JEE Journal of Ecological Engineering

Journal of Ecological Engineering 2022, 23(6), 189–204
https://doi.org/10.12911/22998993/148195
ISSN 2299–8993, License CC-BY 4.0

Received: 2022.03.28
Accepted: 2022.04.13
Published: 2022.04.20

Application of Water Quality Index and Multivariate Statistical Techniques to Assess and Predict of Groundwater Quality with Aid of Geographic Information System

Ammar S. Dawood[1*], Mushtak T. Jabbar[2], Hayfaa H. Al-Tameemi[3], Eric M. Baer[2]

[1] Civil Engineering Department, College of Engineering, University of Basrah, Basrah, Iraq
[2] Geology Department, Earth Sciences, HCC, Seattle, Washington, 98198, USA
[3] Department of Soil Science and Water Resources, University of Basrah, Iraq
* Corresponding author's e-mail: ammars.dawood@yahoo.com

ABSTRACT
In this study, the groundwater quality and spatial distribution of the Basra province in the south of Iraq was assessed and mapped for drinking and irrigation purposes. Groundwater samples (n = 41) were collected from deep wells in the study area to demonstrate, estimate and model the Water Quality Index (WQI). The analysis of water samples integrated with GIS-based IDW technique was used to express the spatial variation in the study area with consideration of WQI. The physicochemical parameters, including pH, sodium (Na^+), electrical conductivity (EC), chloride (Cl^-), total dissolved solids (TDS), calcium (Ca^{2+}), nitrate (NO^{3-}), sulfate (SO_4^{2-}), magnesium (Mg^{2+}), and bicarbonate (HCO_3^-) were identified for groundwater quality assessment. The results of calculated WQI classify groundwater into three sorts. The results of WQI showed that 2.5%, 2.5% and 95% of the groundwater samples were classified as poor/very poor/unsuitable for drinking, respectively. The GIS tools integrated with statistical techniques are utilized for spatial distribution and description of water quality. Correlation analysis of groundwater data revealed that some parameters have actually a relationship that is strong with the other parameters and they share a common source of origin. Multivariate statistical techniques, especially cluster analysis (CA) and factor analysis (FA), were applied for the evaluation of spatial variations of forty-one selected groundwater samples. Cluster analysis confirmed that some different locations of wells have comparable sourced elements of water pollution, whereas factor analysis yielded three factors which are accountable for groundwater quality variations, clarifying more than 72% of the total variance of the data and permitted to group the preferred water quality. Multi-Layer Perceptron (MLP) models were applied in modeling the water quality index. Comparing different result values of the MLP network suggested that the values of MSE and r for the selected model are 0.1940 and 0.9998, respectively. Finally, it can be revealed that the MLP network precisely predicted the output, i.e. the WQI values.

Keywords: cluster analysis, water quality, groundwater, factor analysis, WQI, GIS, multi-layer perceptron.

INTRODUCTION

In comparison to surface water, groundwater is considered to be cleaner and safer. Groundwater is more usually consumed than surface water in arid and semi-arid climate zones. Increased population density as a result of urbanization and agricultural pesticides, the quality of groundwater is degrading. Other factors contributing to deterioration include construction activities such as landslides and the slowing down of rainwater seepage into the earth [Ramesh and Elango, 2012]. Groundwater quality can be assessed for a variety of purposes, such as irrigation or drinking, using a variety of applications [Wilcox, 1955; Ayers and Westcot, 1985; Aller et al., 1987; Simsek and Gunduz, 2007; Boyacioglu, 2010]. The World Health Organization (WHO) and the Environmental Protection Agency (EPA) have established criteria for the quality monitoring of water resources, which can be followed by tracing the parameters relevant to those requirements. The traditional process allows for a simple and complete assessment based on the comparison of water quality criteria. However, it is also extremely difficult to evaluate existing data in order for decision-makers to plan water resource plans.

The principle of water quality index (WQI) was suggested firstly by [Horton, 1965] to represent the water quality class. WQI was improved in 1970 by the National Sanitation Foundation (NSF). WQI offers a single value that is numerical to the user. It is a reflection of a composite effect of the factors that contributed to the water quality for any water-system [Kakate and Sarma, 2007]. It is necessary for analyzing water quality parameters at a certain area and time; overall, water quality is dependent on varied physicochemical parameters. WQI is an effective tool to reveal the water quality as well as it is one of the well-known strategies to communicate the water quality information to the citizens and related policy makers. For that reason, it becomes an important parameter to evaluate and manage groundwater [Chauhan et al., 2010; Saha et al., 1991]. For decision-making and management aims in water resources, WQI is very helpful. The WQI development for each area is a procedure that is important for the progress of that area.

Multi-Layer Perceptron (MLP), a type of neural network (NN), is a design that is simplified biological structure observed in a human brain. Three factors are essential in any MLP model: nodes structure, network topology, and learning algorithm which is applied to establish the weights of the neural network (Rojas, 2013). In terms of MLP topology structures, NN can easily be split into two sorts: feed-forward networks and recurrent networks. Two types of a learning algorithm, which are used in the NN, include supervised and unsupervised learning. One of the most important NN models is the Back-propagation (BP). In the BP networks, the learning rule makes use of the steepest descent technique by which the output errors of the BP network have always been to transmit back to change the weights of interconnections in the network, so they can minimize the total error. The standard type of MLP has an input layer, a hidden layer, and an output layer. The primary variations between the various types of MLP are the arrangement of neurons in the network architecture, the strategy to obtain the weights and the function of inputs as well as the function of the outputs.

In recent years, it was found that there great attention is drawn to using NN which can be a kind of data-driven models. This attention is a result of the fact that those MLP models are simplified and have high levels of forecast accuracy. MLP models are able to find the complex forms and the connections between the inputs and the output data of the model. This ability of MLP models is due to its efficiency of learning non-linear static or dynamic actions between the variables.

Geographic information system (GIS) is an extensive computer-dependent technique, which appears in pollution studies and water resource management. It has a significant role in mapping the suitability of water in any area. Multivariate statistical methods including principal component analysis (PCA), factor analysis, and cluster analysis contribute to the reliable and accurate modelling of water quality parameters and spatial distributions through the combination of laboratory analysis and geographic data. [Vega et al. 1998; Reghunath et al. 2002; Simeonov et al. 2004; Sofie et al. 2006].

Therefore, the aims of the current study are using GIS, WQI and statistical techniques for the characterization of groundwater WQI. Several physicochemical characteristics of groundwater were analyzed from different wells in the Zubair district. The physicochemical data of groundwater samples was examined using SPSS version 22.0 to prepare the statistical analysis. For a pictorial representation of data, Inverse Distance Weightage (IDW) approach has been utilized. The other aim of this research was to examine the water quality factors (such as pH, TDS, EC, Na^+, Ca^{2+}, K, Mg^{2+}, SO_4^{2-}, NO_3^-, Cl^-, and HCO_3^-) that affect the water quality of the Shatt Al-Arab River through the application of WQI and estimate it using MLP.

MATERIAL AND METHODS

Study area

The investigation area was located southwestern part of Basrah (in Zubair district) between the 30° 10′ to 30°26′ N and longitudes 47°39′ to 47°56′ E (Fig. 1). The investigated water samples from wells in this study area are located within the Zubair district in the south of Mesopotamian Zone which is located in the Stable Shelf area. The southern boundary of the Zubair district is either situated in the Al-Batin Fault or long and transversal fault in Kuwait.

In the study area, there is a desert climate. In winter, there is more rainfall than in summer. The temperature of the Zubair district ranges from 17.5 to 43.6°C as average high temperatures and ranging from 6.5 to 27.6°C as average low temperatures. The climate is semi-arid with the annual average minimum temperature (17.4°C) and annual average maximum temperature (31.72°C) values in winter and summer, respectively. The average annual temperature in the Zubair district is 24.5°C. The average rainfall of the Zubair district is 139 mm with negligible rains in summer.

Groundwater sampling

The groundwater sampling was performed in the Zubair district, south of Iraq. The sampling well locations (n = 41) were established inside the study region, so that the regional distribution of water quality can be mapped. In 2015, samples were gathered from a variety of locations.. The analysis of physicochemical for water samples was completed utilizing the standard methods recommended by APHA (1998) in this study. The water quality of the samples composed of 11 physicochemical parameters. These parameters include pH, sodium (Na+), electrical conductivity (EC), chloride (Cl^-), total dissolved solids (TDS), calcium (Ca^{2+}), nitrate (NO_3^-), sulfate (SO_4^{2-}), magnesium (Mg^{2+}), and bicarbonate (HCO_3^-). Some parameters such as pH, EC and TDS were measured in the field, making use of portable digital meters, whereas the major ions such as Na^+, Ca^{2+}, K, Mg^{2+}, SO_4^{2-}, NO_3^-, Cl^-, and HCO_3^- were analyzed in the laboratory. The color composite maps were prepared using Arc GIS 10.3. The data were statistically analyzed using SPSS version 22.0 software. The prediction of WQI by using neural networks was done with utilizing the Weka software.

