International Journal of Engineering Research And Management (IJERM)
ISSN : 2349- 2058, Volume-03, Issue-05, May 2016

# Artificial Neural Network Modeling of the Groundwater Quality (Case Study: Zahrez Basin in Algeria)

**Fatah Bouteldjaoui, Mohamed Bessenasse, Ahmed Kettab**

*Abstract*— **Artificial neural networks (ANNs) model are widely used in water resources applications to predict and forecast water resources' variables. The objective of this study is to investigate the abilities of an artificial neural networks' model to predict the total dissolved solid (TDS) and electrical conductivity (EC). Water quality variables such as pH, calcium ($Ca^{2+}$), magnesium ($Mg^{2+}$),sodium ($Na^{+}$), potassium ($K^{+}$), bicarbonate (HCO3-), chloride ($Cl^{-}$), nitrate ($NO_3^{-}$) and sulfate ($SO_4$) were used as the input data to obtain the output of the neural network(TDS and EC). Performance of the ANN models was evaluated using correlation coefficient (R), Nash-Sutcliffe coefficient of efficiency (NASH), root mean square error (RMS), Normalised Root Mean Square Error (NRMSE) and Mean absolute error (MAE). computed from the measured and model computed values of the dependent variables.The results of this study reveal that the ANN- MLP (9, 9, 1) model gives the best estimates for the TDS prediction. The results of neural network modeling to predict electrical conductivity (EC) indicate that the ANN- MLP (9, 12, 1) model showed better predictive ability in the determination of EC. The identified ANN models can be used as tools for the computation of groundwater quality parameters in Zahrez basin.**

*Key words*— **Artificial neural network (ANN), Groundwater Multilayer perceptron, Zahrez basin**

## I. INTRODUCTION

Zahrez basin is located in the High Plateaus of the north Algeria, characterized by a semi-arid climate where annual rainfall is highly irregular. Because of the scarcity of surface water, groundwater is a major source of water supply in different cities around the study area. Groundwater resources in the study area occur in four main hydrogeological units: (1) Mioplioquaternary consists of conglomerate and clay, (2) Turonian made up of fractured limestone, (3) Albian, and (4) Barremian aquifer made up of sandstone [1]. The mioplioquaternary aquifer is a source of fresh water for the city of Djelfa with a population of 1,491,370 inhabitants. Groundwater quality is affected by a wide range of natural and anthropogenic factors. Natural processes (hydrological, physical, and chemical) may affect the characteristics and concentration of chemical elements in groundwater [2-3]. In addition, there are also anthropogenic impacts due to urbanization, industrial and agricultural activities in the basin. Assessment of the groundwater quality as well as development of management strategies for the protection of water resources is one of the essential objectives for the future development of a country, especially when the rising demand for clean drinking water is considered [4-5]. Artificial neural networks (ANNs) models have been successfully used in hydrological processes, water resources, water quality prediction [6-9]. Niroobakhsh et al. [8] used two ANN networks, multilayer perceptron (MLP) and radial basis function (RBF) to compute the total dissolved solid (TDS) concentrations for the Jajrood River of Iran. In their study, they found that MLP and RBF are able to simulate water quality variables of Jajrood River with more than 90% accuracy. Singh et al. [8] computed dissolved oxygen (DO) and biochemical oxygen demand (BOD) levels in the Gomti River in India using three-layer feed forward neural networks with back propagation learning. The coefficient of determination for modeled values and observed DO values were 0.70, 0.74, and 0.76 for the training, validation and test sets, respectively. The main objective of this study is to construct an artificial neural network (ANN) model for the prediction of total dissolved solids (TDS) and electrical conductivity (EC) in Zahrez basin and demonstrate its application to complex water quality data as how it can improve the interpretation of the results. Here, we have investigated the possibility of training ANN models correlating the primary water quality variables (independent) with their secondary attribute (dependent variable). The TDS and EC of the groundwater were taken as the dependent variables here and set of other parameters constituted the independent variables.

**Manuscript received May 20, 2016**.
**Fatah Bouteldjaoui,** Ecole Nationale Polytechnique, Alger, Algérie
**Mohamed Bessenasse**, Université Saad Dahlab, Blida, Algérie.
**Ahmed Kettab,** Ecole Nationale Polytechnique, Alger, Algérie

## II. MATERIALS AND METHODS

### A. Study area

The Zahrez basin (Fig.1) is one of the endorheic basins of the vast steppes region in the central northern part of Algeria. The Zahrez hydrological basin covers approximately 8,989 $km^2$. Topography of the area is relatively flat with an elevation ranging from 900 to 1330 meters above mean sea level [10]. The catchment lies between longitudes 2° 15' to 4° 08'E and latitudes 34° 35' to 35° 30'N. The area is characterized by a semi-arid climate, typically Mediterranean, with an irregular annual rainfall. The mean annual rainfall and potential evapotranspiration are 250 and 1380 mm, respectively, exceeding rainfall for most of the year. The mean monthly temperature varies between 3°C and 25°C. The precipitation period in a typical year is between October and March and the dry period can extend from April to September [1]. The ephemeral rivers of the region, locally called ''wadi'', have an intermittent flow regime, because the dry season is typically very long (6–8 months) every year. The main wadis in this basin are the Melah and Hadjia rivers which receive many important flow tributaries. The drainage density of the area ranges between 1.4 and 1.8 km/$km^2$ [10].

