Desalination and Water Treatment
www.deswater.com
doi: 10.5004/dwt.2022.28335

255 (2022) 145–156
April

Groundwater quality modeling using geostatistical methods and artificial neural networks: a case study of the Western Middle Cheliff alluvial plain in Algeria

Yamina Elmeddahi[a,b,*], Djamel-Eddine Moudjeber[c], Hacene Mahmoudi[c], Mattheus F.A. Goosen[d]

[a]*Department of Hydraulic, Civil Engineering and Architecture Faculty, University of Hassiba Benbouali, Chlef, Algeria, email: y.elmeddahi@univ-chlef.dz*
[b]*Vegetal Chemistry – Water-Energy Laboratory (LCV2E)*
[c]*Faculty of Technology, University of Hassiba Benbouali, Chlef, Algeria, emails: d.moudjeber@univ-chlef.dz (D.-E. Moudjeber), h.mahmoudi@univ-chlef.dz (H. Mahmoudi)*
[d]*Office of Research & Innovation, Alfaisal University, Riyadh, Saudi Arabia, email: mgoosen@alfaisal.edu*

Received 15 September 2021; Accepted 7 November 2021

ABSTRACT

The aim of the current research was to determine the most appropriate model for estimating potable groundwater in a geographical area based on the water quality index (WQI) using the Western Middle Cheliff alluvial plain in Algeria as a case study. The spatial distribution of the WQI in a graphical display was determined using geostatistical ordinary kriging (OK), and artificial neural networks (ANN) which were integrated with a geographical information system. Results indicated that the ANN model with its high correlation coefficient (*R*) and low error rate had a greater accuracy than OK in estimating the WQI. Based on the WQI classification index, 60% of water samples were found to be poor and 34% in the excellent and good categories. Among the kriging models, the Gaussian version was specified as the best for determining the WQI. The findings indicated that the ANN model with 11 hidden layers had the greatest accuracy. The root mean square error, the mean absolute error, and *R* values were 2.347, 0.71, and 0.9998, respectively compared to the kriging model. The geostatistical model, with its relatively lower precision, was limited by the number of samples. It was difficult to discover the relationship between the spatial location of sampling and the variable. Whereas intelligent models such as ANN, were more capable of obtaining this connection. The significance of this analysis shows that for semi-arid regions, modeling using ANN is an important tool for effective groundwater quality management.

Keywords: Middle Cheliff alluvial plain; Groundwater quality: Geostatistics; Artificial neural network; Prediction of water quality index; Kriging model

1. Introduction

Groundwater quality is usually affected by natural processes such as climate, geology, soil type, rock interactions, precipitation, and evaporation as well as anthropogenic or man-made factors such as agricultural practices including fertilizer application, pesticide spraying, and irrigation, as well as industrial growth and urbanization [1]. Hence, studying and assessing groundwater quality in specific areas can help in the improved management of water resources. Recently, much research has been conducted on understanding groundwater pollution and in trying to identify the

* Corresponding author.

Presented at the International Virtual Conference on Water and Energy (ICWE 21), Hassiba Benbouali University of Chlef, Algeria, 25–27 May 2021

sources. Sargazi et al. [1], for example, employed geographical information systems (GIS) for assessment of groundwater quality. Their results showed that water quality declined in the case study area of southeast Iran. The authors reported that the reasons for the decline in groundwater conditions included the lack of rainfall, incomplete feeding of the aquifer and the geology of the surrounding area.

In arid and semi-arid regions groundwater plays an important role in satisfying domestic, agricultural irrigation, and industrial water needs [2,3]. Groundwater is a vital resource. However, it is characterized by high variability depending on climate and location. Most regions of the North African country of Algeria, for example, are facing groundwater depletion due to the combined effects of climate change, population growth and increases in irrigation and urban needs. The decrease in rainfall over the past decades has also exerted further pressure on groundwater resources [4,5]. In addition, the escalating demand has led to a deterioration. Therefore, it is crucial to be able to monitor and to better understand any changes in groundwater quality.

The combination of GIS and geostatistical analysis has become an important tool for groundwater studies. Recently, geostatistical techniques have been widely used in hydrochemical analyzes, particularly, in predictions at unsampled locations, and to describe the groundwater level [6–8]. Currently, the developed world is using sophisticated tools such as mathematical models for predicting water quality to overcome the drawback of outdated data processing methodologies [1]. The kriging interpolation approach (i.e., Gaussian process regression) for instance is considered the most efficient method for estimating an unknown point by taking advantage of known points [9]. Kriging is one of several methods that use a limited set of sampled data points to estimate the value of a variable over a continuous spatial field. Geostatistics has been widely utilized to investigate the spatial distribution of groundwater quality data [10,11]. This approach based on the modeling and estimating of the spatial dependence structure, is usually described using statistical tools like the variogram or the covariance, calculated on the whole field of interest and under a stationarity hypothesis.

To know the groundwater quality in a specific region, the use of a model becomes essential because surveillance is expensive and time-consuming, and in many cases, samples cannot be collected from certain points. A mathematical model such as a water quality index, for example, is an efficient indicator, that can be widely applied to determine, classify, and manage overall groundwater quality as a single parameter [6–8]. The water quality index (WQI) is a simple, powerful, and widely used technique to evaluate overall water quality, especially groundwater [12,13]. Various WQIs have been formulated worldwide, which can reflect the overall effect of several parameters on water quality [12–15]. The WQI calculation is performed using a weighted arithmetic indicator. It is based on selected water parameters, converted into a single unitless value.

The artificial neural network (ANN) is an intelligent modeling technique that can help to solve many prediction and classification problems related to groundwater quality [16,17]. The functioning principle is like that of the human brain. ANN can learn generalization and prediction skills which can be utilized for non-linear and complex data [16]. It is an excellent tool for visualizing groundwater quality details and for predicting the sources of pollution [16,17]. The artificial neural network (ANN) can be considered as a nonlinear model of data processing which has similarities to the problem-solving process of the human brain [18]. The ability to simulate complex functional relationships with high precision makes it in high demand. The multilayer perceptron (MLP) is the most utilized model in feed-forward ANN due to its simplicity. It is a flexible method consisting of three layers: input, hidden, and output. The input layer consists of neurons which receive input values from the network. This data is then transmitted to hidden neurons. The concealed layers in which each neuron receives information from several previous layers, performs the weight-weighted summation, and then transforms it according to its activation function which is generally a sigmoid function, and the output layer which plays the same role as the hidden layers, the only difference between these two types of layers is that the output of the neurons in the output layer is not linked to any other neuron. Furthermore, in many applications, there may be more than one hidden layer in the network, and the number of hidden layers and the number of neurons in the layer change according to the problem. Using fewer neurons in this layer may result in less sensitive output data. Similarly, if more neurons are used than necessary, problems may arise in treating the new types of data groups in the same network. Therefore, it is generally suggested to use one hidden layer within the network. The model performance function should be based on error measurements [19].

Several researchers assessed the quality of groundwater in Algeria [20–24]. Studies have employed, the water quality index [15,25–27], and have adopted the geostatistical approach [28–30]. However, there is a lack of reports dedicated to groundwater quality space-temporal variation in the Middle Cheliff Basin. Such assessments are essential for effective development of future management plans and the sustainable growth of groundwater resources. In the absence of direct measurements, it would be helpful to be able to use a predictive model to estimate the quality of the groundwater.