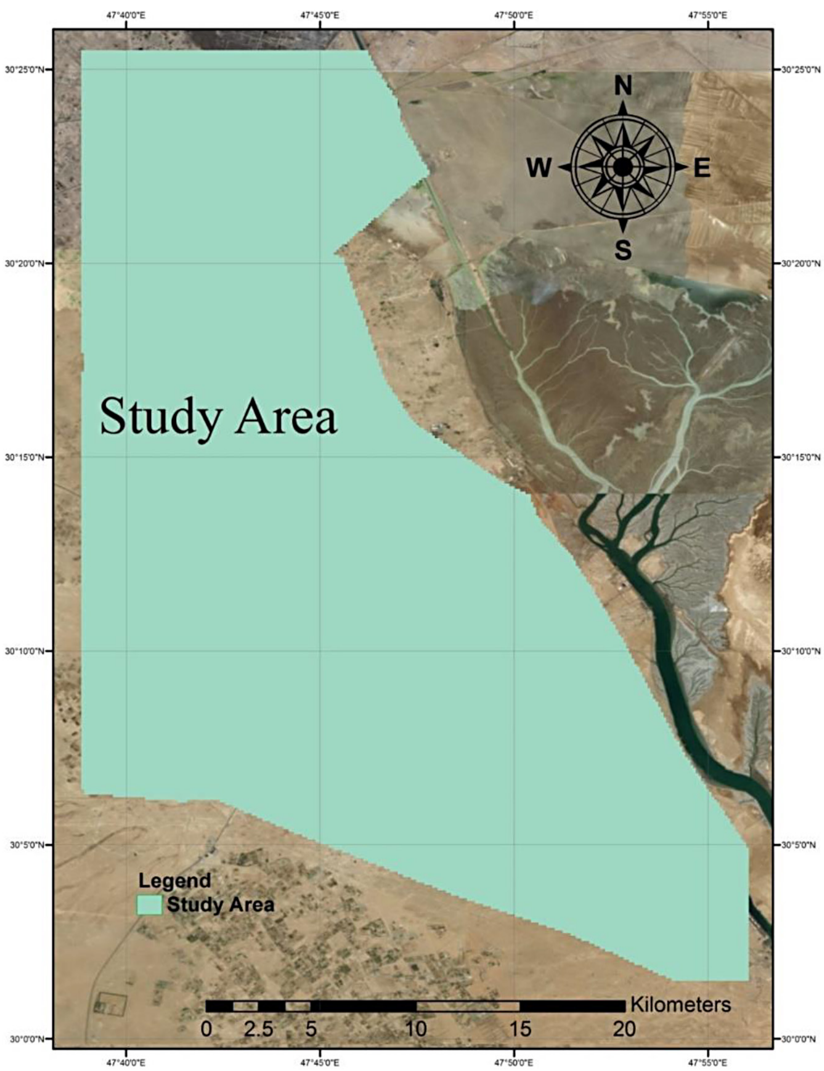

Figure 1. Location map of the study area

Water quality index calculation

The process of water quality index calculation assigned a weight for each of the water quality parameters. These weights are assigned in accordance with its related significance with total quality of water for drinking. The computed Wi value of each physicochemical parameter was given in Table 1. A weight of 5, which is a maximum weight in Table 1, is allocated to nitrate parameter, considering its necessity in water quality assessments. A weight of 3 is given to water quality parameters particularly pH, sodium, chloride and sulfate, whereas a minimum weight of 2 is given to other water quality parameters which can include calcium, magnesium, and bicarbonate because they play a relatively less appreciable role in the assessment of water quality [Srinivasamoorthy et al., 2008; Vasanthavigar et al., 2010]. The relative weight is computed as follows (Eq. 1).

$$W_i = \frac{w_i}{\sum_{i=1}^{n} w_i} \tag{1}$$

where: W_i is denoted the relative weight, w_i expresses the weight of each parameter, used and n is the number of physicochemical parameters

The scale of quality rating (q_i) for each physicochemical parameter is applied by dividing the concentration of the selected groundwater sample by its specific standard as reported by WHO standards, and then the results are multiplied by 100. Finally, the water quality index is calculated by the sum of the sub-indices according to the subsequent equations

$$SI_i = W_i \times q_i \tag{2}$$

$$\mathrm{WQI} = \sum SI_i \tag{3}$$

where: q_i is the quality rating scale, SI_i indicates the sub-index of the i^{th} parameter. The calculated WQI values tend to be categorized into five types, as shown in Table 2.

Table 1. Physicochemical parameter relative weights

Physico-chemical parameters	WHO standards	Weight (wi)	Relative weight (Wi)
pH	8.5	3	0.1304
Na^+ (mg/L)	200	3	0.1304
Ca^{2+} (mg/L)	75	2	0.0870
Mg^{2+} (mg/L)	50	2	0.0870
SO_4^{2-} (mg/L)	250	3	0.1304
Cl^- (mg/L)	250	3	0.1304
NO_3^- (mg/L)	45	5	0.2174
HCO_3^- (mg/L)	120	2	0.0870
Total	-	29	1.0000

Table 2. Water quality sorts matching WQI values

Type of water	WQI
Excellent	< 50
Good	50–100
Poor	100–200
Very poor	200–300
Unfit for drinking	> 300

Statistical analysis

The simple correlation coefficient refers to the degree of linear relation concerning any of two parameters. The Pearson linear correlation matrix can be generated by analyzing the results of the studied parameters. The main purpose of using cluster analysis, which is a type of multivariate method, is to categorize the element system into groups or clusters according to their similarities. The popular approach in cluster analysis, in which clusters tend to be created sequentially, is called hierarchical clustering. Probably, the most elements being comparable are first sorted, and these groups which can be preliminary merged associated with their particular similarities. Ultimately, the similarity reduces all the subgroups and merged tend to be combined into a single cluster. Single linkage method was used as cluster analysis to water quality data. A hierarchical clustering procedure, usually referred to as a dendrogram, is graphically displayed in the results using a tree diagram [Johnson and Wichern 2002; Alvin 2002]. Factor analysis is developed for the transforming the variables in the original study into new un-correlated variables, which are linear combinations of the original variables. Principal component method (PCA) is applied on different factors for extraction process. The PCA axes are rotated so that less important variables don't make as much of a difference (Alvin 2002).

Multi-layer perceptron modeling

The modeling of Multi-Layer Perceptron is often indicated by (i, j, k) network structure, in which i represents the number of neurons in the first layer (input layer); j represents the number of neurons in the second layer (hidden layer) and k represents the number of neurons in the third layer (output layer). MLP model works by using a supervised learning technique, in which the learning process is depending on the concept of error minimizing anywhere between the output data developed by the network. The error minimizing strategy is carried out through optimizing the weights and the real output data. To make predictive models using the previously mentioned systems, the Na^+, Ca^{2+}, K, Mg^{2+}, SO_4^{2-}, NO_3^-, Cl^-, and HCO_3^- were used as input parameters, and the water quality index was chosen as the only output parameter. This research can provide various strategies compared to comparable studies for investigating the application of MLP models based on these eight parameters which are related to water quality.

Figure 2 indicates the structure of the developed MLP network. In the present research, the MLP structure has one hidden layer. The layer that is first in MLP network is an input layer that comprises the input vectors. In this layer, the input vectors consists of Na^+, Ca^{2+}, K, Mg^{2+}, SO_4^{2-}, NO_3^-, Cl^-, and HCO_3^-. The second layer in the MLP model is a hidden layer, in which the number of neurons could be varied to achieve the most advantageous model structure and additionally to develop its forecasting ability. In the last layer of the MLP model, it is called the output layer.