*B. Water quality data set*

A tot al of 47 water samples groundwater samples were collected during sampling campaigns carried out in October 2012 at wells in various parts of the study area (Fig. 1). Electrical conductivity (EC), temperature and pH were measured in the field, using the portable Orion EC and pH meters. Water samples were filtered through a 0.45 µm cellulose membrane and collected in 100 ml polyethylene bottles for major and minor element analysis which have been done at the National Agency for Water Resources (ANRH). Cations ($Ca^{2+}$, $Mg^{2+}$, $Na^{+}$, $K^{+}$) were analyzed by atomic absorption spectrometry, anions ($Cl^{-}$, $SO_4^{2-}$ and $NO_3^{-}$) by high performance ionic liquid chromatography (HPILC). Bicarbonates ($HCO_3^{-}$) were determined by acid-base titration method [11].

*C. Data processing*

Before the network training, the original data were normalized in accordance with the requirements of the BP algorithm. The values applied in the input and output layers were normalized by the following formula in the range of (0–1).

$$x_i^{norm} = \frac{x_i - x_i^{min}}{x_i^{max} - x_i^{min}} \quad (1)$$

Where xi, $x_i^{min}$, $x_i^{max}$, and $x_i^{norm}$ denote, respectively, values of input (output) variables i, minimum value of input (output) variable, maximum value of input (output) variable and the normalized value of i.

The de-normalized value of the ANN was computed using:

$$y_t = y_{min} + y_{norm}\left(y_{max} - y_{min}\right) \quad (2)$$

Where $y_t$, $y_{min}$,$y_{max}$ and $y_{norm}$ are, respectively, real valued output variable, minimum and maximum values of real-valued output and the normalized output value from the neural-ANN model.

The proportion of ANN training set from the available data ranged generally from 25% to 80%.The proportion of ANN testing set from the available data is about 15 to 20% and the proportion of the validation data set is a bout (5 to 15%) [12].

*D. Artificial neural networks modeling (ANNs)*

ANN models have been used successfully to model complex nonlinear input–output relationships particularly in situations where the explicit form of the relation between the variables involved is unknown [13-14]. As a nonlinear statistical technique, ANNs can be used to solve problems that cannot be addressed by traditional approaches [15]. The ANN architecture is composed of an input layer, a certain number of hidden layers and an output layer in forward connections. The input layer introduces data into the model and calculates the weighted sum of the input(s). The hidden layer or layers processes data, and the output layer produce the results of the ANN model. Each layer is com-posed of one or more basic element(s) called an artificial neuron or a node, which is connected to a network by a weight factor. A feed- forward neural network is commonly used for predicting and forecasting water quality variables[16-18]. The major steps for development of ANN models include defining the suitable model inputs, specifying network type, pre-processing and partitioning of the available data; determining network architecture; defining model performance criteria; training (optimization of connection weights); and validating the model [19-21].

*E. Multi-layer perceptron (MLP)*

A multilayer feed-forward network or multi-layer perceptrons (MLP), originally proposed by Rumelhart and McClelland [22], are the most commonly used and well-researched class of ANNs [23]. A MLP consists of an input layer, which receives the values of the input variables, an output layer, which provides the model output, and one or more hidden layers. Nodes in each layer are interconnected through weighted acyclic arcs from each preceding layer to the following, without lateral or feedback connections [24].

*F. Activation function*

The activation (transfer) function determines the response of a node to the total input signal it receives. The most commonly used activation function, named logistic sig-moid-type function was used in this study for the hidden layer [25-26]. However, a linear-type activation function was used for the output layer, as suggested by Maier and Dandy [21] and Rumelhart et al. [27]. The sigmoid function is a bounded, monotonic, non-decreasing function that provides a graded, non-linear response [28], whereas a linear-transfer function calculates a neuron's output by simply returning the value passed to it. The mathematical expressions for these two functions are as follows:

Linear function: $$f(n) = n \quad (3)$$

Logistic sigmoid function: $$f(n) = \frac{1}{1 + e(-n)} \quad (4)$$

Another sigmoid function is the tan-sigmoid transfer function, defined by

Tan sigmoid function:

$$f(n) = \frac{1 - e(-n)}{1 + e(-n)} \quad (5)$$

*G. Modeling performance criteria*

To determine the performance of each of the selected network model, five different criteria were used: the root mean square error (RMSE, the normalized Root Mean Square Error (NRMSE), the Nash-Sutcliffe Efficiency Index (NASH), and the mean absolute absolute error (MAE), and the correlation coefficient (R). The five indices are computed according to the following equations:

a) Root Mean Square Error is RMSE:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2} \quad (6)$$

b) Normalized Root Mean Square Error (NRMSE):

$$NRMSE = \frac{RMSE}{x_{max} - x_{min}} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2}}{x\ max - x\ min} \quad (7)$$

c) Nash-Sutcliffe coefficient of efficiency:

$$NASH = 1 - \frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (8)$$

d) Mean absolute error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i| \quad (9)$$

e) Correlation Coefficient ( R )

$$R = \frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}\sqrt{\sum_{i=1}^{N}\left(y_i - \bar{y}\right)^2}} \quad (10)$$

where, $y_i$ and $x_i$ are actual and obtained values of output, and N is the number of values.
$y_i$ is the mean estimation from the observed records for site i, $\hat{y}_i$ is the mean estimation obtained from the model for site i, and $\bar{y}$ is the mean of the mean estimation from the observed records of the n sites.
Where, yi and xi are actual and obtained values of output, y is the mean of actual output values.