The aim of the current research was to determine the most appropriate model for estimating potable groundwater in a geographical area based on the WQI using the Western Middle Cheliff alluvial plain in Algeria as a case study. The best method for estimating the spatial distribution of the WQI in a graphical display was determined using Gaussian process regression (i.e., kriging), and artificial neural networks (ANN) which were integrated with a geographical information system (GIS). The physicochemical water quality variables such as electrical conductivity (EC), pH, total dissolved solids (TDS), Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, Cl^{-}, SO_4^{2-}, NO_3^{-}, and HCO_3^{-} were assessed as part of this process. The results of ANN and geostatistical models were compared.

2. Methodology

2.1. Case study area characteristics

The case study area was in north western Algeria, about 200 km west of Algiers. It covered 270 km^2 in the Western Middle Cheliff basin which is itself formed by

three sub-watersheds (Fig. 1). This basin belongs to the large Cheliff-Zahrez (CZ) hydrographic basin, which covers approximately 56,227 km^2 (more than 22% of the area of northern Algeria) and is located between 36°01′ and 36°17′ north latitude and 0°58′ and 01°27′ east longitude. The region's climate is semi-arid Mediterranean with very irregular rainfall and temperature. The Plio-Quaternary formations forms a major part of the Cheliff valley and is constituted mainly of stressed water aquifers. The lithological variation and the existence of fractures make these formations a multilayer aquifer with thick clay intercalations. The substratum is formed by the blue marls of the Upper Miocene and the Pliocene, which outcrop on the northern and southern borders of the plain. Generally, the study site is agricultural with pedology heterogeneity, the nature and distribution of the soils would promote basic exchanges and enrichment of the waters with calcium and magnesium.

In the current analysis, the physicochemical groundwater parameters such as EC, TDS, sodium (Na), calcium (Ca), magnesium (Mg), potassium (K), chlorine (Cl), sulfate (SO_4), bicarbonate (HCO_3), nitrate (NO_3) and the pH were used to determine the drinking water quality. The data employed were collected in 2014 from 54 observation wells in the Western Middle Cheliff plain. The groundwater quality was assessed for its suitability for drinking purposes by comparing the physical and chemical parameters of different samples in the study area with the drinking water standards recommended by the World Health Organization [31].

2.1. Determination of the groundwater quality index

The first step in the computation of the WQI, consists of determining the weight value (w_i) for each parameter depending on their relative influence on the overall quality of drinking water. The highest weight of 5 was given to EC, TDS and nitrate. Bicarbonate was assigned a minimum weight of 1 because it has a less significant role in water quality evaluation [12,32]. The other parameters were given a weight between 4 and 2 according to their influence on the assessment of the quality of drinking water.

In the second step, the relative weight (W_i) of each parameter is calculated through the following equation:

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

where W_i is the relative weight, w_i is the parameter weight and n is number of parameters.

The third step involves the calculation of the quality rating scale (q_i) for each parameter using the equation:

$$q_i = \frac{C_i}{S_i} \times 100 \tag{2}$$

where (q_i) is the quality ranking, (C_i) is the concentration of each chemical parameter in the analyzed sample in milligrams per liter, and (S_i) is the admissible limit of each parameter in drinking water according to World Health Organization (WHO) standard in milligrams per liter.

The sub-index of each parameter SI_i is determined in the fourth step, using the equation:

$$SI_i = W_i \times q_i \tag{3}$$

Finally, the WQI is calculated by adding up the SI_i values of all the parameters as follows:

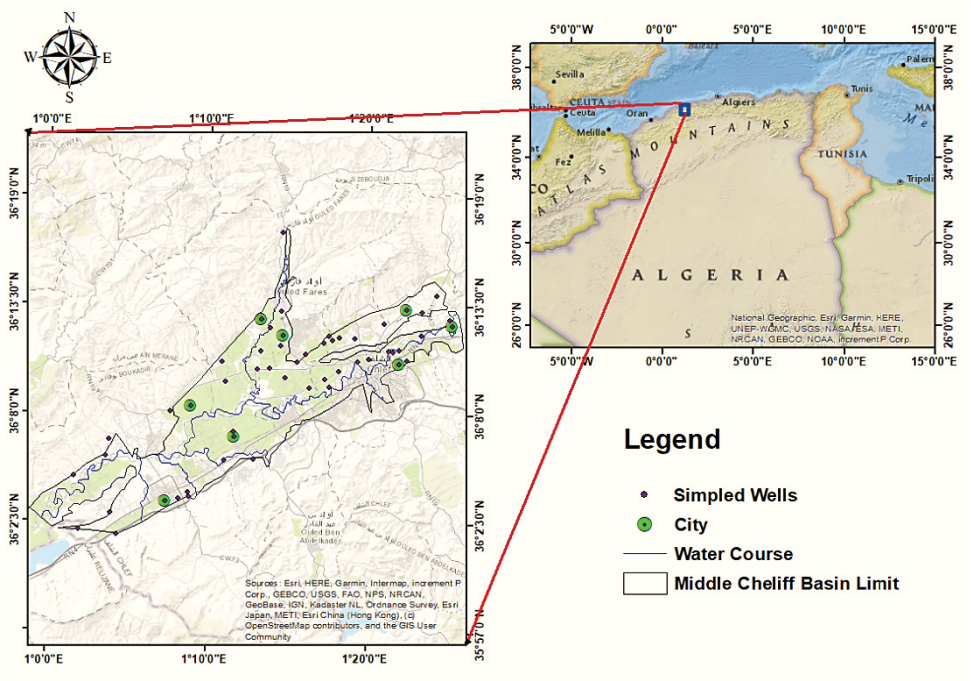

Fig. 1. Situation map and sampling well locations of the case study area.

$$WQI = \sum_{i=1}^{n} SI_i \quad (4)$$

Computed WQI values are generally classified into five groups. The WQI range and categories of water have been classified and are represented in Table 1.

2.2. Application of geostatistics in assessing the spatial distribution of groundwater quality data

Ordinary kriging and semi-variogram models were applied to determine the spatial distribution of the groundwater WQI in the Western Middle Cheliff plain. In geostatistics, the variogram is the most substantial tool to identify the spatial structure as well as the three-dimensional correlation between data. The best-fitted semi-variogram models were selected based on the mean absolute error (MAE) and root mean square error (RMSE) values. The efficiency of the model is the most accurate when MAE and RMSE are minimum. A cross-validation technique was used to determine the correlation between the measured and estimated values. The aim of a correlation coefficient (R) is to evaluate how accurately the quality parameters are estimated in the non-sampled sites of the study area.

Ordinary kriging (OK) was applied to interpolate predictive maps of groundwater quality index for unsampled locations. This method uses the semi-variogram to describe the spatial continuity or auto-correlation. The semi-variogram γ (h) determines the strength of the statistical correlation as a function of distance. It is defined as half the mean quadratic difference between two observations of a variable separated by a distance vector h [11]. It is defined as:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[z(\mu_i + h) - z(\mu_i) \right]^2 \quad (5)$$

where $\gamma(h)$ is the variogram for distance h; $N(h)$ denote the number of data pairs for that lag h, and $z(\mu_i)$ and $z(\mu_i+h)$ are the values of the regionalized variable of interest at location μ_i and μ_i+h, respectively.

Table 1
Groundwater classification based on the quality water index (Sahu and Sikdar [14])

Class	WQI value	Type of water quality
1	<50	Excellent
2	50–100	Good
3	100.1–200	Poor
4	200.1–300	Very poor
5	>300	Unsuitable for drinking

2.3. Evaluating water quality index using artificial neural networks

The ANN model was applied to predict the groundwater quality index (WQI) in 54 samples from the study area. Fig. 2 shows the structure of an artificial neuron model. Multi-layer perceptron (MLP) network models are the common network architectures utilized in most research applications. In MLP, the weighted sum of the inputs and the term bias to the activation level are passed through the transfer producing output and the units are arranged in a layered feed-forward neural network [33]. The Levenberg–Marquardt algorithm was used to train the data and to determine the performance of neural network. Several architectures of the ANN were tested with the error measures for validation. The number of neurons in the input and output layers is determined by the nature of the problem under study. The number of neurons in the hidden layer was determined by trial and error to reduce the whole error between the observed and estimated values by tuning weights, and these weights were combined and processed through an activation function and edited into the output layer. The process initiated with a few neurons, and additional neurons continued to be added until the increase in neurons had no effect on error correction.