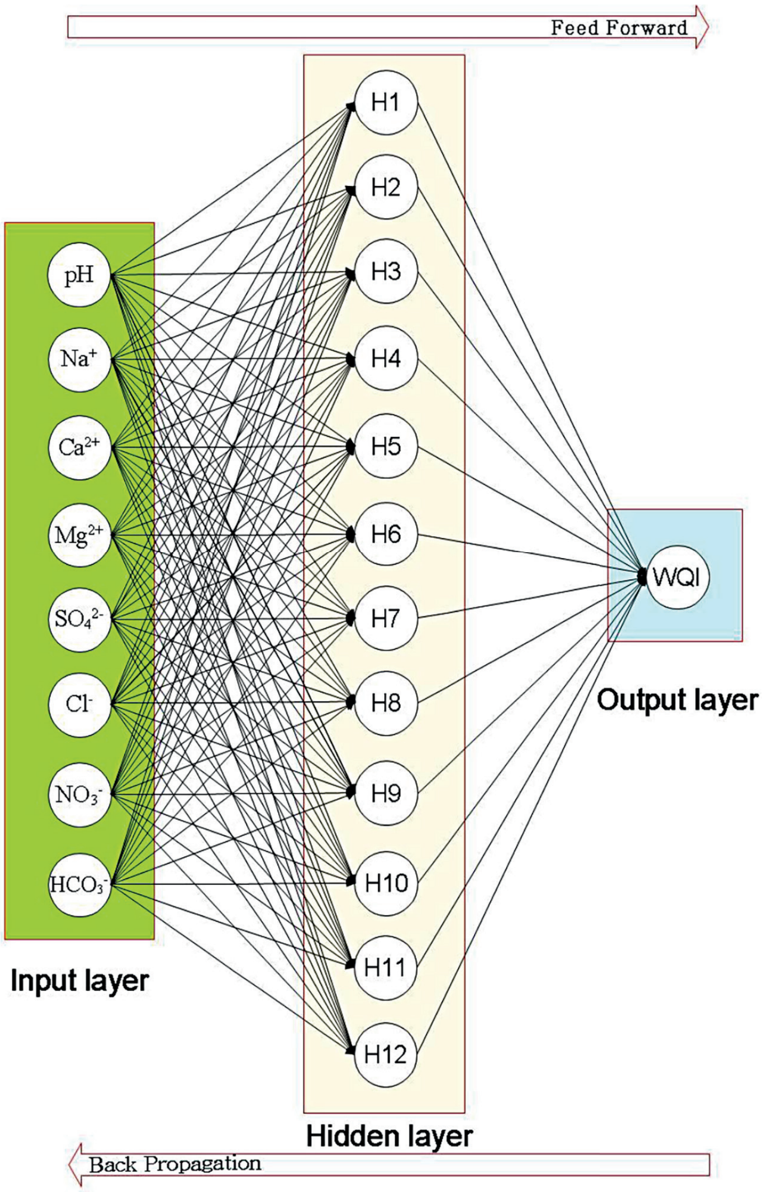

Figure 2. The developed MLP network's structure

Table 3. MSE value according to HLN on a hidden layer

Order	Model	MSE	R	Number of neurons on hidden layer
1	MLP1	0.036360	0.9998	4
2	MLP2	0.025212	0.9998	6
3	MLP3	0.030736	0.9998	8
4	MLP4	0.067456	0.9981	10
5	MLP5	0.089063	0.9995	12
6	MLP6	0.027414	0.9999	14
7	MLP7	0.222014	0.9899	16

There is only one output in this layer which is WQI.

In this study, the proposed MLP network uses an algorithm in the training system, known as backpropagation, in which the error is based on the comparison of the model output and the required output. Then, the produced error is coming back to the hidden and input layers for the next training processes. The network training operation is complete when the error is decreased to the value provided by the user [Faizollahzadeh Ardabili et al., 2017].

In this research, firstly, the percentages of data are assigned for training, testing, and validating that were selected as 70, 15, and 15, for training, testing, and validating process. The back propagation algorithm was carried out in the training process and the function of sigmoid transfer, was applied as a transfer function, since it permits the network to estimate the non-linear actions of the procedure. One of the abilities that are primarily of the Levenberg-Marquardt algorithm is actually its best feature to solve the fitting problems [Kipli et al., 2012]. The Levenberg-Marquardt algorithm, which usually utilizes the trainlm rule, is applied to solve the nonlinear least squares problems and it is also used for solving different problems concerning curve-fitting [Hines et al., 1997].

The output layer, which a linear function, used the purelin transfer function determines a layer output from its net input [Faizollahzadeh Ardabili, 2014]. The hidden-layer neurons (HLN) in the second layer of the MLP network are determined by selecting different number for the HLN and the performance factor of the network was acquired in the training process (Table 3). The most difficult process when using the MLP is the HLN selection. Finally, the MLP structure with 14 neurons in the hidden layer was preferred as the best MLP network with the highest network performance in this study.

Prediction performance of neural network

For the evaluation of the proposed MLP network based on prediction performance, the correlation coefficient (R), an absolute fraction of variance (R2), the mean absolute percentage error (MAPE), mean square error (MSE), and root mean square error (RMSE) were applied and expressed below:

$$R = \frac{\sum_{i=1}^{n}(t_i - \bar{t})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^{n}(t_i - \bar{t})^2 \sum_{i=1}^{n}(o_i - \bar{o})^2}} \tag{4}$$

$$R^2 = \frac{(n\sum t_i o_i - \sum t_i \sum o_i)^2}{[n\sum {t_i}^2 - (\sum t_i)^2][n\sum {o_i}^2 - (\sum o_i)^2]} \tag{5}$$

$$MAPE = \frac{1}{n}\left[\frac{\sum_{i=1}^{n}|t_i - o_i|}{\sum_{i=1}^{n} t_i} x100\right] \tag{6}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(t_i - o_i)^2 \tag{7}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(t_i - o_i)^2} \tag{8}$$

where: t_i signifies the target value, n is the total number, O_i is the output value, $\bar{t}$ and $\bar{o}$ are the average values of each target and output, respectively.

RESULTS AND DISCUSSION

Physical-chemical properties of groundwater

Statistics of the groundwater quality parameters are presented in Table 4, for the study area. The pH of groundwater samples ranges from 6.6–8.7 in the study area, with the average value of 7.34. The average pH implies that the majority of

Table 4. Descriptive statistics of physicochemical parameters (n = 41)

Parameters	Minimum	Maximum	Mean	Std. Dev.
pH	6.6	8.7	7.34	0.373
TDS	1200.0	8990.0	4985.10	1759.475
Na^+	12.2	3997.0	917.85	949.232
Ca^{2+}	257.0	1002.0	592.72	179.791
K	0.0	267.2	57.87	69.112
Mg^{2+}	4.0	316.4	150.75	70.256
SO_4^{2-}	290.0	3408.0	1713.92	811.534
Cl^-	249.6	4651.0	1880.86	1051.327
NO_3^-	6.0	86.0	27.97	14.158
HCO_3^-	61.0	3100.0	871.12	918.107
EC	1720.0	13890.0	7869.56	2674.375

All values in ppm excepting the pH (–) and EC (µS/cm).

the samples met WHO (1996) standard of 6.5–8.5 except one sample (pH = 8.7) within the study location. The TDS for the groundwater samples in the range of 1200–8990 mg/L are above the admissible limit of 1000 mg/L. This indicates that the groundwater in the study region considerably suffers from anthropogenic sources. Spatial distribution maps of physical parameters (pH and TDS) were created for the study area (see Figure 3a and b).

In the study area, it is observed that sodium is the first dominant cation. The sodium concentration is varied from 12.2–3997.0 mg/L with an average value of 917.85 mg/L. The high availability of sodium concentration could be caused by water coming from the textile industries. Calcium is the second dominant cation in the study area, with a range of 257.0–1002.0 mg/L with an average value of 592.72 mg/L. The spatial distribution plots of Na^+ and Ca^{2+} for the study area are shown in Figure 4 a and b, respectively.

Magnesium of the groundwater samples varies from 4.0 mg/L to 316.4 mg/L. The drinking