## III. Results and Discussion

### A. Descriptive statistics

The details of descriptive statistics for groundwater quality parameter s are given in Table 1. Recorded groundwater pH varies from 7.6 to 10, indicating that the groundwater samples are mainly alkaline. The TDS values in the groundwater ranged from 300 to 4,012 mg/l, with a mean value of 1546 mg/l. TDS in the study area varying over two orders of magnitude from fresh (TDS < 1000 mg/l) to brackish (1,000 mg/l < TDS < 10,000 mg/l). The most dominant major cations are $Na^+$ and $Ca^{2+}$, while major anions are dominated by $Cl^-$ followed by $SO_4^{2-}$. Also, Table 1 reflects a moderate to high variability (standard deviation and coefficient of variation) of samples parameters. The highest variability was for $CO_3^{2-}$, followed by $K^+$, and $Na^+$ with a coefficient of variation greater than 1.0, reflecting the spatial variation of groundwater quality in the Zahrez basin.

### B. Artificial neural network (ANN)

In order to construct an artificial neural network ANN model for the total dissolved solids (TDS) and electrical conductivity (EC), the available 47 measured data set including $Ca^{2+}$, $HCO_3^-$, $Mg^{2+}$, $Na^+$, $K^+$, $Cl^-$, $NO_3^-$ and $SO_4^{2-}$; and pH variables were divided into three phases : 75%, 15% and 10 % of data set were chosen for training, testing, and validation phase, respectively. Different ANN models were constructed and tested in order to determine the optimum number of nodes in the hidden layer and transfer functions. Selection of an appropriate number of nodes in the hidden layer is very important aspect as a larger number of these may result in over-fitting, while a smaller number of nodes may not capture the information adequately. The suitable number of nodes (neurons) in hidden layers ranges from $(2n^{1/2}+m)$ to $(2n+1)$, to (2n + 1), where n is the number of input nodes and m is the number of output nodes [29]. To confirm the optimum structure of the AN N model, several models were constructed. The results are provided in Table 2 and 3.

### C. Total dissolved solids (TDS) models

The architecture of the best ANN models for the total dissolved solids (TDS) and electrical conductivity (EC) in the Zahrez groundwater is presented in Table 2. The best ANN model for the TDS is composed of one input layer with nine input variables, one hidden layer with nine nodes and one output layer with one output variable. It can be seen from Table 2 that the MLP (9,9,1) model provided a best fit model for the training and test data sets. The respective values of RMSE, NRMSE, and MAE for the two data sets are 99.10, 0.024 and 81.08 for training, and 140.42, 0.034 and 111.47 for testing. The correlation coefficients between the observed and predicted TDS values were 0.995, 0.976 and 0.984 for the training, test and validation sets, respectively. The NASH values corresponding to the training and testing sets are 0.95 and 0.94, respectively, suggesting good fit of the model to the data set. The comparison of the measured and predicted TDS values for the training, testing, and validation data sets are shown in Fig. 2 and 3. The correlation coefficients of training, testing, and validation were 0.995, 0.976, and 0.984, respectively, suggesting good fit of the model to the data set.

