multiple ground-based activities such as the rush of agricultural land pesticides have rendered important contributions in the water bodies in the region to pollutant concentrations (Ewaid et al., (2020a).

The consistency of water sources have been measured by direct measurement of contaminant levels most scientific studies in Iraq.

For example, Kadhum, et al., (2020) measured the physicochemical proprieties of the Euphrates River, Iraq. Statistical methods and descriptive statistics have been used by to evaluate the sources of contaminants and other nutrients in surface and groundwater (Abed et al., 2019; Ewaid et al., 2020b). In addition, a combination of descriptive and multivariate statistical technology is a valuable method for characterizing levels of pollutants in water basins with compound forms of land use and pollution history (Ewaid, 2018).

Some of multivariate statistical techniques used in previous evaluations of water quality are principal components analysis (PCA), hierarchical cluster analysis (HCA), and discriminant analysis (DA) (Ewaid, 2018; Kadhum, et al., 2020; Ewaid et al., 2020a; Abed et al., 2019; Ewaid et al., 2020b). Multivariate statistical techniques are now gaining ground in explanation of hidden structures in water quality data in Iraq (Ewaid and Abed, 2017). Therefore, the key reason for this paper is to explore the use of multivariate statistical methods in the assessment of water quality in northern Dhi-Qar in the south of Iraq.

2. Materials and Methods

2.1 The Study Area

Due to increased demand for a resource with a gradually more reduced availability, the hydrological system in North Dhi-Qar has undergone considerable changes over the past 40 years. The climate has affected by the Tigris River and the Gharraf River, which crosses the area, in various ways.

These rivers are the primary source of water. With the loss of their water quality, they have dropped to less than a third of their usual capacity. The groundwater at North Dhi-Qar Province southern part of Iraq is situated between 46° 00 / – 46° 30 / E and 31° 00 / – 32° 00 / N, (Ewaid, 2018) (Figure 1).
2.2 Samples Preparation and Analysis
To study the chemical parameters, samples were taken from 16 wells from January to October 2020 in the study area. After 15 minutes of pumping, the samples have been collected after stabilizing the water temperature to eliminate groundwater in the hydraulic structure. The samples were gathered with 1 L clean bottles of polyethylene. The bottles were first washed and labeled with an identification number by washing acidified. In line with the samples position, the bottle number was then reported on the sample datasheet. GPS was used to determine the location of the sample field.

All the samples were placed in an ice box at < 5°C and then moved to the laboratory for chemical parameter analysis using standard test procedures (Baird, 2017). The pH and EC are measured in the field using a multi parameter immediately after sampling by WTW (P3 Multi Line pH/LF-SET). The EDTA titration method for Ca and Mg is used for the laboratory study, while flame photometry tested the K and Na contents. UV-visible spectrophotometry was used to measure sulfate levels.

For determining bicarbonate and chloride, titrimetric has been used. The gravimetric analysis procedure used for measuring TDS.

2.3 Multivariate Statistical Analysis
Using various multivariate statistical techniques, the representation of complex data matrices can be simplified and large datasets organized to provide significant insight (Abed et al., 2019). In the current analysis, the chemical parameters of groundwater samples have been assessed by two multivariate statistical techniques. The statistical packages SPSS 25 has been used for the descriptive and multivariate statistical analysis.
2.3.1 The principal component analysis (PCA).
For the treatment of hydro chemical data, the PCA method was used. The PCA technique is commonly used to transform the original data set into a smaller set of uncorrelated variables known as the principle components (PCs). It uses own values and prophetic vectors linked to a covariance matrix to produce PCs by multiplying the correlated variables by the prophetic device (Kumar et al., 2018). By extracting the most relevant variables, the PCs can be used to interpret the original data sets with minimal knowledge loss (Ling et al., 2017).
Although PCA is an investigative and descriptive tool, its purpose is to define the principal factors controlling soil-water chemistry (Zafar and Ahmad 2018). Ten hydro chemicals measured variables (EC, TDS, Ca$^{2+}$, Mg$^{2+}$, Na$^+$, K$^+$, Cl$^-$, SO$_4$, and HCO$_3$) were utilized in this analysis.

2.3.2 The cluster analysis (CA).
Cluster Analysis is a multivariable technique used to organized groups known as clusters for a number of observations or objects. The findings within each group are identical, but the clusters are different. This technique can be used to group the popular data on water quality where the water of a specific quality is indicated by each cluster.
A cluster analysis for water chemistry data was conducted to group the wells in terms of water quality (Herojeet et al., 2017). The objects are clustered in such a way that related wells fall into the same class. The classification scheme for measuring the similarity, with Ward's method for linkage, produces the most distinctive groups, where each member of the group is identical to its fellow members rather than to any member outside the group, using Euclidean distance. (Stand line distance between two points in the dimensional area, identified by the variables) (Bouguerne, 2017). In this study, the 16 wells included in the study were used.