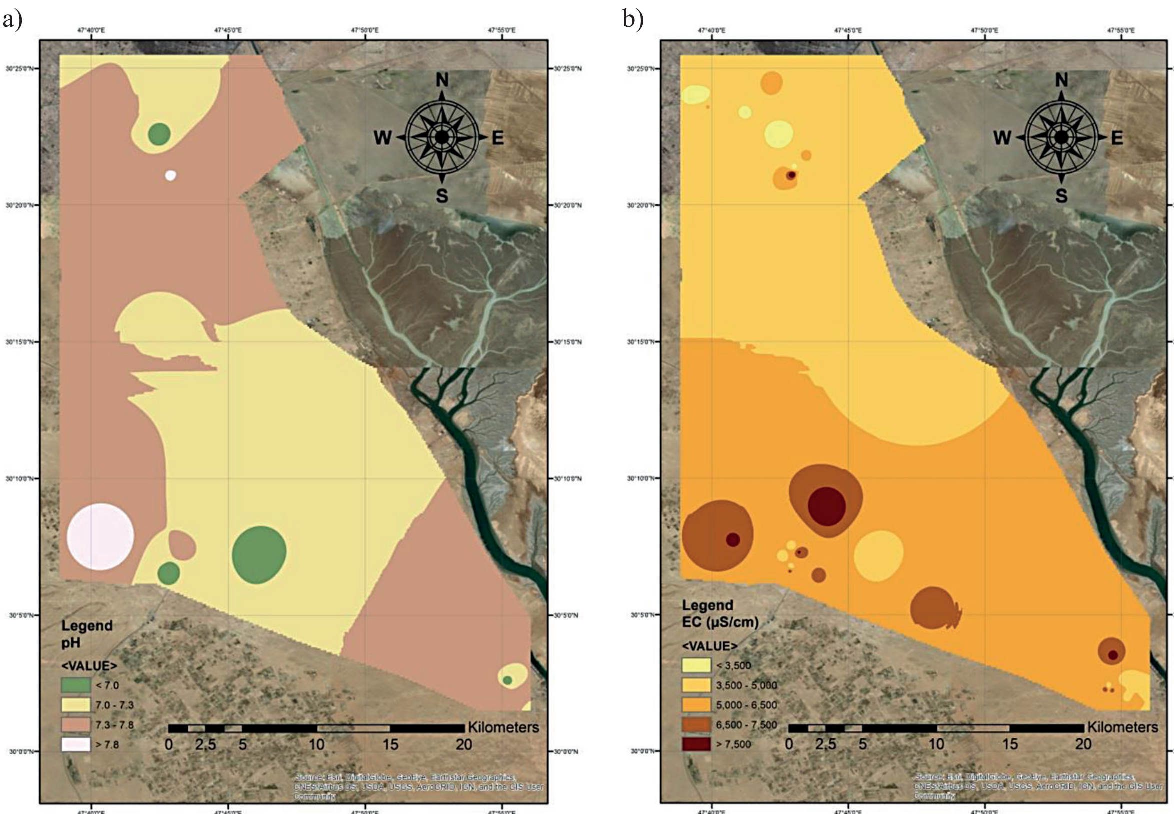

Figure 3. Spatial distribution maps of physical parameters (a) pH and (b) TDS

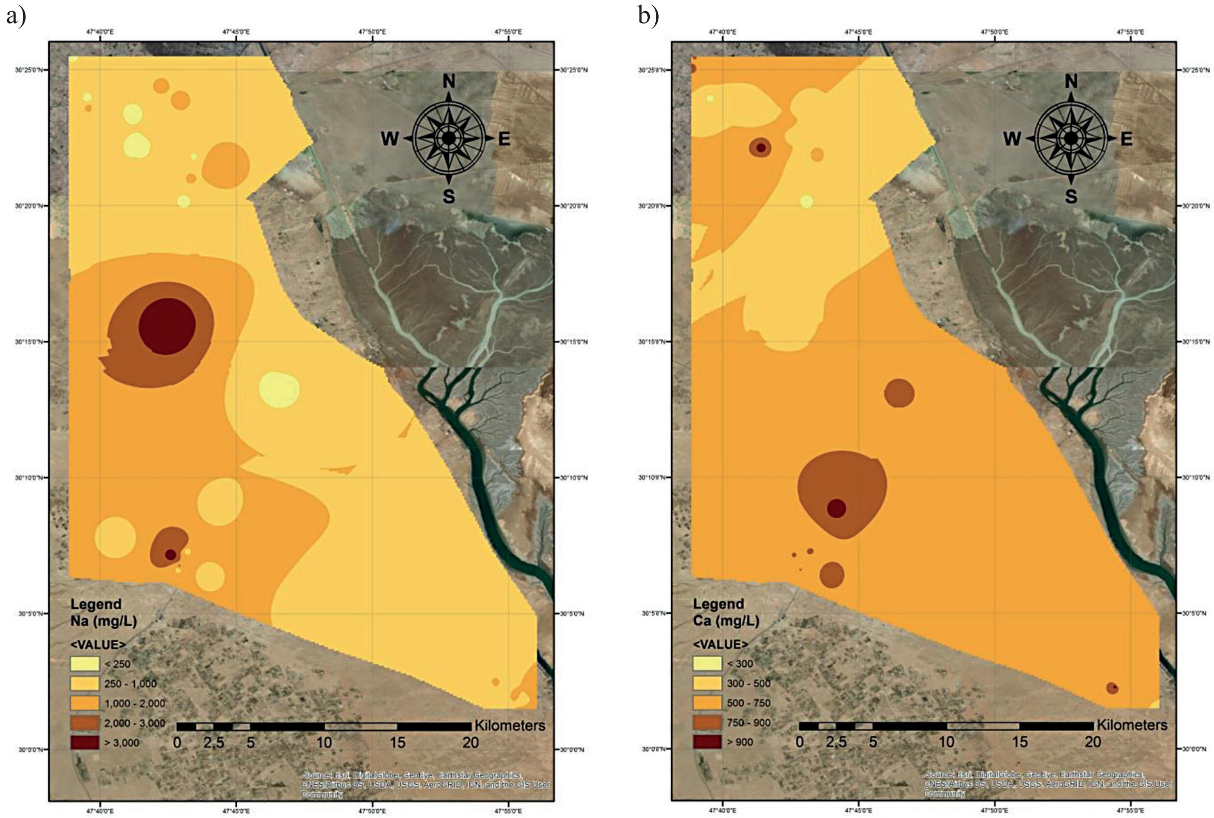

Figure 4. Spatial distribution maps of physical parameters (a) Na^{+} and (b) Ca^{2+}

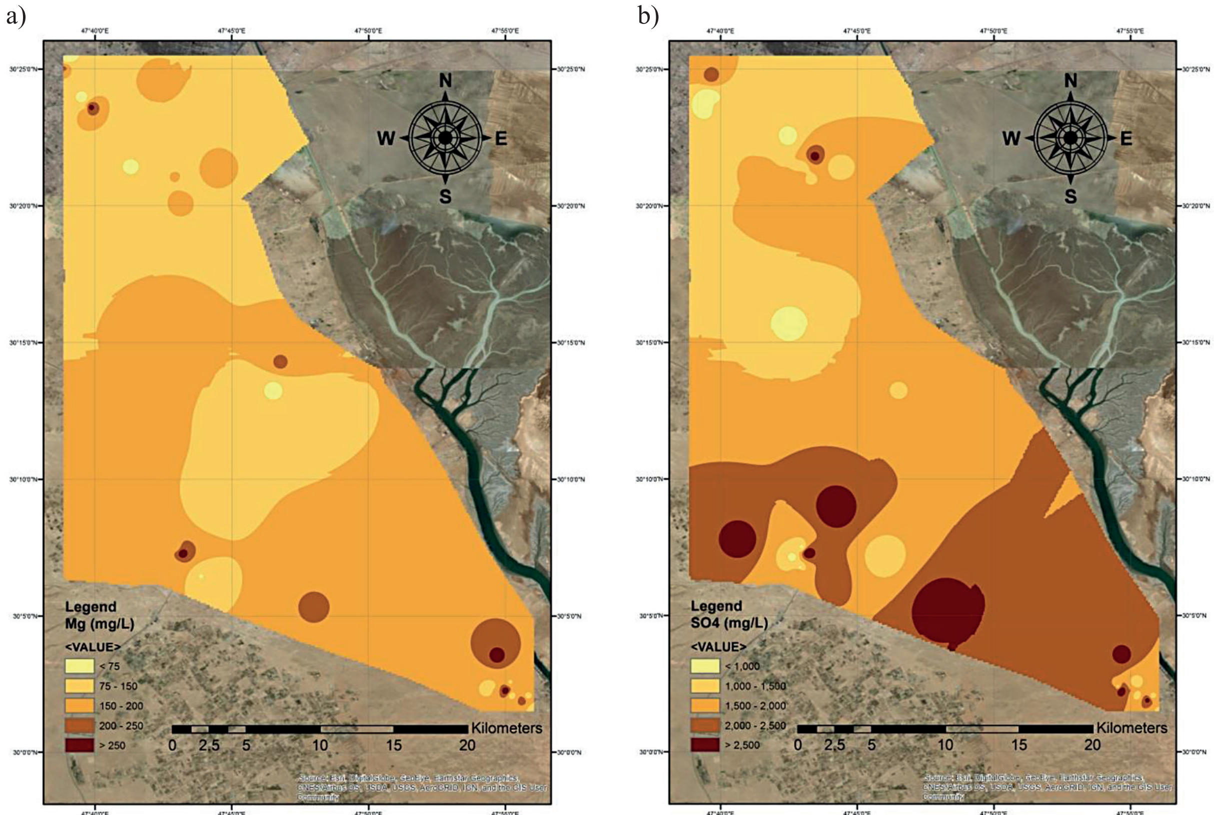

Figure 5. Spatial distribution maps of physical parameters (a) Mg^{2+} and (b) SO_4^{2-}