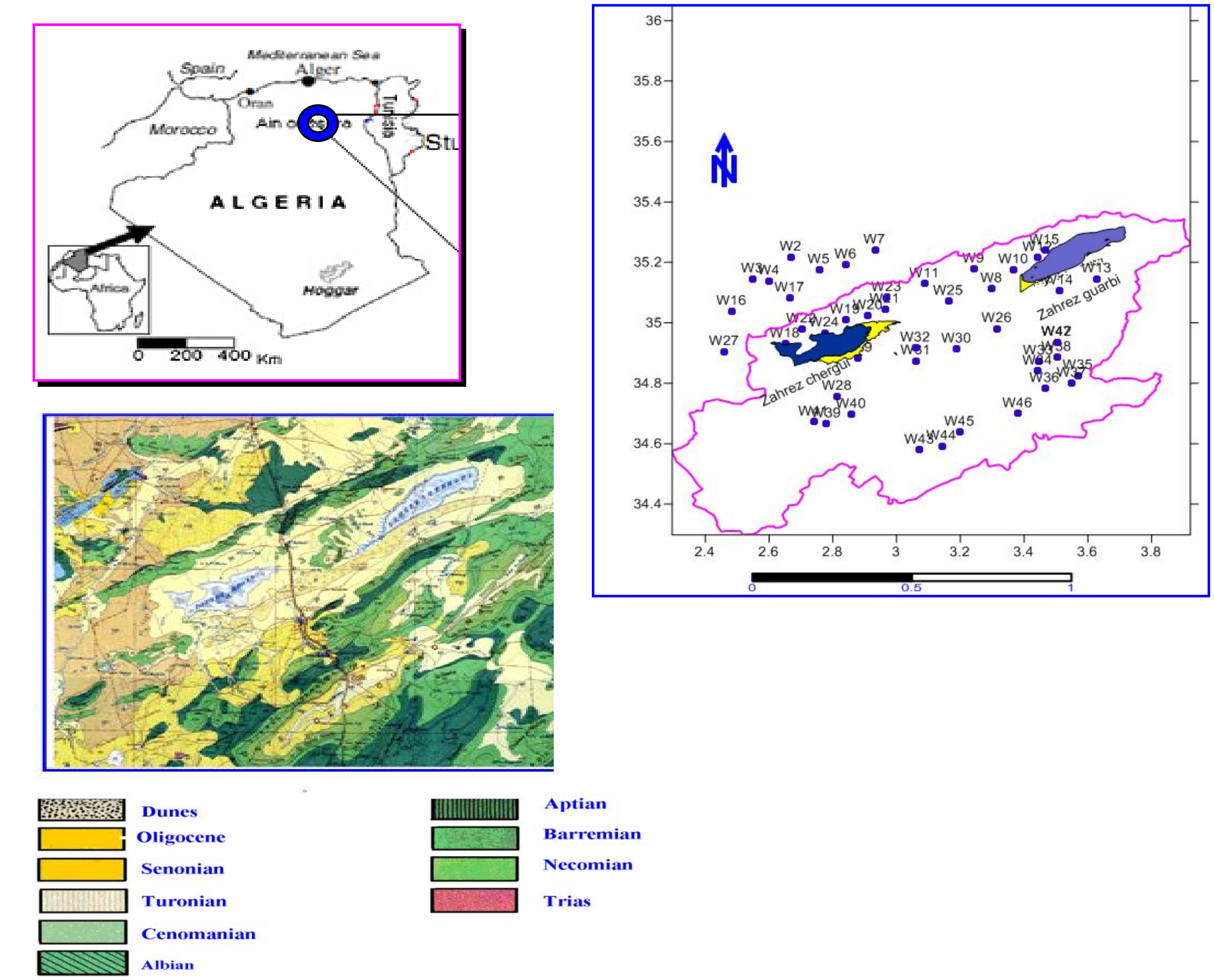


**Fig. 1.** Location map and geological formations of the Zahrez

**Table 1.** Descriptive statistics of groundwater quality parameters measured in the study area

| Parameter | Min | Max | Mean | Variance | SD | CV |
|---|---|---|---|---|---|---|
| EC (µS/ cm) | 450 | 7780 | 2829 | 3155617 | 1776.5 | 0.63 |
| Ca (mg/l) | 22.9 | 377 | 166.2 | 9167.2 | 95.75 | 0.58 |
| Mg (mg/l) | 20 | 149 | 68.2 | 1355.2 | 36.81 | 0.54 |
| Na (mg/l) | 17.6 | 933.9 | 228.2 | 59505.9 | 243.94 | 1.07 |
| K (mg/l) | 3.5 | 60.8 | 11.2 | 141.9 | 11.91 | 1.07 |
| Cl (mg/l) | 40 | 1712 | 506.7 | 178691 | 422.72 | 0.83 |
| $HCO_3$ (mg/l) | 0 | 275 | 83.3 | 4496.5 | 67.06 | 0.8 |
| $NO_3$ (mg/l) | 1 | 93.5 | 38.9 | 898.4 | 29.97 | 0.77 |
| pH | 7.6 | 10 | 8.4 | 0.4 | 0.6 | 0.07 |
| $CO_3$ (mg/l) | 0 | 24 | 4.7 | 51.3 | 7.16 | 1.51 |
| TDS (mg/l) | 300 | 4012 | 1546 | 914242 | 956.16 | 0.62 |
| $SO_4$ (mg/l) | 32 | 1250 | 411.1 | 90580.5 | 300.97 | 0.73 |