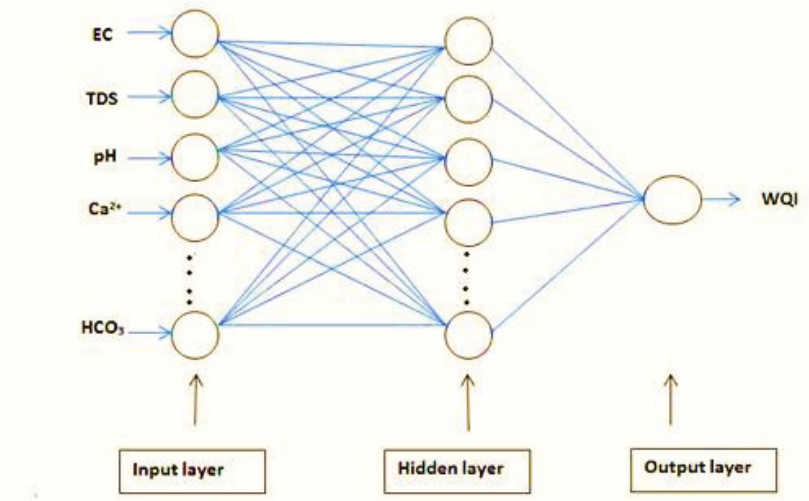

Fig. 2. Structure of a multi-layer feed forward ANN model.

In this present paper, eleven parameters were selected as inputs (EC, pH, TDS, Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, Cl^{-}, SO_4^{2-}, NO_3^{-} and HCO_3^{-}) and one as output (WQI). For the training, and testing and validation process, 70%, and 30% of collected data, respectively, were utilized. During the creation of a neural model, a tangent sigmoid activation function was applied between the input layer to the hidden layer, and a linear transfer function was employed between the hidden to the output layer. To determine the optimal number of hidden nodes for the input layer, the trial-and-error methods were the most frequently used (RMSE, *R*, and MAE). If more neurons are used than necessary, problems may arise in treating the new types of data groups in the same network. Therefore, it is generally suggested to use one hidden layer within the network. After training all the architectures and comparing the results, a single hidden layer was considered in the current study, with a tangent sigmoid activation function between the input and hidden layers, and a linear activation function between hidden and output layers, The model performance function was based on error measurements [19].

3. Results and discussion

3.1. Analysis of water quality parameters

The groundwater hydro chemical properties of the study area are summarized in Table 2. The World Health Organization [31] drinking standards, the statistical database of physicochemical parameters, and their assigned weights (w_i) are also illustrated in this table. The results showed that the pH ranged from 6.9 to 8.3 in the wet season and from 7.0 to 7.8 in the dry season, this is within the permissible limits of the WHO standards. The concentration of calcium indicated that more than 29% of samples were below the WHO's norm, with a mean of about 147 and 251 mg/L in the wet and dry season, respectively. The highest values were observed in the center of the plain with values ranging from 328.6 to 385 mg/L in wet and dry seasons, respectively. It can be argued that the high value of Ca^{2+} could be due to either the dissolution of carbonate formations ($CaCO_3$), or the dissolution of gypsum formations ($CaSO_4$). For the magnesium, the values are comparable to those of calcium, because they come from the dissolution of carbonate formations with high magnesium contents (magnesite and dolomite) from the Triassic of Ouarsenis [34].

Many sampled water points (approximately 30%) provided water with magnesium contents above the drinking water standard of 75 mg/L, while 70 % of the observations were below the norm. The average value for Mg^{+2} was 65.7 and 69.2 mg/L in wet and dry seasons, respectively. Magnesium and sodium have the same homogeneous spatial distributions, which increase in the northwest and eastern parts of the plain (Fig. 3a and b). The concentration of sodium shows that about 57% of the samples were below

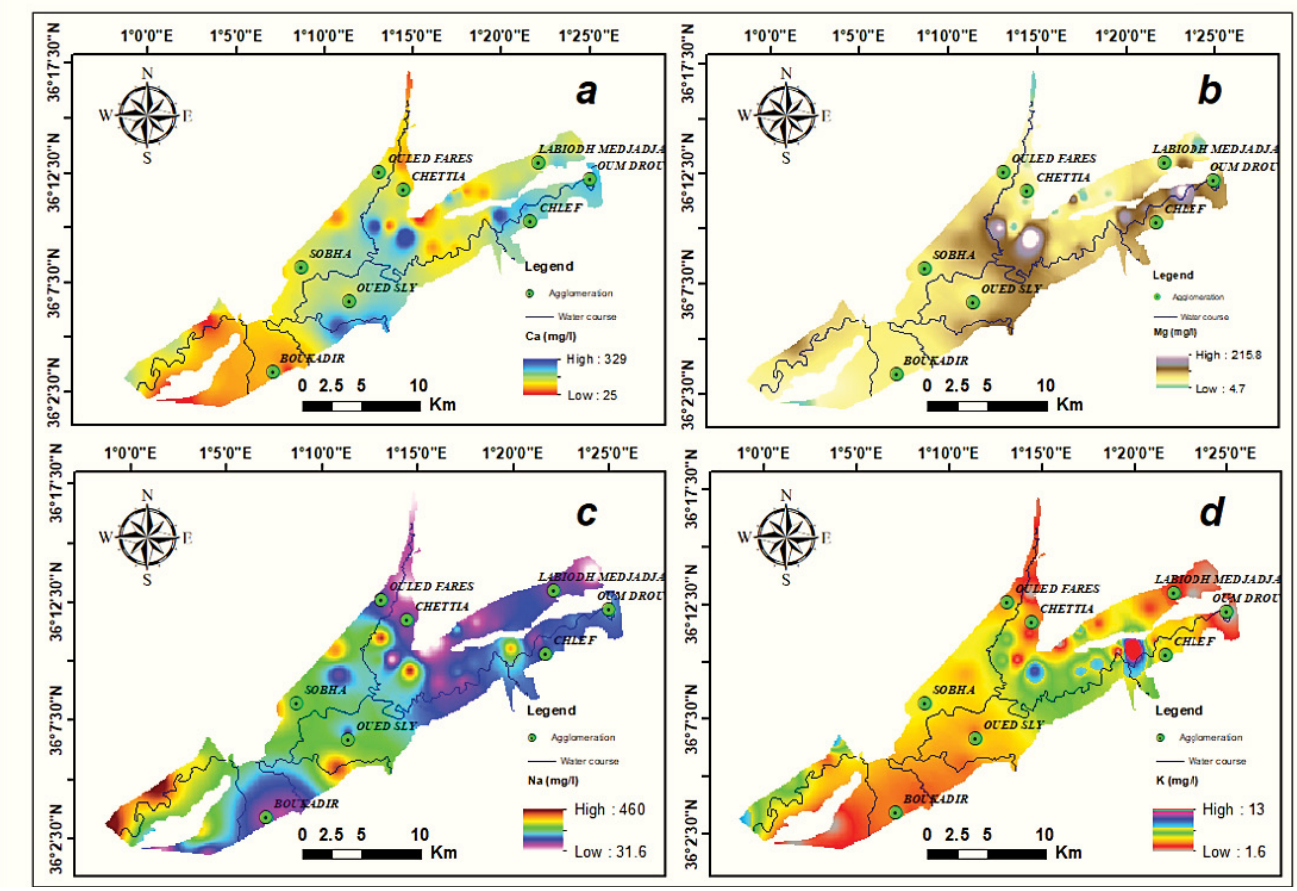

Fig. 3. Spatial distribution maps for the concentrations of major cations: (a) calcium, (b) magnesium, (c) sodium, and (d) potassium.