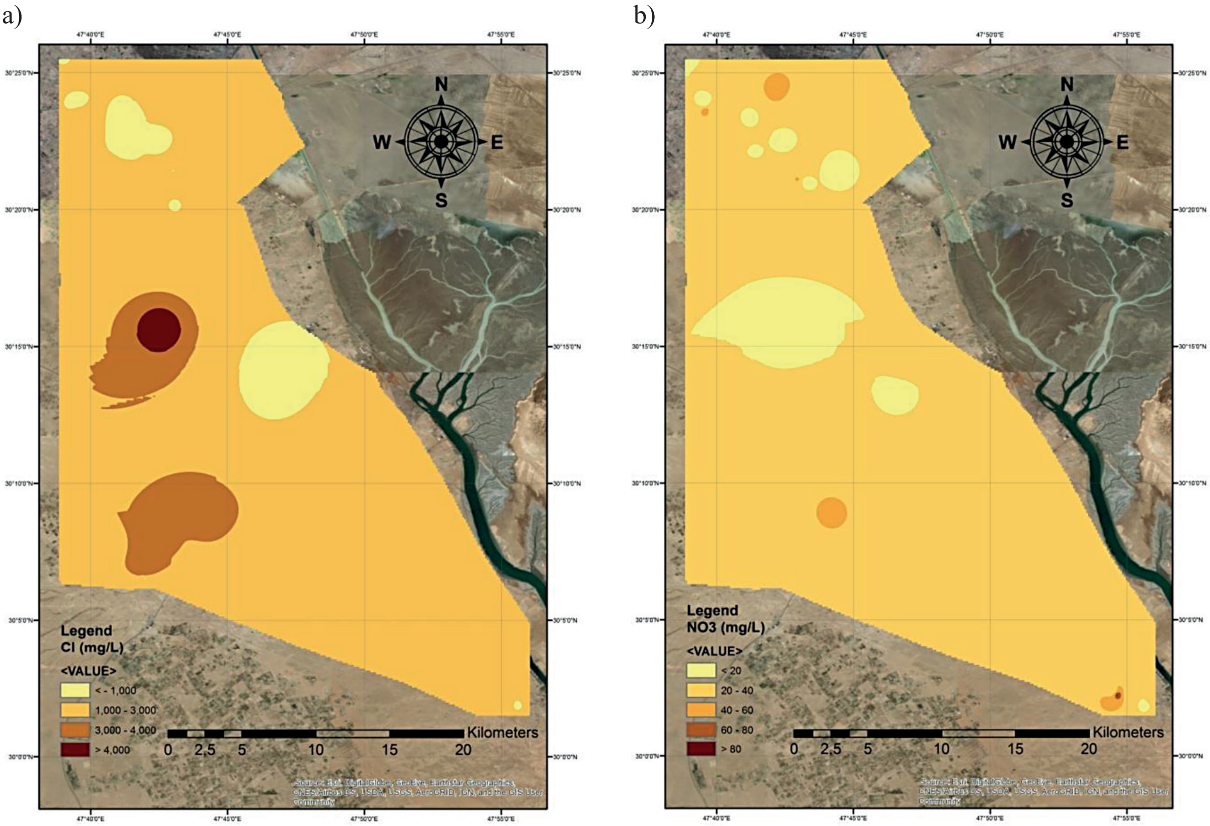

Figure 6. Spatial distribution maps of physical parameters (a) Cl^- and (b) NO_3^-

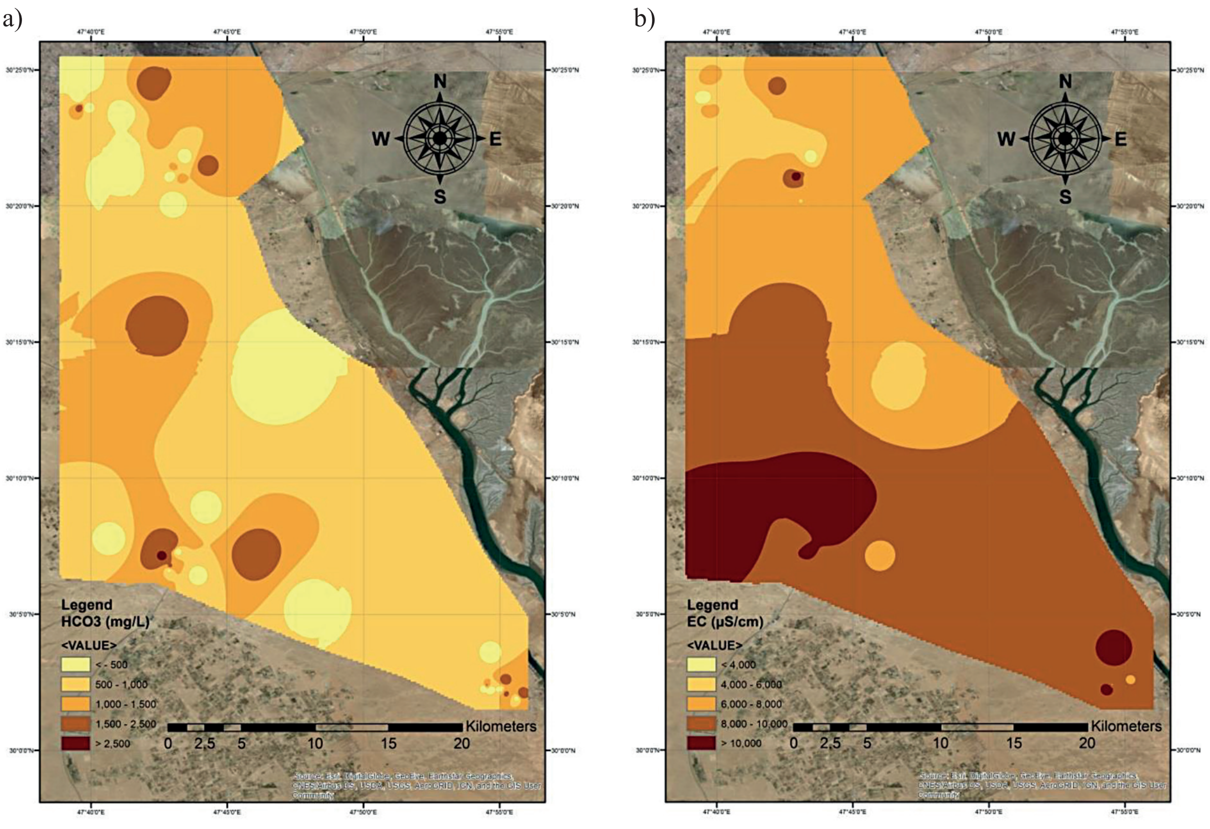

Figure 7. Spatial distribution maps of physical parameters (a) HCO_3^- and (b) EC

water standards recommended the allowable limit of magnesium is 50 mg/L [WHO, 1996]. In was noticed the study region, the average magnesium concentration which is equal to 150.75 mg/L, surpassed the permissible limit. However, sodium slightly dominates over magnesium. During the study period, the average potassium was 69.112 mg/L. Concerning the origin of those cations, potassium and sodium are the ions which are typically crystalline stones. Potassium mineral is less dissolvable within the water that is natural. It may be effortlessly attached by clay minerals. This is actually the significant reason behind the lack of this ion in groundwater. The sulfate concentration is high in all the locations, with the highest value equal to 3408.0 mg/L. High values of ions such as sulfate and calcium indicate a dissolution that is possible gypsum. Spatial distribution maps of cation and anion parameters (Mg^{2+} and SO_4^{2-}) were created for the study area (see Figure 5 a and b).

Chloride signifies the first dominant anion, in all of the locations. The concentration of chloride varies between 249.6–4651.0 mg/L with the average value of 1880.86 mg/L. Generally, the normal source of chloride is principally from halite dissolution. The primary origin of chloride may be also related to textile wastewater and the domestic sewage. Usually, chloride is highly dissolvable in water. The admissible limit of nitrate for drinking water is 45 mg/L [WHO, 2011]. Nonetheless, within the study region, some groundwater samples exceed the permissible limit. Its concentration is in the range of 6–86 mg/L. Nitrate is significant because of the influence of this ion on the human health. Nitrate is available in both surface water and groundwater, as a result of farming activity, wastewater disposal, septic tanks leaching, including oxidation of the nitrogenous waste material in animal and human excreta [WHO, 2011]. Nevertheless, the abnormal values of sodium, chloride and nitrate could be caused by anthropogenic chemistry. Spatial distribution maps of chloride and nitrate parameters were created for the study area (see Figure 6 a and b).