**Table 2.** Performance parameters of the artificial neural network for predicting the TDS

| ANN Structure | R | | | Nash | | | RMSE | | |
|---|---|---|---|---|---|---|---|---|---|
| | Training | Test | validation | Training | Test | validation | Training | Test | validation |
| MLP 9-8-1 | 0.988 | 0.958 | 0.987 | 0.98 | 0.89 | 0.96 | 157.40 | 216.92 | 243.51 |
| MLP 9-12-1 | 0.995 | 0.971 | 0.987 | 0.99 | 0.92 | 0.94 | 105.68 | 177.73 | 296.11 |
| MLP 9-5-1 | 0.990 | 0.964 | 0.961 | 0.98 | 0.89 | 0.89 | 143.06 | 218.90 | 415.67 |
| MLP 9-13-1 | 0.990 | 0.976 | 0.956 | 0.98 | 0.91 | 0.87 | 143.51 | 197.79 | 434.12 |
| MLP 9-9-1 | 0.995 | 0.976 | 0.984 | 0.99 | 0.95 | 0.94 | 99.10 | 140.42 | 289.77 |
| MLP 9-6-1 | 0.987 | 0.960 | 0.998 | 0.97 | 0.86 | 0.95 | 164.76 | 239.43 | 281.85 |
| MLP 9-3-1 | 0.985 | 0.958 | 0.991 | 0.97 | 0.89 | 0.96 | 174.341 | 217.907 | 233.903 |
| MLP 9-4-1 | 0.987 | 0.959 | 0.988 | 0.97 | 0.89 | 0.95 | 160.145 | 216.306 | 272.020 |
| MLP 9-7-1 | 0.994 | 0.967 | 0.992 | 0.99 | 0.91 | 0.98 | 106.85 | 198.73 | 192.68 |
| MLP 9-10-1 | 0.989 | 0.958 | 0.992 | 0.98 | 0.86 | 0.97 | 151.26 | 244.59 | 206.28 |
| MLP 9-11-1 | 0.995 | 0.968 | 0.956 | 0.99 | 0.84 | 0.88 | 105.42 | 257.32 | 416.45 |

**Table 3.** Performance parameters of the artificial neural network for predicting the TDS concentration, in training, testing, and validation phase (NRMSE, RMSEr and MAE)

| ANN Structure | NRMSE | | | RMSEr | | | MAE | | |
|---|---|---|---|---|---|---|---|---|---|
| | Training | Test | validation | Training | Test | validation | Training | Test | validation |
| MLP 9-8-1 | 0.038 | 0.053 | 0.059 | 0.120 | 0.117 | 0.185 | 129.55 | 127.66 | 163.79 |
| MLP 9-12-1 | 0.026 | 0.043 | 0.072 | 0.163 | 0.132 | 0.297 | 87.96 | 135.74 | 223.28 |
| MLP 9-5-1 | 0.035 | 0.053 | 0.101 | 0.203 | 0.182 | 0.470 | 122.80 | 146.14 | 359.83 |
| MLP 9-13-1 | 0.035 | 0.048 | 0.106 | 0.200 | 0.159 | 0.465 | 123.88 | 139.68 | 379.24 |
| MLP 9-9-1 | 0.024 | 0.034 | 0.071 | 0.131 | 0.100 | 0.317 | 81.08 | 111.47 | 211.77 |
| MLP 9-6-1 | 0.040 | 0.058 | 0.069 | 0.136 | 0.128 | 0.231 | 139.92 | 135.12 | 247.47 |
| MLP 9-3-1 | 0.043 | 0.053 | 0.057 | 0.135 | 0.121 | 0.260 | 145.46 | 131.61 | 160.64 |
| MLP 9-4-1 | 0.039 | 0.053 | 0.066 | 0.144 | 0.122 | 0.259 | 129.27 | 120.79 | 194.47 |
| MLP 9-7-1 | 0.026 | 0.048 | 0.047 | 0.114 | 0.122 | 0.151 | 82.96 | 137.50 | 158.78 |
| MLP 9-10-1 | 0.037 | 0.060 | 0.050 | 0.13 | 0.14 | 0.20 | 124.81 | 151.32 | 179.36 |
| MLP 9-11-1 | 0.026 | 0.063 | 0.102 | 0.11 | 0.17 | 0.37 | 85.97 | 183.38 | 319.05 |

**Table 4.** Performance parameters of the artificial neural network for predicting the EC concentration, in training, testing, and validation phase (R, Nash and RMSE).

| ANN Structure | R | | | Nash | | | RMSE | | |
|---|---|---|---|---|---|---|---|---|---|
| | Training | Test | validation | Training | Test | validation | Training | Test | validation |
| MLP 9-8-1 | 0.992 | 0.969 | 0.957 | 0.99 | 0.93 | 0.94 | 231.56 | 305.15 | 624.02 |
| MLP 9-12-1 | 0.995 | 0.966 | 0.982 | 0.99 | 0.94 | 0.96 | 184.47 | 297.39 | 497.78 |
| MLP 9-5-1 | 0.989 | 0.959 | 0.841 | 0.99 | 0.94 | 0.78 | 263.59 | 289.50 | 1175.03 |
| MLP 9-13-1 | 0.974 | 0.995 | 0.945 | 0.97 | 0.99 | 0.91 | 405.86 | 112.65 | 738.04 |
| MLP 9-9-1 | 0.993 | 0.974 | 0.947 | 0.99 | 0.96 | 0.92 | 218.53 | 236.98 | 696.49 |
| MLP 9-6-1 | 0.994 | 0.982 | 1.000 | 0.99 | 0.97 | 0.96 | 199.28 | 214.24 | 518.88 |
| MLP 9-3-1 | 0.975 | 0.985 | 0.926 | 0.97 | 0.97 | 0.86 | 403.55 | 192.35 | 926.96 |
| MLP 9-4-1 | 0.969 | 0.958 | 0.966 | 0.95 | 0.92 | 0.95 | 472.70 | 341.77 | 553.00 |
| MLP 9-7-1 | 0.994 | 0.977 | 0.990 | 0.99 | 0.95 | 0.98 | 201.78 | 267.32 | 384.79 |
| MLP 9-10-1 | 0.994 | 0.979 | 0.973 | 0.99 | 0.96 | 0.94 | 200.88 | 228.50 | 612.46 |
| MLP 9-11-1 | 0.992 | 0.967 | 0.922 | 0.99 | 0.94 | 0.89 | 229.55 | 296.54 | 843.20 |