Table 2
Statistics description of water quality (WQ) parameters and assigned weights of each

Parameter	WHO [31]	Descriptive statistics								Weight (w_i)	Percentage of samples below the permissible limits (%)
		Wet season				Dry season					
		Max.	Min.	Mean	CV	Max.	Min.	Mean	CV		
pH	8.5	8.3	6.9	7.3	4.6	7.8	7	7.3	2.9	3	100
EC	1,500	5,350	650	2,338.4	42.1	5,540	886	3,539	48.6	5	23
TDS	1,000	3,424	416	1,525.5	43.9	3,540	567	2,260	49.1	5	25
Na^+	150	460	31.6	175.8	61.7	520	59	245	59.2	3	57
Ca^{2+}	100	328.6	25	147.0	40.5	385	52	251	52.4	2	29
Mg^{2+}	75	215.8	4.7	65.7	62.0	106	11	69.2	52.3	3	70
K^+	12	13	1.6	3.7	47.3	13	2	4.6	77	2	98
Cl^-	250	2,312.5	75.9	347.5	88.7	1,295	88	547	67.8	4	48
SO_4^{2-}	200	653	13	249.8	77.3	1,190	25	458	85.3	4	52
NO_3^-	50	81.6	0.1	22.6	43.0	75	3	30	77.9	5	95
HCO_3^-	300	683.2	137	324.5	31.9	445	198	345	26.5	1	37

All values are in mg/L except pH, and EC (μS/cm)

the standards for drinking water, with a wide variation from 31.6 to 460 mg/L and 59 to 520 mg/L with an average of 175.8 and 245 mg/L in wet and dry season, respectively. These values varied with a decrease in rainfall. Fig. 3c shows an increasing trend of Na^+ in the west and center of the valley. The potassium concentration exceeded 12 mg/L in three wells in the center of the plain (Fig. 3d). One can speculate that this came from the alteration of potassium clays and the dissolution of chemical fertilizers (NPK) which were used by farmers in agriculture. The presence of this may also be associated with wastewater effluents discharge. The mean value of k^+ was 3.7 and 4.6 mg/L in wet and dry seasons, respectively.

The chloride concentrations showed that 48% of sampled water points were below the WHO's norm and 52% of observations were above the norm with a content ranging from 544.6 to 662.9 mg/L The highest value (2,312 mg/L) was recorded in a single well in the wet season (Table 2). The elevated content of Cl^- was presumably due to discharge of chlorinated fertilizers, the dissolution of evaporite deposits (the dissolution of the halite) the permanent interaction of the groundwater with the marly substratum; the association of the marly miocene soils [35] and the gypsiferous formation which mainly outcrops north of Ouled Fares. This formation is responsible for the salinity of certain runoff water and consequently for the salinity of aquifers [35].

Additionally, the sources of SO_4^{2-} in groundwater can be attributed to dissolution and oxidation of sulfate minerals, discharges from industrial and domestic sewers, as well as leaching of waste deposits. The average values of SO_4^{2-} varied from 249 to 458 mg/L in wet and dry seasons, respectively. Chloride and sulfate have similar spatial distributions, increasing towards the northwest, east and in the central part of the study area (Fig. 4a and b). Likewise, a few samples had nitrate levels above the standard limit of 50 mg/L (Fig. 4c). The sources of nitrate in groundwater were presumably industrial wastewaters, nitrogenous fertilizers, infiltration of surface water, return sewage water, and agricultural sewage usage [36].

Fig. 4d shows an increasing trend of HCO_3^- in the east of the valley, the values range from 324.5 to 345 mg/L in wet and dry seasons, respectively. The concentration of bicarbonate in water depending on the types of soil it crosses (infiltration) or its flow (runoff). Furthermore, TDS values ranged from 416 to 3,424 mg/L in the wet season and from 430 to 3,590 mg/L in the dry season. These findings far exceeded WHO standards.

The outcomes showed that most samples (77%) had the highest values of EC and exceeded the WHO's norm (1,500 μS/cm). Only 34% of samples exceeded the Algerian norm (2,800 μS/cm). Spatial variability in CE and TDS (Fig. 5) shows that mineralization rises toward the northwest, east, and in the central part of the study area. The high values of EC in the center of the plain were likely due to the anthropological or man-made pollution of groundwater or due to the water–rock interaction (i.e., the geology of the aquifer). Also, the EC was strongly dependent on the chemical composition of water and its temperature [12].

3.1. Assessment of water quality using WQI

In this investigation, WQI values of groundwater samples varied from 37-259 for the wet season (Fig. 6). The highest WQI values were observed at sample well F07, and the lowest values were observed at sample well 105/452. Fig. 7 summarizes the potable water quality variation which could be classified from excellent to very poor, with 6% of water samples in the "excellent" category, 28% as "good", 60% in the "poor" category and 6% as "very poor".

3.2. Geostatistical evaluation of spatial distribution of groundwater WQI using ordinary kriging and semi-variogram models

In the kriging method, different experimental variograms, containing Gaussian, exponential, circular,