Bicarbonate may be the anion that dominant next to chloride and sulfate. The dissolution of carbonates is the origin for this ion (HCO_3^-). The mean concentration of EC is 7869.56 µS/cm. The concentration of EC in the groundwater samples varies from 1720.0 µS/cm to 13890.0 µS/cm. Spatial distribution maps of HCO_3^- and EC parameters were created for the study area (see Figure 7 a and b).

WQI and its spatial distribution

The computing of WQI presents the composite impact of specific groundwater parameters in the total quality of water that used for human consumption [Mitra, 1998]. For this objective, eight groundwater quality parameters, including pH, Na^+, Ca^{2+}, Mg^{2+}, SO_4^{2-}, NO_3^-, Cl^-, and HCO_3^- are considered. In line with the preceding method that is discussed, the WQI of groundwater for Zubair district was studied. This research targets the application of WQI for human usage. The calculations of WQI are available on the basis of the standards limits suggested by WHO (1996). The calculated WQI in the study location is shown in Figure 8. In the study area, WQI varies from 124 to 1477. In this region, WQI showed that only one sample indicated good quality among the 41 samples analyzed another sample presented very poor water quality among the total analyzed samples. On the basis of the calculated WQI, 2.5% samples are good for drinking, 2.5% samples are very poor for drinking; the rest of samples which have 95% of the samples were classified as unfit for drinking quality. The class of poor water quality is caused by the high values of WQI in sulfate, nitrate and chloride parameters. The unfit water quality is a result of the very high values of WQI and in all parameters. These high values of WQI reveal that the industrial activities in the study area tend to affect the groundwater quality.

Correlation coefficient analysis

The correlation data for water quality is shown in Table 5. There are strong positive correlations exhibited between TDS and each of EC, SO_4^{2-} and NO_3^-, whilst medium correlation is noticed between EC and each of NO_3^- and Cl-. This suggests that EC has a strong relationship with other parameters, and it offers a prevalent origin source. Na^+ was observed to correlate with each of HCO_3^- and Cl^- in strong and medium correlation, respectively. HCO_3^- exhibited medium negative correlation with SO_4^{2-}. The other considerable correlation is exhibited amongst the studied parameters is shown in Table 5.

Factor and cluster analyses

Factor analysis was performed for the data set (11 variables) to make a comparison of the compositional patterns amongst the examined samples (groundwater quality parameters) and to

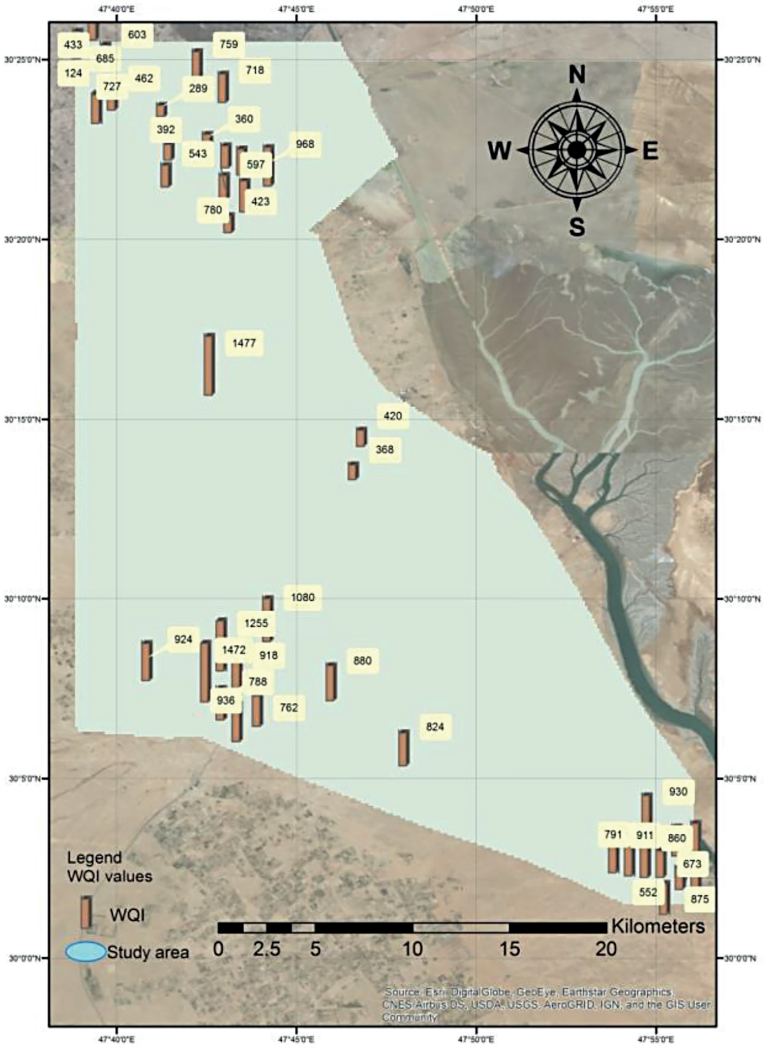

Figure 8. WQI values in the study area for the groundwater samples

Table 5. Matrix of correlations between several groundwater parameter

Parameter	pH	Ca^{2+}	Mg^{2+}	Na^{+}	K	HCO_3^-	SO_4^{2-}	Cl^-	TDS	EC	NO_3^-
pH	1.00										
Ca^{2+}	-0.02	1.00									
Mg^{2+}	0.20	0.00	1.00								
Na^{+}	-0.23	-0.13	0.16	1.00							
K	0.20	0.44	0.15	-0.12	1.00						
HCO_3^-	-0.40	-0.23	0.01	0.80	-0.25	1.00					
SO_4^{2-}	0.42	0.48	0.30	-0.43	0.23	-0.64	1.00				
Cl^-	0.01	0.30	0.25	0.67	0.06	0.33	0.12	1.00			
TDS	0.47	0.46	0.42	-0.12	0.25	-0.38	0.78	0.46	1.00		
EC	0.32	0.35	0.43	0.26	0.21	-0.38	0.48	0.68	0.83	1.00	
NO_3^-	0.27	0.30	0.25	-0.06	0.03	-0.07	0.49	0.26	0.83	0.54	1.00

determine the variation sources. From the analysis, factor analysis produced three factors with eigenvalue more than one, explaining 72.1% of the total variance in the dataset. The eigenvalues, the accounted percentage of variance as well as the cumulative variance percentage for the three determined factors are given in Table 6. The results from the factor analysis are given in Table 7 of groundwater samples, showed the parameter loadings. This table shows that the majority of the variables related to each factor was defined, making certain contributions with other factors. The three resulted factors could be caused by the three different and considerable sources of anthropogenic activities or contaminating units.

The first factor explained 37.1% of the total variance, and it is positively correlated with SO_4^{2-}, TDS and EC. This factor could be recognized as major industrial pollution. The contaminants released are frequently high in toxic chemical compounds, with a high level of organic waste concentration. The second factor explained 24.0% of the total variance, and it is positively correlated with Na^+, HCO^3, and Cl^- positively.This factor may be known as the factor in charge of domestic sewage contamination. The third factor accounted for 11.1% of the total variance, and it is positively correlated with Ca^{2+} and K positively.

Table 6. Extracted values of different analysis for groundwater samples

Component	Sum's extracted from squared loadings		
	Eigenvalue	% of variance	Cumulative %
1	4.0758	37.1	37.1
2	2.6422	24.0	61.1
3	1.2178	11.1	72.1

Table 7. Factor analysis results of groundwater samples in the study locations

Parameter	Factors		
	F1	F2	F3
pH	0.257	-0.138	-0.396
Ca^{2+}	0.286	-0.035	0.633
Mg^{2+}	0.227	0.167	-0.329
Na^+	-0.083	0.579	0.050
K	0.191	-0.098	0.514
HCO_3^-	-0.212	0.500	0.002
SO_4^{2-}	0.415	-0.204	-0.026
Cl^-	0.221	0.475	0.144
TDS	0.473	0.046	-0.080
EC	0.397	0.285	-0.043
NO_3^-	0.331	0.077	-0.193

Cluster analysis is applied to discover the similarity groups amongst the well locations. Cluster analysis delivered a dendrogram, as demonstrated in Figure 9. It could be observed that cluster analysis is unsuccessful in revealing distinct groups and no clear-cut class structure is evident, because of some locations in the study area are clustered under similar group.