**Table 5.** Performance parameters of the artificial neural network for predicting the EC concentration, in training, testing, and validation phase (NRMSE, RMSEr and MAE)

| ANN Structure | NRMSE | | | RMSEr | | | MAE | | |
|---|---|---|---|---|---|---|---|---|---|
| | Training | Test | validation | Training | Test | validation | Training | Test | validation |
| MLP 9-8-1 | 0.032 | 0.042 | 0.085 | 0.161 | 0.138 | 0.228 | 193.69 | 260.84 | 508.66 |
| MLP 9-12-1 | 0.025 | 0.041 | 0.068 | 0.111 | 0.175 | 0.202 | 141.37 | 200.30 | 432.47 |
| MLP 9-5-1 | 0.036 | 0.039 | 0.160 | 0.191 | 0.169 | 0.320 | 218.59 | 260.20 | 864.29 |
| MLP 9-13-1 | 0.055 | 0.015 | 0.101 | 0.205 | 0.093 | 0.250 | 295.65 | 98.12 | 556.85 |
| MLP 9-9-1 | 0.030 | 0.032 | 0.095 | 0.125 | 0.129 | 0.253 | 174.31 | 181.78 | 563.19 |
| MLP 9-6-1 | 0.027 | 0.029 | 0.071 | 0.131 | 0.105 | 0.190 | 148.90 | 180.91 | 393.13 |
| MLP 9-3-1 | 0.055 | 0.026 | 0.126 | 0.183 | 0.104 | 0.232 | 329.40 | 154.97 | 693.82 |
| MLP 9-4-1 | 0.064 | 0.047 | 0.075 | 0.228 | 0.201 | 0.375 | 377.33 | 317.11 | 531.46 |
| MLP 9-7-1 | 0.028 | 0.036 | 0.052 | 0.20 | 0.18 | 0.19 | 159.03 | 228.19 | 332.24 |
| MLP 9-10-1 | 0.027 | 0.031 | 0.084 | 0.15 | 0.13 | 0.30 | 154.09 | 175.58 | 546.17 |
| MLP 9-11-1 | 0.031 | 0.040 | 0.115 | 0.19 | 0.16 | 0.30 | 186.86 | 216.21 | 662.85 |

*D. Electrical conductivity (EC) model*

Twelve architectures ANN were used to predict the conductivity in the Zahrez groundwater. The performance parameters of the training, test and validation sets are shown in Table 3. The selected ANN (MLP-9-12-1) provided a best fit model for the training and test data sets. The constructed ANN model (EC) was trained using the BFGS quasi Newton Algorithm (BFGS 55). For the best MLP network model A non-linear transfer function (Tanh) was used in the hidden layer and a non-linear transfer function (Logistic) in the output layer.

The respective values of RMSE, NRMSE, and MAE for the two data sets are 184.47, 0.025 and 141.37 for training, 297.39, 0.041 and 200.30 for testing. The correlation coefficients (R) for the training, test and validation sets were 0.995, 0.966 and 0.982, respectively. The respective values of NASH for the training and testing sets were, 0.993 and 0.935 respectively, suggest for a good-fit of the selected EC model to the data set. The scatter plot of observed versus modeled values of EC are shown in Figure. 4 and 5. The coefficient of correlation (R) values for the training, test, and validation sets were 0.995, 0.966 and, 0.982, respectively, suggesting a good-fit of the EC model (MLP-9-12-1) to the data set.