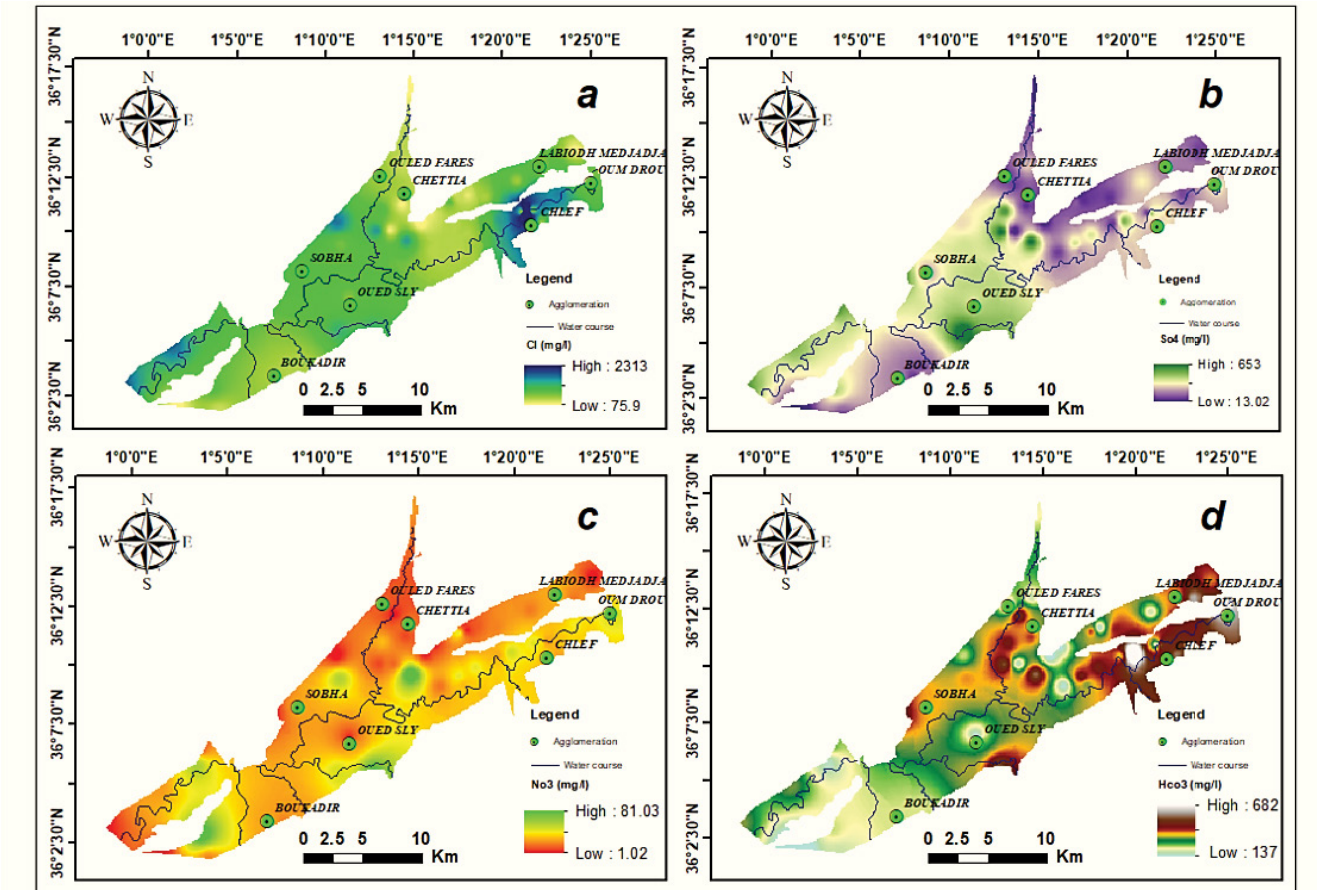

Fig. 4. Spatial distribution maps for the concentrations of major anions: (a) chloride, (b) sulfate, (c) nitrate, and (d) bicarbonate.

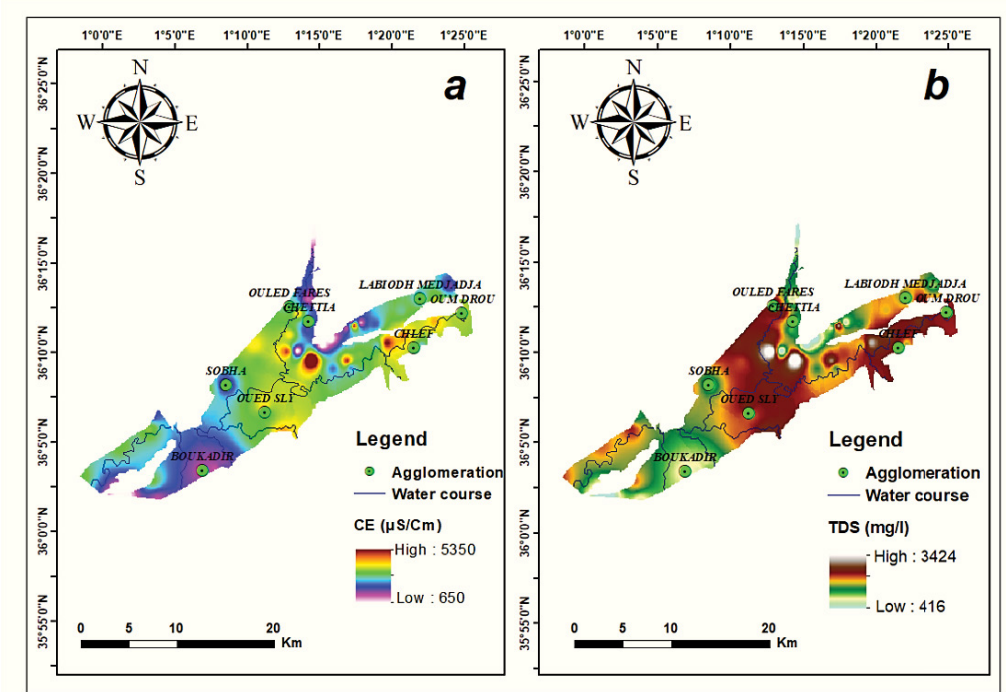

Fig. 5. Spatial distribution map for the concentrations of: (a) electrical conductivity and (b) total dissolved solids.

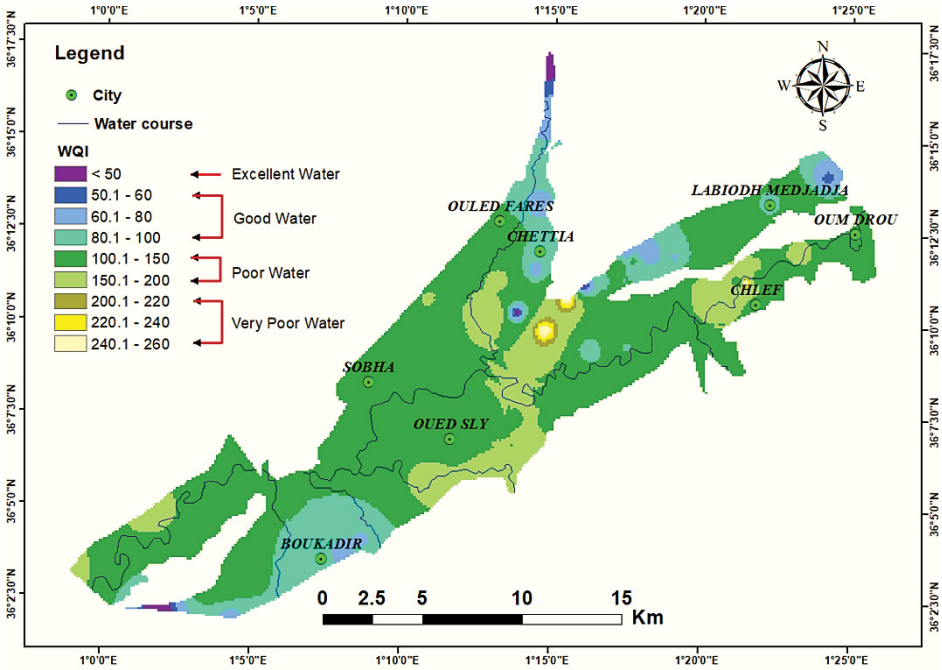

Fig. 6. Spatial distribution of water quality index.

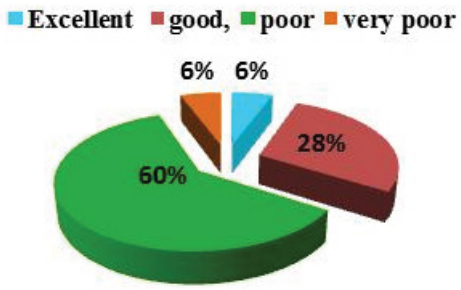

Fig. 7. Distribution of WQI by class or category (%) in the case study area.

spherical, and stable, were selected based on the lowest error. Results are presented in Table 3. The Gaussian model was identified as the best for the WQI. It had a lower error and more accuracy than the other four models used in this study with an RMSE value of 37.036, an MAE value of 26.447 and a correlation coefficient of 0.52. According to the water quality map (Fig. 8), it is important to note that the spatial distribution of the WQI could be divided into three scattering levels or classes. In the northeast and southwest, the WQI was good, ranging from 52–100, while it was the lowest quality (poorest) in the rest of the case study area with WQI greater than 100.