3. Results and discussion
3.1 Statistical analysis

Ten chemical parameters were examined. The statistical analysis of parameters are summarized in Table 1. Some parameters have shown high values of standard differences to fluctuate arbitrarily in their concentration in groundwater.

| Wells | pH  | EC   | TDS  | Ca$^{2+}$ | Mg$^{2+}$ | Na$^+$ | K$^+$ | Cl$^-$ | SO$_4$ | HCO$_3$ |
|-------|-----|------|------|-----------|-----------|--------|-------|--------|--------|---------|
| 1     | 7.3 | 13150| 28993| 834       | 793       | 786    | 244   | 104    | 741    | 373     |
| 2     | 7.5 | 58500| 40400| 300       | 1213      | 5539   | 897   | 7384   | 1017   | 6100    |
| 3     | 6.9 | 68509| 47956| 914       | 88        | 16293  | 99    | 201    | 9161   | 282     |
3.2 Correlation analysis

In order to carry out factor analysis, the correlations between parameters must be determined (Table 2). In order to quantify the variability between the individual variables of water quality pairs, the coefficients are calculated.

In the promotion of science, determination of correlational coefficients between different variables is crucial and opens new boundaries to expertise, thereby reducing the level of ambiguity in decision making. Correlation analysis also acts as a statistical instrument for water quality determination to know the ions that govern water chemistry (Khaledian et al., 2018).

The positive pH-Ca correlation shows that, as pH increases, Ca increases or vice versa. This demonstrates that when the pH rises, the concentration of hydrogen ions (H<sup>+</sup>) decreases which replaces Ca in the mineral (clay) structure, thus reducing H<sup>+</sup> and increasing Ca in water is the mechanism of the interchange of ions. This can only be attributed to a mechanism of ion exchange.
EC shows a strong Ca-K correlation and a very good TDS and Mg correlation. This indicates the similar hydro chemical features of these parameters in the research field (Zheng et al., 2015).

TDS is closely linked to Ca, Mg, Na and EC. Ca has a positive pH and K affinity and a positive EC and Mg association. Mg and sulfate, pH and Ca are in a positive relationship. The relationship between ionic character parameters was higher than that of low Ion character parameters. The variance in a relationship generally shows the complexity in the consistency of the soil and the impact of soil-water interactions.

**Table 2. Correlation coefficients for hydro chemical parameters**

|       | pH  | EC  | TDS | Ca   | Mg   | Na   | K    | Cl   | SO4  | HCO3 |
|-------|-----|-----|-----|------|------|------|------|------|------|------|
| pH   |     |     |     |      |      |      |      |      |      |      |
| EC   | -0.378 |     |     |      |      |      |      |      |      |      |
| Sig. (2-tailed) | 0.149 |     |     |      |      |      |      |      |      |      |
| TDS  | -0.254 | .768* |     |      |      |      |      |      |      |      |
| Sig. (2-tailed) | 0.342 | 0.001 |     |      |      |      |      |      |      |      |
| Ca   | 0.011 | 0.307 | .669* |      |      |      |      |      |      |      |
| Sig. (2-tailed) | 0.967 | 0.247 | 0.005 |      |      |      |      |      |      |      |
| Mg   | 0.043 | 0.181 | .535* | 0.229 |      |      |      |      |      |      |
| Sig. (2-tailed) | 0.874 | .502* | 0.033 | 0.394 |     |      |      |      |      |      |
| Na   | -0.295 | .669* | .854* | .650* | 0.261 |      |      |      |      |      |
| Sig. (2-tailed) | 0.267 | 0.005 | 0.000 | 0.006 | 0.329 |     |      |      |      |      |
| K    | 0.127 | .616* | .498* | -0.035 | 0.423 | 0.116 |      |      |      |      |
| Sig. (2-tailed) | 0.640 | 0.011 | 0.049 | 0.897 | 0.102 | 0.668 |     |      |      |      |
| Cl   | 0.108 | .547* | 0.338 | -0.243 | 0.384 | 0.046 | .900* |      |      |      |
| Sig. (2-tailed) | 0.690 | 0.028 | 0.201 | 0.365 | 0.142 | 0.865 | 0.000 |     |      |      |
| SO4  | -0.373 | 0.350 | .606* | .599* | 0.149 | .764* | -0.099 | -0.089 |      |      |
| Sig. (2-tailed) | 0.155 | 0.184 | 0.013 | 0.014 | 0.583 | 0.001 | 0.716 | 0.742 |     |      |
| HCO3 | 0.130 | .593* | 0.366 | -0.207 | 0.325 | 0.051 | .953* | .966* | -0.160 |      |
| Sig. (2-tailed) | 0.632 | 0.016 | 0.163 | 0.442 | 0.219 | 0.852 | 0.000 | 0.000 | 0.554 |     |