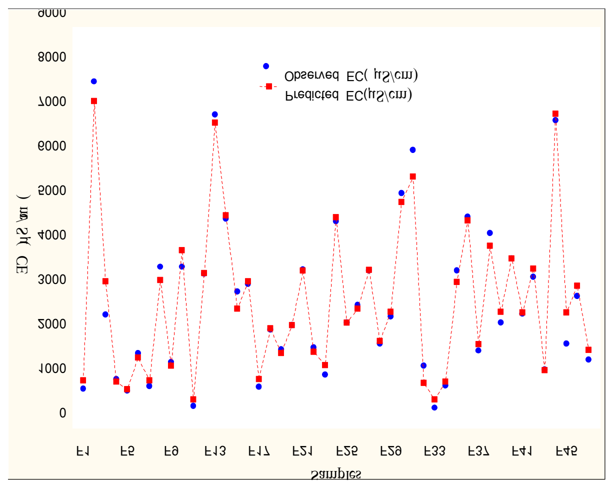

**Fig 2.** Measured and predicted TDS concentrations by MLP (9, 9, 1) model in training, testing, and validation phase.

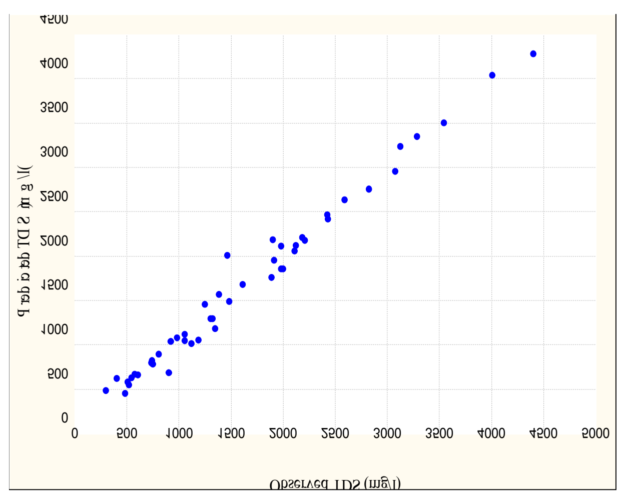

**Fig. 3.** Scatter diagram of the predicted values versus measured values for the training, testing, and validation data sets

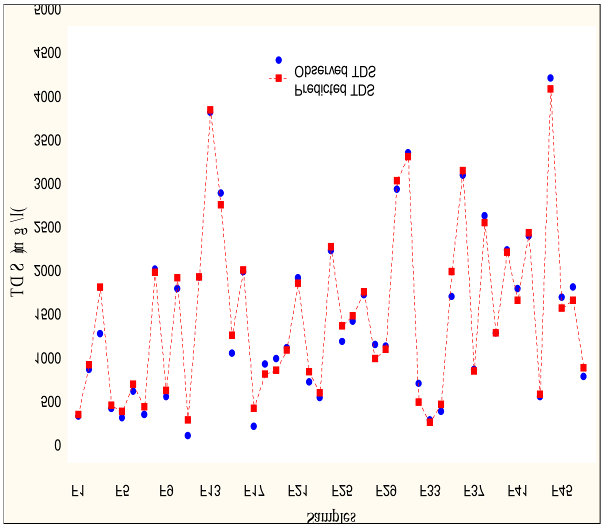

**Fig 4.** Measured and predicted EC by MLP (9, 12, 1) model in training, testing, and validation phase.

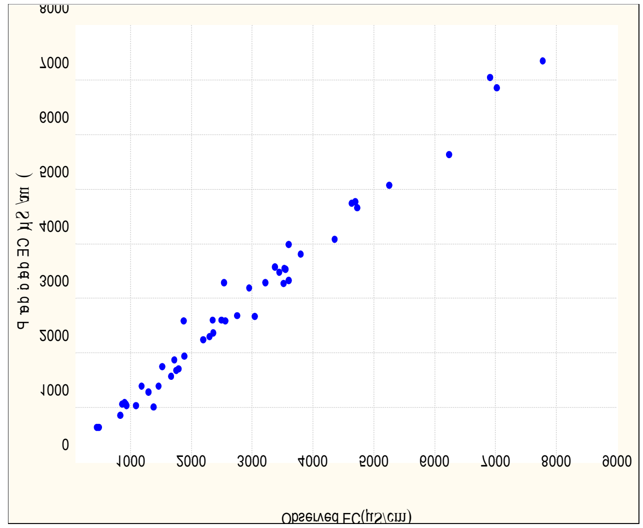

**Fig 5.** Scatter diagram of the predicted values versus measured values for the training, testing, and validation data sets.

## IV. CONCLUSION

In this study artificial neural network (ANNs) was developed to predict total dissolved solid (TDS) and electrical conductivity (EC) in groundwater of the Zahrez basin. The results indicate that, the ANN- MLP (9, 9, 1) model provided a best accuracy for prediction of the TDS concentration. It is found that the coefficient of correlation (R) values for the training, testing, and validation sets were 0.995, 0.976, and 0.984, respectively, the respective values of RMSE, NRMSE, and MAE for the two data sets are 99.10, 0.024 and 81.08 for training, and 140.42, 0.034 and 111.47 for testing, and 289.77, 0.071 and 211.77 for validation. The results of the predictive ANN models of electrical conductivity (EC) showed that the ANN- MLP (9, 12, 1) model provides the best accuracy , with the coefficient of correlation (R) of 0.995, 0.966 and, 0.982 for the training, test and validation sets, respectively. The respective values of RMSE, NRMSE, and MAE for the two data sets are 184.47, 0.025 and 141.37 for training, 297.39, 0.041 and 200.30 for testing, and 497.78, 0.068, and 432.47 for validation. Finally, from the results obtained, an ANN model appears to be a useful tool for prediction of the groundwater quality parameter in the Zahrez basin.

## ACKNOWLEDGEMENTS

The authors wish to thank the anonymous reviewers for their fruitful contribution in the improvement of the manuscript, through their revisions, suggestions and critical comments.