Application of MLP models in prediction

In this research, three-layer network architecture, containing only one single hidden layer, was applied to prepare the MLP network. The HLN in the hidden layer, momentum coefficient, and the learning rate were determined by the trial and error process. The training procedure in the MLP is the most significant important step for generating the MLP network. The network training started with different neurons (HLN) in the hidden layer and in the next phase of training, another HLN were added to the number of neurons. The HLN in the hidden layer was tried in the range of 4 to 16. The MSE value was reported as an efficiency

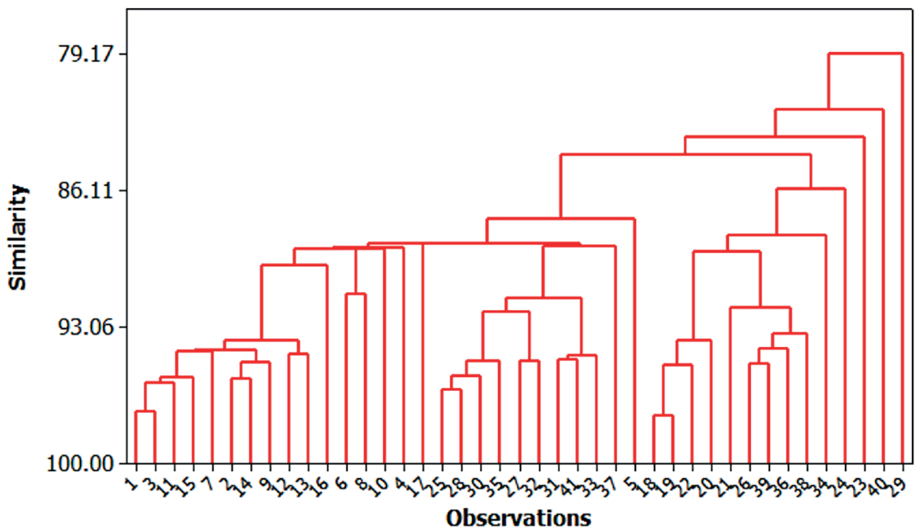

Figure 9. Dendrogram displaying clustering of wells locations relating to groundwater quality parameters

factor for making the decision about which model is better than other, as seen in Table 4.

As reported in Table 3, the efficiency factor of order 6 with MSE value 0.027414 has the minimum value, whereas the efficiency factor of order seven with the value of 0.222014 has the most worth value; depending on the description, MSE is a factor of difference between predicted and what exactly is predicted [Ardabili et al., 2016], consequently a small value of MSE implies the accuracy of the forecasted network. Here, the order 6 of the HLN was selected to train the network.

As the activation function, the tangent sigmoid and logarithmic sigmoid functions were applied on the hidden and output layers, respectively. The learning rate (lr) plus the momentum coefficient (mc) were examined in the range of 0.1 to 0.9. The number of iterations in the established MLP models was selected as 500.

Figure 8 shows the linearity outputs of the MLP network with the target values. The linearity is represented by the coefficient of determination (R^2). After application of the MLP in this study, the results of output values have 99.98% of linearity with the target values. The overall results of applying the MLP network are shown in Table 8.

In an overview of Table 8, relating to testing data from trials, probably the most effective MLP structure as a result of the largest R^2 value was MLP6 (8, 14, 1) model, in which all the chemical parameters were utilized. Throughout this model, 8 symbolizes the inputs consisting of Na^+, Ca^{2+}, K, Mg^{2+}, SO_4^{2-}, NO_3^-, Cl^-, and HCO_3^-; 14 symbolizes the number of neurons in the hidden layer, and 1 symbolizes the output. As reported by MLP6 network, the learning rate (lr) and momentum coefficient (mc) for the adopted model are 0.2 and 0.2, respectively.

On the basis of the results of Fig.10 and evaluating the coefficient values of the proposed model, it may be stated that the MLP network (with r = 0.9999) has a minimum linearity between target and output values, because this figure is plotted depending on the total data series to contrast the overall forecast proficiency of the

Table 8. The model structure, R^2, Ir, mc, and MAPE of the developed models

Order	R2	R2	Ir	mc	MAPE
1	MLP1(8,4,1)	0.9996 004	0.3	0.2	0.040625
2	MLP2(8,6,1)	0.9996 004	0.2	0.7	0.027121
3	MLP3(8,8,1)	0.9996 004	0.1	0.3	0.034722
4	MLP4(8,10,1)	0.9962	0.7	0.5	0.070284
5	MLP5(8,12,1)	0.9990	0.5	0.2	0.107946
6	MLP6(8,14,1)	0.9998	0.2	0.2	0.030983
7	MLP7(8,16,1)	0.9799	0.7	0.6	0.239764

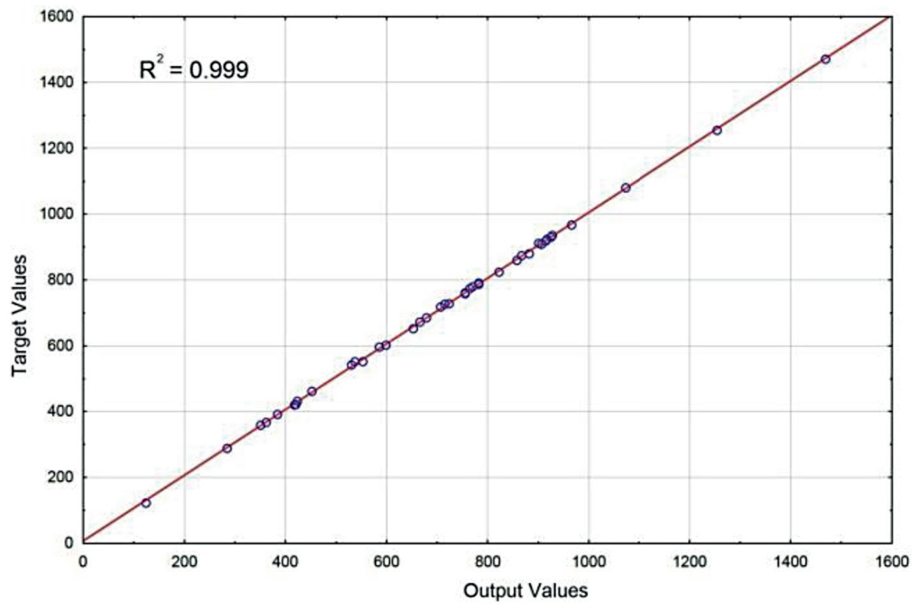

Figure 10. The results of the MLP6 network

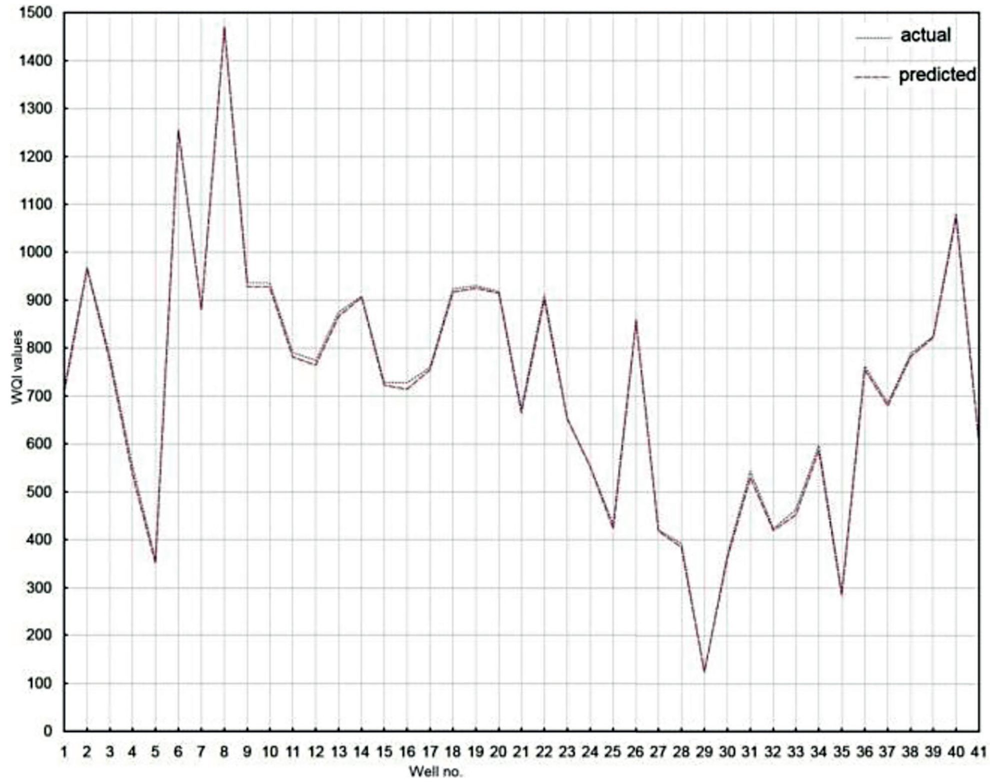

Figure 11. The actual and predicted WQI values using the MLP6 network

models. Figure 11 shows the actual and predicted WQI values using the MLP6 network.