3.3. Artificial neural networks model evaluation for WQI estimation

The applicability of ANN to predict WQI values was verified in 54 wells from the study area. Model performance was evaluated by monitoring the error between the pattern output and measured data set. The model performance results are presented in Table 4 and Fig. 9. The best ANN model at the training stage was found at 11 hidden nodes with the values for RMSE, MAE and *R* of 5.72, 3.47, and 0.9386, respectively.

This was verified at the testing stage with values of 2.347, 0.71 and 0.9998, respectively in validation. The performance criteria shown in Fig. 9 indicate accurate results for training, validation, and testing periods with more exactness in the validation and testing stage with values of 0.9998 and 0.9813 respectively. At 13 hidden nodes reveals a lowest accuracy at the training and testing stage with the values for RMSE, MAE and *R* of 18.85, 6.29, and 0.7737, respectively. The predicted WQI values of the ANN model were verified with the measured values and hence, it demonstrates that ANN is an effective tool in the assessment of the WQI.

The results obtained showed that 34% of the samples were in the excellent and good category, 7% in the very poor class and 59% were in the poor category. Excellent and good quality water was observed in the Ouled Fares, Labiodh Medjadja and Boukadir locations (Fig. 10) while very poor water was found in the south west, central and northeast of the case study area.

A very salty surface layer was observed near the Tsighaout Wadi, close to Chlef. The spatial distribution of the WQI in Fig. 10 shows that the excellent and good categories are in the northeast and southwest of the study area. On the other hand, the very poor category was observed in the location of Chettia, which is known for its high population density, in the southwest of the study area which encloses a very salty surface layer and near the Tsighaout Wadi presumably due to wastewater discharge. It can be argued that the WQI values indicating poor quality of

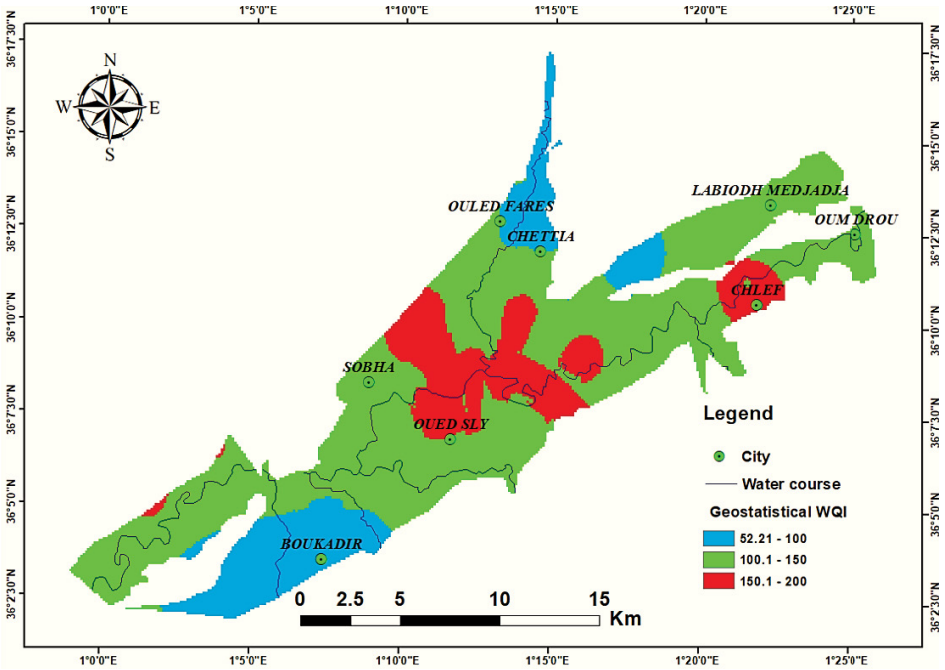

Fig. 8. Spatial distribution of water quality index based on ordinary kriging model.

groundwater for consumption are mainly the results of domestic and/or agricultural discharges.

3.4. Comparing the outcomes of the geostatistical and ANN models

According to the results of Tables 3 and 4, it can be concluded that among methods of groundwater quality prediction, the ANN model had the greatest accuracy in the estimation method with the highest R (0.99) and the lowest error rate. Also, when expected values were analyzed in detail (Fig. 11), the ANN model had the higher performance compared to the geostatistical method. For example, Fig. 11 shows the comparison of the WQI values measured and predicted using ANN and geostatistical (Gaussian) model. WQI values of 50 groundwater samples out of the 54 testing water samples were predicted with an error ratio of less than 5% by ANN model. While 3 samples were predicted between 5% to 10% error ratio, and 1 groundwater sample 10 to 20% error ratio. Whereas WQI values predicted by geostatistical (Gaussian model) reveal an error ratio lower than 5% with 12 samples out of the 54 testing water samples, 8 samples were predicted between 5% to 10% error ratio and 16, 12 and 5 groundwater samples between 10% to 20% error ratio, 20% and 50% error ratio and error ratio of greater than 50%, respectively. Thus, it can be concluded that ANN is the better model compared to the geostatistical model.

Table 3
Performance parameters for ordinary kriging model

Model	Structure	MAE	RMSE	R
Ordinary kriging (OK)	Gaussian	26.447	37.036	0.52
	Spherical	32.56	45.267	0.41
	Stable	32.65	45.01	0.48
	Circular	32.80	45.42	0.44
	Exponential	39.90	54.41	0.28

Table 4
Performance statistics of ANN model for groundwater quality index estimation

Number of hidden nodes	Training			Testing		
	RMSE	MAE	R	RMSE	MAE	R
9	26.30	11.61	0.73521	14.582	6.258	0.7911
10	20.638	3.764	0.9091	8.541	1.681	0.8620
11	5.72	3.47	0.9386	2.347	0.71	0.99986
12	16.54	12.07	0.9464	17.325	7.139	0.7648
13	18.856	6.929	0.7737	11.448	4.609	0.1041
15	20.658	3.764	0.9955	12.488	3.576	0.6634