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
Bold labels strong significant correlation**
3.3 Factor analysis
The FA approach is used for the water quality assessment, in order to consider the causes of changes in the water quality by means of various parameters (Cronk, 2017). FA produces a variable without observation that compiles differences into three variables or more. The factors were extracted and rotated by varimax using the principle component (PC) analysis process.

Table 3 shows that three factors explained 85% of the total variance, using the Kaiser criterion with eigenvalues more than or equal to 1 (Zhang, et al., 2018). A scree plot in Fig. 2 displays the own values sorted as a function of the factor number from large to small. It was observed that other elements were omitted after the third factor (start the cole in the downward curve) (Fig. 2; Table 3). In order to classify the key causes of fluctuation in the hydrochemical properties of the well samples, the obtained three major components clarified approximately 85 percent of the total variance. The PC1 represents about 36.75% of the variance and has a high power load for electrical conductivity, HCO₃, Cl, K, EC. The PC2, which explains 35% of the variance, has high loadings for EC, Na, TDS, Ca, and SO₄. PC3 shows high loadings for pH, which accounts for 13.2% of the hydro chemical variance.

Table 3 shows the description of the loads factor, their values and variances. In decreasing order, the factors are given depending on the variance percentage.

Table 3. Loadings of 10 variables on 3 significant PCs in the varimax rotated component matrix.

|       | 1      | 2      | 3      |
|-------|--------|--------|--------|
| HCO₃  | .985   | -.058  | .041   |
| Cl    | .964   | -.054  | .041   |
| K     | .961   | .083   | .119   |
| EC    | .651   | .520   | -.431  |
| Na    | .107   | .904   | -.231  |
| TDS   | .436   | .876   | -.068  |
| Ca    | -.188  | .869   | .208   |
| SO₄   | -.135  | .808   | -.273  |
| pH    | .042   | -.213  | .867   |
| Mg    | .410   | .422   | .438   |
| Eigenvalue | 4.304 | 3.015 | 1.181 |
| % of Variance | 36.748 | 35.012 | 13.235 |
| % Cumulative of Variance | 36.748 | 71.760 | 84.995 |

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization
3.4 Cluster analysis
The research area was defined by multivariate statistical techniques. These statistical methods assist in the selection of water samples from various locations for many forms of grouping studies.

Positions for the sampling of wells are clustered with the chemical similarity (Zhang, et al., 2018). In this analysis, the SPSS 25 program was used in order to group the water qualities of the wells in clusters and generate a dendrogram (Fig. 3). A tree-structured diagram is the result of a hierarchical equation. The diagram displays the results. The product of a clustering is either the distance between the clustered rows or columns or the similarity between them, depending on the values chosen or on the definition of each measured distance.

The values of water quality parameters tested using mean and standard deviation values have been standardized. This was achieved to reduce variations in each parameter's quantitative values. Using square Euclidean distance, the linking distance between the wells was calculated and Ward's clustering method used. The 16 wells are broken down into 3 main clusters using the cluster analysis, as shown in Fig. 3.
3.5 Conclusion
The analysis of hydrochemical data from sixteen wells was using multivariate statistical techniques including correlation analysis, factor analysis, and cluster analysis to obtain knowledge that was not available in the study field at first glance. Ten parameters including pH, EC, TDS, Ca$^{+2}$, Mg$^{+2}$, Na$^+$, K$^+$, Cl$^-$, SO$_4$ and HCO$_3$ were determined. Most ions are concentrated within the reasonable limits provided by WHO except where TDS is often slightly higher than the permissible limits. Correlation analyzes in the study area show that the relation between parameters with high ion character is higher than those with less ionic character and that the shift in relation indicates the complexity in the quality of soil water and the effects of soil-water interaction. FA was used to split the 16 wells analyzed into three key factors representing the minimum information loss data collection.

The parameters were grouped into three main clusters through hierarchical cluster analysis. This research therefore demonstrates the efficacy of multivariate statistical methods as an efficient instrument of exploration for an analysis of complex water quality knowledge sets, and establishes key factors and mechanisms, which are important and successful for water quality control in the study area to track the chemical composition of groundwater.
It should, however, be noted that hydrochemical processes cannot arise directly from multivariate statistical technologies, but they may provide insight into the factors controlling such processes.