CONCLUSIONS

The assessment and estimation of groundwater quality in the Basrah province were identified using 41 bore wells. Water quality assessment of groundwater samples in the Basrah province revealed that the groundwater is deteriorated in most studied locations of the study area. Average concentrations of EC and TDS were above the admissible limit in all groundwater samples. The concentration of Na over Ca is high that denotes the occurrence of a cation exchange. Following analysis of various physicochemical parameters, it was observed that the WQI ranged from 123 to 1477. WQI revealed that a percentage of 2.5%, 2.5% and 95% from the groundwater samples were classified as poor, very poor, unsuitable for drinking classes. The analysis of physiochemical water quality parameters of groundwater samples agrees with the same result of WQI. For these reasons, water from most wells is not used for human consumption.

From the results, it can be stated that the application of multivariate statistical methods has confirmed as effective and efficient tools to clarify the total variance in the data under consideration. Also, FA recognized three factors trusted for clarifying more than 72% of the total variance in the studied data. Therefore, the multivariate statistical methods, specifically CA and FA, can simply categorize the available source of pollution.

The results of applying the MLP model with 14 HLN in the hidden layer have a small value of MSE (0.027414) is a determining element of efficiency during the training period. Finally, by this study, the MLP model with secure availability an efficient tool in the forecast process; however, when it comes to requiring high precision, intelligent techniques must certainly be utilized.

Acknowledgements

We thank anonymous reviewers for the valuable suggestions, which helped us improve the quality of our paper.

REFERENCES

1. Abbas A.H.A., Dawood A.S., Al-Hasan Z.M. 2017. Evaluation of groundwater quality for drinking purpose in basrah governorate by using application of water quality index. Kufa Journal of Engineering, 8(1).

2. Aller L., Bennett T., Lehr J., Petty R.J., Hackett G. 1987. DRASTIC: A standardized system for evaluating ground water pollution potential using hydrogeologic settings. US Environmental Protection Agency. Washington, DC, 455.
3. Alvin C.R. 2002. Methods of multivariate analysis. Wiley INC, USA.
4. American Public Health Association, American Water Works Association, Water Pollution Control Federation and Water Environment Federation. 1912. Standard methods for the examination of water and wastewater. American Public Health Association, 2.
5. Ardabili S.F. 2014. Simulation and comparison of control system in mushroom growing rooms environment (Doctoral dissertation, University of Tabriz).
6. Ardabili S.F., Mahmoudi A., Gundoshmian T.M., Roshanianfard A. 2016. Modeling and comparison of fuzzy and on/off controller in a mushroom growing hall. Measurement, 90, 127–134.
7. Ardabili S.F., Najafi B., Ghaebi H., Shamshirband, S., Mostafaeipour A. 2017. A novel enhanced exergy method in analyzing HVAC system using soft computing approaches: A case study on mushroom growing hall. Journal of Building Engineering, 13, 309–318.
8. Ayers R.S., Westcot, D.W. 1985. Water quality for agriculture. Rome: Food and Agriculture Organization of the United Nations, 29, 174.
9. Boyacioglu H. 2010. Utilization of the water quality index method as a classification tool. Environmental Monitoring and Assessment, 167(1), 115–124.
10. Chauhan A., Pawar, M., Lone S.A. 2010. Water quality status of Golden Key lake in clement town, Dehradun, Uttarakhand. J. Am. Sci, 6(11), 459–464.
11. Dawood A.S., Ahmed A.N. 2016. Using GIS for Assess the Groundwater Quality in Southwest Side of Basrah City. Muthanna Journal of Engineering and Technology (MJET), 4(2), 75–87.
12. Hines J.W., Tsoukalas L.H., Uhrig R.E. 1997. MATLAB supplement to fuzzy and neural approaches in engineering. John Wiley & Sons, Inc.
13. Horton R.K. 1965. An index number system for rating water quality. J Water Pollut Control Fed, 37(3), 300–306.
14. Johnson R.A., Wichern D.W. 2002. Applied multivariate statistical analysis. Prentice-Hall, London.
15. Kakati S.S., Sarma H.P. 2007. Studies on water quality index of drinking water of Lakhimpur District. Indian Journal of Environmental Protection, 27(5), 425.
16. Kipli K., Muhammad M.S., Masra S.M.W., Zamhari N., Lias, K., Mat, D.A.A. 2012. March. Performance of Levenberg-Marquardt backpropagation for full reference hybrid image quality metrics. In Proceedings of the International MultiConference of Engineers and Computer Scientists, 1, 14–17.
17. Mahmood A.A., Eassa A.M., Muayad H.M., Israa Y.S. 2013. Assessment of ground water quality at Basrah, Iraq by water quality index (WQI). Journal of University of Babylon, 21(7), 2531–2543.
18. Milovanovic M. 2007. Water quality assessment and determination of pollution sources along the Axios/Vardar River, Southeastern Europe. Desalination, 213(1–3), 159–173.
19. Mitra B.K., Sasaki C., Keijirou E. 2006. Spatial and temporal variation of ground water quality in sand dune area of aomori prefecture in Japan. In 2006 ASAE Annual Meeting. American Society of Agricultural and Biological Engineers, 1.
20. Pawar R.S., Panaskar D.B., Wagh V.M. 2014. Characterization of groundwater using water quality index of solapur industrial belt, Maharashtra, India. Int J Res Eng Technol, 2(4), 31–36.
21. Ramakrishnaiah C.R., Sadashivaiah C., Ranganna G. 2009. Assessment of water quality index for the groundwater in Tumkur Taluk, Karnataka State, India. E-Journal of Chemistry, 6(2), 523–530.
22. Ramesh K., Elango L. 2012. Groundwater quality and its suitability for domestic and agricultural use in Tondiar river basin, Tamil Nadu, India. Environmental Monitoring and Assessment, 184(6), 3887–3899.
23. Reghunath R., Murthy T.S., Raghavan B.R. 2002. The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. Water Research, 36(10), 2437–2442.
24. Reza R., Singh G. 2010. Assessment of ground water quality status by using Water Quality Index method in Orissa, India. World Appl Sci J, 9(12), 1392–1397.
25. Rojas R. 2013. Neural networks: a systematic introduction. Springer Science & Business Media.
26. Sahu B.K., Panda R.B., Sinha B.K., Nayak, A. 1991. Water quality index of the River Brahmani at Rourkela industrial complex of Orissa. J. Ecotoxicol. Environ. Monit., 1(3), 169–175.
27. Shamim M.A., Ghumman A.R., Ghani U. 2004. December. Forecasting groundwater contamination using artificial neural networks. In 1st International Conference on Water Resources and Arid Environments, 5–8.
28. Simeonov V., Simeonova P., Tzimou-Tsitouridou R. 2004. Chemometric quelity assessment of surface waters: two case studies. Chemia i Inżynieria Ekologiczna, 11(6), 449–469. (in Polish)
29. Simsek C., Gunduz O. 2007. IWQ index: a GIS-integrated technique to assess irrigation water quality. Environmental Monitoring and Assessment, 128(1), 277–300.
30. Sinha A.K., Saxena A.S., Saxene R. 2004. Water quality index for Ram Ganga river water at Moradabad. Pollut. Res, 23, 527–531.

31. Srinivasamoorthy K., Chidambaram S., Prasanna M.V., Vasanthavihar M., Peter J., Anandhan P. 2008. Identification of major sources controlling groundwater chemistry from a hard rock terrain – a case study from Mettur taluk, Salem district, Tamil Nadu, India. Journal of Earth System Science, 117(1), 49–58.
32. Tyagi S., Sharma B., Singh P., Dobhal R., 2013. Water quality assessment in terms of water quality index. American Journal of Water Resources, 1(3), 34–38.
33. Vandendriessche S., Vincx M., Degraer S., 2006. Floating seaweed in the neustonic environment: a case study from Belgian coastal waters. Journal of Sea Research, 55(2), 103–112.
34. Vasanthavigar M., Srinivasamoorthy K., Vijayaragavan K., Rajiv Ganthi R., Chidambaram S., Anandhan P., Manivannan R., Vasudevan S. 2010. Application of water quality index for groundwater quality assessment: Thirumanimuttar sub-basin, Tamilnadu, India. Environmental Monitoring and Assessment, 171(1), 595–609.
35. 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.
36. WHO. 1996. Guidelines to Drinking Water Quality [Z]. World Health Organisation, Geneva. 2, 989.
37. WHO. 2011. Guidelines for Drinking-Water Quality Fourth Edition [Z]. World Health Organization, Geneva, Switzerland.
38. Wilcox L. 1955. Classification and use of irrigation waters (No. 969). US Department of Agriculture.