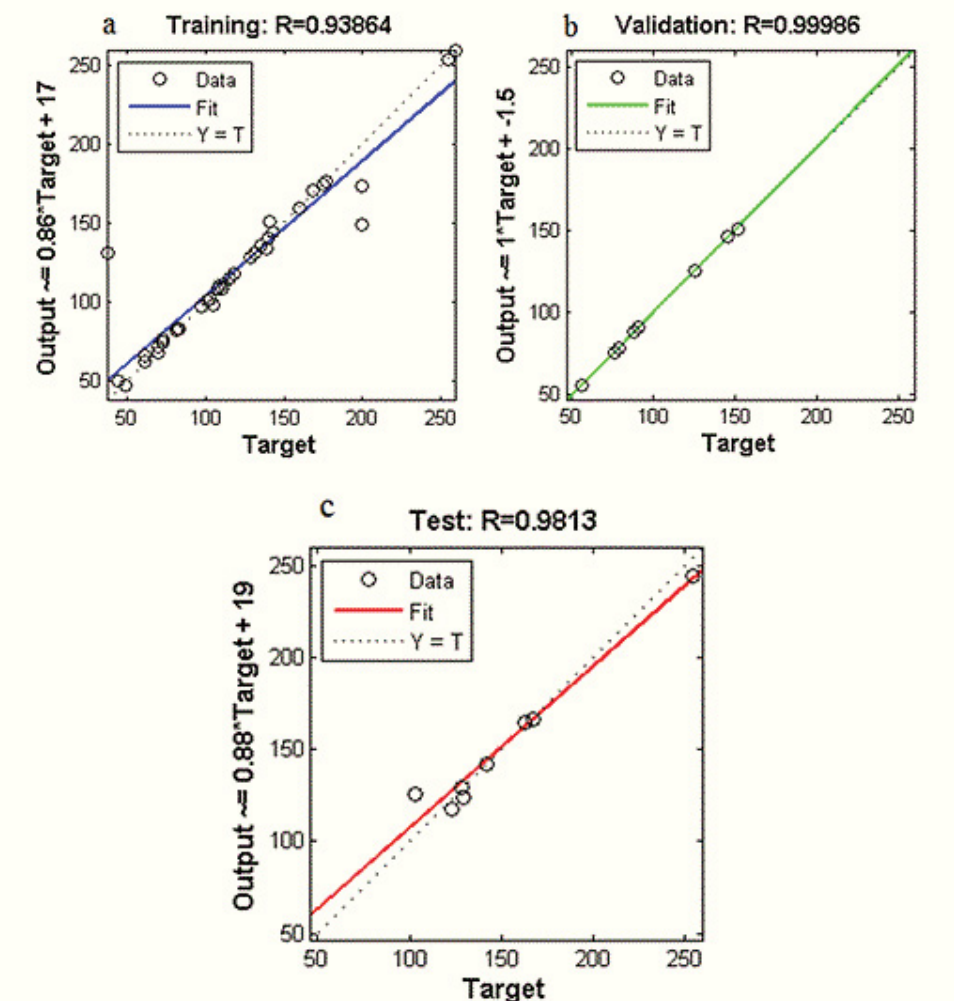

Fig. 9. The measured values against estimated value of WQI; (a) training data, (b) validation and (c) testing.

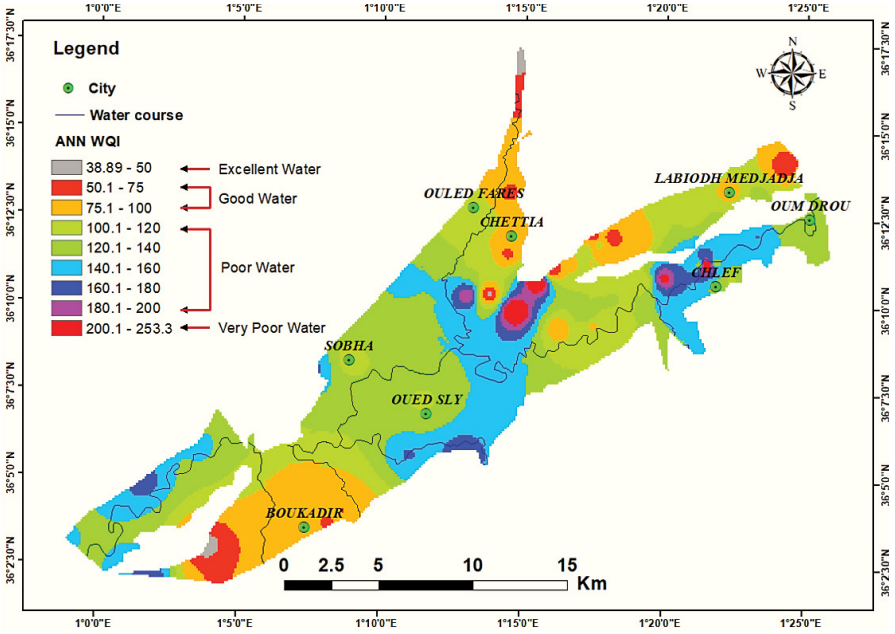

Fig. 10. Spatial distribution of water quality index using the ANN model.

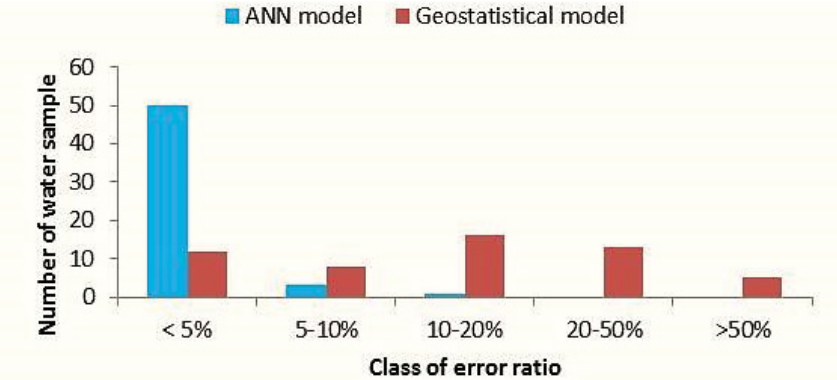

Fig. 11. Comparison of error ratios between ANN and geostatistical models.

4. Conclusions

The ANN model with its high *R* (0.99) and low error rate had a greater accuracy than the geostatistical ordinary kriging (OK) system in estimating the groundwater quality (WQI). The geostatistical model, with its relatively lower precision, depends on the spatial location of the variables. However, correctness in approximating the variable depends on the number of samples that can be obtained from the region. Therefore, in areas like the current case study which are limited by the number of samples, it was difficult to discover the relationship between the spatial location of sampling and the variable. Whereas intelligent models such as ANN, were more capable of obtaining this connection.

The assessment of groundwater quality is very important in arid areas where resources are limited. The result of the present research showed that excellent water quality (i.e., WQI) is found in the northeast and southwest of the case study area. On the other hand, very poor groundwater was observed in the high population density region, as well as in the southeast which had a very salty surface layer and contained wastewater discharge sites. It can be argued that the WQI values indicating poor quality of groundwater for consumption are mainly the result of domestic and/or agricultural releases. The multi-layer perceptron (MLP) model in feed-forward ANN gave the most reliable results. The significance of this analysis showed that for semi-arid regions, modeling groundwater quality using ANN is an important tool for helping decision-makers to manage drinking water supplies more effectively. It will assist them, for example, in choosing the best sites for drilling new groundwater boreholes.