References

[1] Ewaid, S. H. (2018). Irrigation water quality of Al-Gharraf Canal, south of Iraq. In Journal of Physics: Conference Series (Vol. 1003, No. 1, p. 012006). IOP Publishing.

[2] Ewaid, S. H., & Abed, S. A. (2017). Water quality assessment of Al-Gharraf River, South of Iraq using multivariate statistical techniques. Al-Nahrain Journal of Science, 20(2), 114-122.

[3] Kadhum, S. A., Abed, S. A., Ewaid, S. H., Chabuk, A., Al-Ansari, N., & Jassim, A. K. (2020). Multivariate Analysis and Geochemical Assessment of heavy Metals Pollution in Surface Sediment from Euphrates River, Iraq: Multivariate Analysis and Geochemical Assessment of heavy Metals Pollution in Surface Sediment from Euphrates River, Iraq. Pollution Research, 39, S262-S267.

[4] Ewaid, S. H., Abed, S. A., Al-Ansari, N., & Salih, R. M. (2020a). Development and Evaluation of a Water Quality Index for the Iraqi Rivers. Hydrology, 7(3), 67.

[5] Al-Dabbas, M. A., Mhajej, K. G., & Kadum, W. H. (2020). Ground Water Hydrochemical Assessment at Rifai-Nasiriya district, Thi Qar Governorate-South Iraq. In Journal of Physics: Conference Series (Vol. 1660, No. 1, p. 012076). IOP Publishing.

[6] Salam Hussein Ewaid et al 2020 J. Phys.: Conf. Ser. 1664 012143.

[7] Ahmed Sabah Al-Jasimee et al 2020 J. Phys.: Conf. Ser. 1664 012141.

[8] Salam Hussein Ewaid et al 2021 IOP Conf. Ser.: Earth Environ. Sci. 722 012008

[9] Salwan Ali Abed et al 2019 J. Phys.: Conf. Ser. 1294 072025.

[10] Kumar, B., Singh, U. K., & Ojha, S. N. (2018). Evaluation of geochemical data of Yamuna River using WQI and multivariate statistical analyses: a case study. Inter. J. of River Basin Manag. 1-13.

[11] Ling, T. Y., Soo, C. L., Liew, J. J., Nyanti, L., Sim, S. F., & Grinang, J. (2017). Application of multivariate statistical analysis in the evaluation of surface river water quality of a tropical river. Journal of Chemistry, 2017.

[12] Zafar MU, Ahmad W (2018) Water Quality Assessment and Apportionment of Northern Pakistan by Multivariate Statistical Techniques, A Case Study. Int J Hydro 2 (1): 00040.

[13] Herojeet, R., Rishi, M. S., Lata, R., & Dolma, K. (2017). Quality characterization and pollution source identification of surface water using multivariate statistical techniques, Nalagarh Valley, Himachal Pradesh, India. Applied Water Science, 7(5), 2137-2156.
[14] Baird, R. B. (2017). Standard Methods for the Examination of Water and Wastewater, 23rd. Water Environment Federation, American Public Health Association, American Water Works Association.

[15] Bouguerne, A., Boudoukha, A., Benkhaled, A., & Mebarkia, A. H. (2017). Assessment of surface water quality of Ain Zada dam (Algeria) using multivariate statistical techniques. Inter. J. of River Basin Management, 15(2), 133-143.

[16] Khaledian, Y., Ebrahimi, S., Natesan, U., Basatnia, N., Nejad, B. B., Bagmohammadi, H., & Zeraatpisheh, M. (2018). Assessment of water quality using multivariate statistical analysis in the Gharaso River, Northern Iran. In Urban ecology, water quality and climate change (pp. 227-253). Springer, Cham.

[17] Zheng, L. Y., Yu, H. B., & Wang, Q. S. (2015). Assessment of temporal and spatial variations in surface water quality using multivariate statistical techniques: A case study of Nenjiang River basin, China. Journal of Central South University, 22(10), 3770-3780.

[18] Cronk, B. C. (2017). How to use SPSS®: A step-by-step guide to analysis and interpretation. Routledge.

[19] Zhang, Y., Li, Y., Da, W., Yu, M., Quan, Q., & Li, Y. (2018). Application of multivariate statistical methods in the assessment of water quality in selected locations in Jialing river basin in Guangyuan, China. Water Science and Technology: Water Supply, ws2018058.