## REFERENCES

[1] Sidi Moussa, M. F. 2000. “Ressources Hydrauliques de la zone du projet GTZ-HCDS.” Coopération Algero-Allmande. 176 P.
[2] Kettab. A, and Metaiche. M. 2005 “Water desalination price from recent performances: Modelling, simulation and analysis.” *International Journal of Nuclear Desalination*
[3] Kettab. A, Bessenasse. M, and A.S. Moulla. (2010) “Seawater desalination: Study of three coastal stations in Algiers region.” *Desalination*
[4] Kettab. A. (2001). “Water resources in Algeria: strategies, investments, and vision.” *Desalination* 136 (1-3): 25-33
[5] Kettab. A, Mitiche. R, and Bennaçar. N. (2008). “Water for a sustainable development: challenges and strategies.” *Revue des sciences de l'eau* 21 (2).
[6] Wen. C.G, and Lee. C.S. (1998). A neural network approach to multiobjective optimization for water quality management in a river basin. *Water Resour Res* 34 (3):427–436.
[7] Lobbrect. A.H, and Solomatine. D.P. (1999). “Control of water levels in polder areas using neural networks and fuzzy adaptive systems.” *Water Industry Systems Modeling and Optimization Applications* vol. 1: 509–518.
[8] Niroobakhsh. M.S.H, Musavi-Jahromi. S.H, Manshouri. M, and Sedghi. S. (2012). “Prediction of water quality parameter in Jajrood River basin: Application of multi-layer perceptron (MLP) perceptron and radial basis function networks of artificial neural networks (ANNs).” *African Journal of Agricultural Research* 7 (29): 4131–4139.
[9] Singh. K.P, Basant. A, Malik. A, and Jain. G. (2009). “Artificial neural network modeling of the river water quality—a case study.” *Ecological Modelling. 220, 888–895.*

[10] ANRH. 2009. Agency for water Resources
[11] Belkhiri. L, Boudoukha. A, Mouni. L, and Baouz T. (2011). “Statistical categorization geochemical modeling of groundwater in Ain Azel plain (Algeria).” *Journal of African Earth Sciences* 59:140–148
[12] Haykin. S. (2009). “Neural networks and learning machine.”, 3rd ed, edited by Pearson
[13] Gallant. S.I. (1993). “Neural Network Learning and Expert Systems.”, edited by MIT press, Massachusetts, USA
[14] Smith. M. (1994). “Neural Networks for Statistical Modelling.”, edited by Van Nostrand Reinhold, NY
[15] Dogan. E, Sengorur. B, and Koklu. R. (2009). “Modeling biochemical oxygen demand of the Melen River in Turkey using an artificial neural network technique.” *Journal Environmental Management* 90 (2): 1229–1235.
[16] Maier. H.R, Jain. A, Dandy. G.C, and Sudheer. K.P. (2010). “Methods used for the development of neural networks for the prediction of water resource variables in river systems: cur-rent status and future directions.” Environmental Modelling & Software *25* (8): 891–90 9.
[17] Palani. S, Liong. S.Y, Tkalich. P. (2008)“An ANN application for water quality forecasting.”*Marine Pollution. Bulletin*. 56 (9): 1586–1597.
[18] Zhao. Y, Nan. J, Cui. F.Y, and Guo. L. (2007). “Water quality forecast through application of BP neural network at Yuqiao reservoir.”*Journal Zhejiang University Science* 8 (9): 1482–14 87.
[19] Dawson. C.W, and Wilby. R.L. (2001). “Hydrological modelling using artificial neural networks.” *Progress in Physical Geography* 25 (1): 80–108.
[20] Govindaraju. R.S. (2000). “Artificial neural networks in hydrology. I, preliminary concepts.” *Journal of Hydrologic Engineering* 5 (2): 115–123.
[21] Maier. H.R, and Dandy. G.C. (2000). “Neural networks for the prediction and forecasting of water resources variables: a review of

modeling issues and applications." *Environmental Modeling and Software* 15 (1): 101–124.
[22] Rumelhart. D.E, and McClelland. J.L. (1986). "Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Foundations.", edited by MIT Press, Cambridge, Mass
[23] Ouarda. T.B.M.J, and Shu. C. (2009). "Regional low-flow frequency analysis using single and ensemble artificial neural networks." *Water Resources Research.* 45: W11428.
[24] Shu. C, Ouarda. T.B.M.J. (2007). "Flood frequency analysis at ungauged sites using artificial neural networks in canonical correlation analysis physiographic space." *Water Resour. Res.* 43: W07438
[25] Haykin. S. (1994). "Neural Networks ." edited by Macmillan College Publishing, New York
[26] Dekker. SC, Bouten. W, and Schaap. MG. (2001). "Analyzing forest transpiration model errors with artificial neural networks." *Journal of Hydrology* 246:197–208
[27] Rumelhart. DE, Durbin. R, Golden. R, and Chauvin. Y. (1995) "Backpropagation: the basic theory." *In: Chauvin Y, Rumelhart DE (eds) Backpropagation: theory, architectures, and applications, edited by* Erlbaum, Hillsdale, NJ
[28] Shamseldin. AY.(1997) "Application of a neural technique to rainfall-runoff modeling." *Journal of Hydrology* 199:272–294
[29] Fletcher. D, and Goss. E. (1993) "Forecasting with neural networks: an application using bankruptcy data." *Information Management* 24 (3): 159–167.

Artificial Neural Network Modeling of the Groundwater Quality (Case Study: Zahrez Basin in Algeria)