References

[1] S. Sargazi, M. Mokhtari, M.H. Ehrampoush, S.A. Almodaresi, H. Sargazi, M. Sarhadi, The application of geographical information system (GIS) approach for assessment of groundwater quality of Zahedan city, Sistan and Baluchestan Province, Iran, Groundwater Sustainable Dev., 12 (2021) 100509, doi: 10.1016/j.gsd.2020.100509.
[2] Y. Elmeddahi, A. Issaadi, H. Mahmoudi, M. Tahar Abbes, G. Mattheus F.A., Effect of climate change on water resources of the Algerian Middle Cheliff basin, Desal. Water Treat., 52 (2014) 2073–2081.
[3] Y. Elmeddahi, H. Mahmoudi, A. Issaadi, M.F.A. Goosen, Analysis of treated wastewater and feasibility for reuse in irrigation: a case study from Chlef, Algeria, Desal. Water Treat., 57 (2016) 5222–5231.
[4] Y. Elmeddahi, H. Mahmoudi, A. Issaadi, M.F.A. Goosen, R. Ragab, Evaluating the effects of climate change and variability on water resources: a case study of the Cheliff basin in Algeria, Am. J. Eng. Appl. Sci., 9 (2016) 835–845.
[5] Y. Elmeddahi, R. Ragab, Assessing the Climate Change Impact on Water Resources and Adaptation Strategies in Algerian Cheliff Basin, In: Water Resources in Algeria - Part I: Assessment of Surface and Groundwater Resources, Hdb Env. Chem. Springer, Cham, Switzerland, 2020, pp. 111–134, doi: 10.1007/698_2019_398.
[6] M. El Baba, P. Kayastha, M. Huysmans, F. De Smedt, Evaluation of the groundwater quality using the water quality index and geostatistical analysis in the Dier al-Balah Governorate, Gaza Strip, Palestine, Water, 12 (2020) 262, doi: 10.3390/w12010262.
[7] S. Abanyie, S. Ampofo, N. Biyogue Douti, T. Boateng, Geospatial assessment of groundwater quality in the Savelugu-Nanton Municipality, Northern Ghana, Afr. J. Appl. Res., 4 (2018) 93–105.
[8] E.D. Sunkari, M. Abu, P.S. Bayowobie, U.E. Dokuz, Hydrogeochemical appraisal of groundwater quality in the Ga West Municipality, Ghana: implication for domestic and irrigation purposes, Groundwater Sustainable Dev., 8 (2019) 501–511.
[9] M. Uyan, T. Cay, Spatial analyses of groundwater level differences using geostatistical modeling, Environ. Ecol. Stat., 20 (2013) 633–646.
[10] A. Shaheen, J. Iqbal, Spatial distribution and mobility assessment of carcinogenic heavy metals in soil profiles using geostatistics and Random Forest, Boruta Algorithm, Sustainability, 10 (2018) 799, doi: 10.3390/su10030799.
[11] R. Webster, M.A. Oliver, Geostatistics for Environmental Scientists, 2nd ed., John Wiley & Sons, Ltd., 2007. Available at: https://www.academia.edu/6871734/Geostatistics_for_Environmental_Scientists_Second_Edition
[12] A. Bouderbala, The impact of climate change on groundwater resources in coastal aquifers: case of the alluvial aquifer of Mitidja in Algeria, Environ. Earth Sci., 78 (2019) 1–13.
[13] N. Rajmohan, M.H.Z. Masoud, B.A.M. Niyazi, Assessment of groundwater quality and associated health risk in the arid environment, Western Saudi Arabia, Environ. Sci. Pollut. Res., 28 (2021) 9628–9646.
[14] P. Sahu, P.K. Sikdar, Hydrochemical framework of the aquifer in and around East Kolkata Wetlands, West Bengal, India, Environ. Geol., 55 (2008) 823–835.
[15] A. Hamlat, A. Guidoum, Assessment of groundwater quality in a semiarid region of Northwestern Algeria using water quality index (WQI), Appl. Water Sci., 8 (2018) 1–13.
[16] T.E. Keskin, M. Düğenci, F. Kaçaroğlu, Prediction of water pollution sources using artificial neural networks in the study

areas of Sivas, Karabük and Bartın (Turkey), Environ. Earth Sci., 73 (2015) 5333–5347.
[17] M.A. Ghorbani, M.T. Aalami, L. Naghipour, Use of artificial neural networks for electrical conductivity modeling in Asi River, Appl. Water Sci., 7 (2017) 1761–1772.
[18] C. Gao, M. Gemmer, X. Zeng, B. Liu, B. Su, Y. Wen, Projected streamflow in the Huaihe River Basin (2010–2100) using artificial neural network, Stoch. Environ. Res. Risk Assess., 24 (2010) 685–697.
[19] N. Naeamikhah, T. Nasrabadi, Z.Z. Sirdari, Role of different parameters in the quantification of generated sludge in the oxylator unit of water treatment plants, using artificial neural network model (case study of Jalalieh water treatment plant, Tehran, Iran), Appl. Ecol. Environ. Res., 15 (2017) 129–142.
[20] E.F. Derradji, F. Benmeziane, A. Maoui, H. Bousnoubra, N. Kherici, Evaluation of salinity, organic and metal pollution in groundwater of the Mafragh Watershed, NE Algeria, Arabian J. Sci. Eng., 36 (2011) 573–580.
[21] L. Belkhiri, L. Mouni, Hydrochemical analysis and evaluation of groundwater quality in El Eulma area, Algeria, Appl. Water Sci., 2 (2012) 127–133.
[22] F. Touhari, M. Meddi, M. Mehaiguene, M. Razack, Hydrogeochemical assessment of the Upper Cheliff groundwater (North West Algeria), Environ. Earth Sci., 73 (2015) 3043–3061.
[23] A. Saou, M. Maza, J.L. Seidel, Surface and groundwater quality, in the mixing area, using hydrochemical and statistical methods, in the Valley of Low Djemaa — Zitouna Basins, Algeria, J. Geol. Soc. India, 95 (2020) 315–320.
[24] Y. Benmoussa, B. Remini, M. Remaoun, Quality assessment and hydrogeochemical characteristics of groundwater in Kerzaz and Beni Abbes along Saoura valley, Southwest of Algeria, Appl. Water Sci., 10 (2020) 1–10.
[25] B. Hallouche, F. Hadji, A. Marok, L. Benaabidate, Spatial mapping of irrigation groundwater quality of the High Mekerra watershed (Northern Algeria), Arabian J. Geosci., 10 (2017) 1–15.
[26] A. Bouderbala, Assessment of groundwater quality and its suitability for domestic and agricultural uses in Low-Isser plain, Boumedres, Algeria, Arabian J. Geosci., 10 (2017).1–13, Internal Report, 277p (in French)
[27] R. Kouadra, A. Demdoum, Hydrogeochemical characteristics of groundwater and quality assessment for the purposes of drinking and irrigation in Bougaa area, Northeastern Algeria, Acta Geochim., 39 (2020) 642–654.
[28] O. Bouteraa, A. Mebarki, F. Bouaicha, Z. Nouaceur, B. Laignel, Groundwater quality assessment using multivariate analysis, geostatistical modeling, and water quality index (WQI): a case of study in the Boumerzoug-El Khroub valley of Northeast Algeria, Acta Geochim., 38 (2019) 796–814.
[29] R. Kechiched, I.E. Nezli, A. Foufou, M.S. Belksier, S.A. Benhamida, R. Djeghoubbi, N. Slamene, O. Ameur-zaimeche, Fluoride-bearing groundwater in the complex terminal aquifer (a case study in Hassi Messaoud area, southern Algeria): hydrochemical characterization and spatial distribution assessed by indicator kriging, Sustainable Water Resour. Manage., 6 (2020) 1–14.
[30] S.M. Leulmi, A. Aidaoui, D.D. Hallal, M.A. Khelfi, J.P.L. Ferraira, A. Ammari, O. Aziez, Spatio-temporal analysis of nitrates and piezometric levels in groundwater using geostatistical approach: case study of the Eastern Mitidja Plain, North of Algeria, Arabian J. Geosci., 14 (2021) 1–15.
[31] WHO, Guideline for Drinking Water Quality, 3rd ed., World Health Organization, Geneva, 2008. Available at: https://www.who.int/water_sanitation_health/dwq/fulltext.pdf
[32] R. Khosravi, H. Eslami, S.A. Almodaresi, M. Heidari, R.A. Fallahzadeh, M. Taghavi, M. Khodadadi, R. Peirovi, Use of geographic information system and water quality index to assess groundwater quality for drinking purpose in Birjand city, Iran, Desal. Water Treat., 67 (2017) 74–83.
[33] B. Bhushan, M. Singh, Y. Hage, Identification and control using MLP, Elman, NARXSP and radial basis function networks: a comparative analysis, Artif. Intell. Rev., 37 (2012) 133–156.
[34] M. Schrambach, L. Mostefa, Exploitation and conservation of groundwater in the upper, middle, lower Cheliff and lower Mina plains, (1966).
[35] M. Elaid, M. Hind, B. Abdelmadjid, M. Mohamed, Contribution of hydrogeochemical and isotopic tools to the management of Upper and Middle Cheliff Aquifers, J. Earth Sci., 31 (2020) 993–1006.
[36] B. Yan, C. Xiao, X. Liang, Z. Fang, Impacts of urban land use on nitrate contamination in groundwater, Jilin City, Northeast China, Arabian J. Geosci., 9 (2016) 1–8.