Moroccan Journal of Chemistry
ISSN: 2351-812X
http://revues.imist.ma/?journal=morjchem&page=login
Ousmane & al. / Mor. J. Chem. 6 N°2 (2018) 245-255

Development of a Neural Network approach for Predicting nitrate and sulfate concentration in three lakes: Ifrah, Iffer and Afourgagh, Middle Atlas Morocco

H. Ousmana [a], A. El Hmaidi [a]*, M. Berrada [b], B. Damnati [c], I. Etabaai [c], A. Essahlaoui [a]

[a] Team *of water sciences and environmental engineering, Department of Geology, Faculty of Sciences, University of Moulay Ismail, B.P. 11201, Zitoune, Meknes, Morocco.*
[b] *Laboratory of Mathematical and Computational Modeling (LM2I), ENSAM of Meknes, University of Moulay Ismail, Marjane II, Meknes, B.P 15290 AL Mansour, 50000, Meknes, Morocco.*
[c] *LEORN, Department of Earth Sciences, Faculty of Sciences and Techniques, B.P 416, Tangier 90 000, Morocco*.

** Corresponding author:*
elhmaidi@yahoo.fr
Received 11 April 2017,
Revised 26 Oct 2017,
Accepted 18 Jan 2013

Abstract

Neural networks are mathematical and computer models to power nonlinear data that play a very important role in various scientific fields. They are specially used for automatic resolution of environmental problems. This study focuses on the prediction of nitrate (NO_3^-) and sulfate (SO_4^{2-}) of lake water in the Moroccan Middle Atlas. Ifrah, Iffer and Afourgagh are taken as case studies by using a number of parameters physic-chemical of water. Two methods were used: Multiple Linear Regression (MLR) and Artificial Neural Networks (ANN) Multilayer Perceptron Model (MLP). In order to choose the best neural network architecture, several statistical tests were used in conjunction with some robustness tests: Mean Square Error (MSE), Mean Absolute Error (MAE) and correlation coefficient (R). The results showed that the models established by artificial neural network Multilayer Perceptron type (ANN-MLP) of configuration [17-8-2] are more efficient compared to those determined by the conventional method based on multiple linear regression. This performance demonstrates the existence of a nonlinear relationship between the physic-chemical characteristics of both nitrates and sulfates in the lakes waters studied that are under investigation.

Keywords: *Neural Networks, Multiple linear regression, Prediction, Middle Atlas, Physic-Chemical, Lake waters.*

1. Introduction

The lakes are aquatic systems in continuous exchange with the external environment on which they rely on. They are also systems of transfer, storing and recycling water, the material (organic and inorganic) and energy (kinetic, thermal, light and potential). Such a feature of these ecosystems makes them very sensitive to environmental and climate change [1]. The region where the study sites is located, commonly called Dayates region because of the juxtaposition of several natural lakes of tectonic-karst origin [2]. The lakes Ifrah, Iffer and Afourgagh are located in the Causse of Middle Atlas (Fig.1). Artificial neural networks (ANN) are an approximation method of complicated systems, especially useful when these systems are difficult to model using classical statistical methods. The RNA provide interesting results due to their ability of learning [3], their parallelism and their ability to solve many problems nonlinear systems [4].

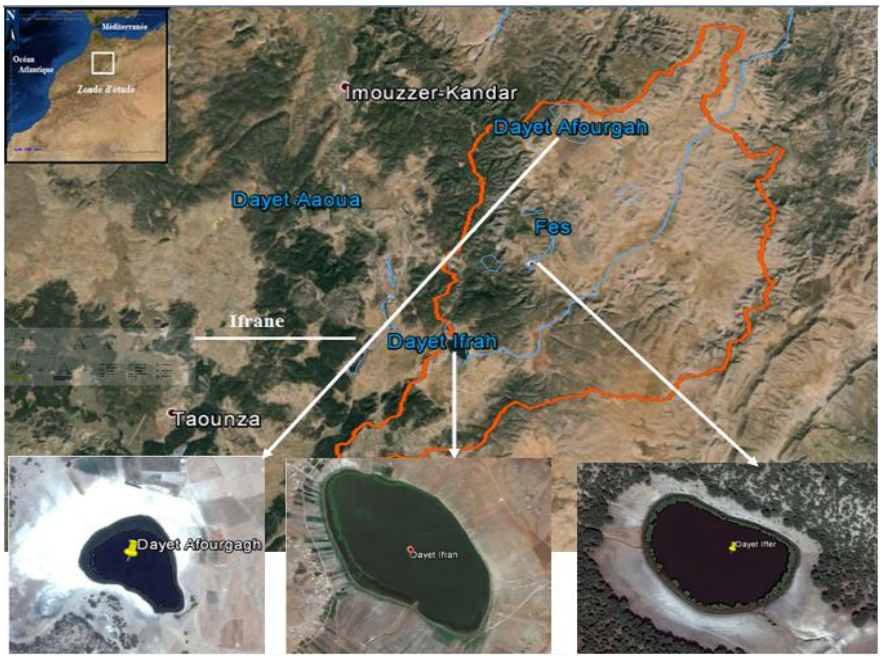

Figure 1. Localisation of lakes Ifrah, Iffer and Afourgagh (Source: Google Earth).

Over the past decade, ANN research has applied into the areas of hydrology, ecology and of environment. Many researchers showed that the ANN model gives a better performance compared to the other model in forecasting water quality. Clementking and Jothi Venkateswaran [5], have made study on an application of Back Propagation Neural Network (BPNN) for water quality attributes variations. Heydari and al [6], have studied the use of a Neural Network technique for the prediction of water quality parameters. Banejad and al [7], made a study to predict Water Quality Indexes Artificial Neural Network. Panda and al [8], employed an ANN approach for estimation water quality in lake using satellite imagery. Abyaneh [9], has reviewed the evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. Diamantopoulou and al [10], have investigated on water quality Parameter s of Axios River in Northern Greece using ANN. Suen and al [11], carried out a study on modeling nitrate concentrations in rivers with ANNs. Beucher and al [12] have done a study on mapping and characterization of acid sulfate soils: Application to Sirppujoki River catchment, southwestern Finland. The main aim of this research is the application of neural network method for the prediction of nitrate (NO_3^-) and sulfate (SO_4^{2-}) from three middle Atlas lakes (central Morocco).

2. Materials and methods

2.1. Presentation of Study Area

The sites that are selected for this work are Ifrah, Iffer and Afourgagh lakes which are located in the north of the Moroccan Central Middle Atlas (tabular Middle Atlas), about 180 km from the Atlantic coast to the East and the Mediterranean to the north (Fig. 2). Ifrah Lake (33 ° 33'N, 04 ° 56'W) is the largest in the region. The watershed of Lake Ifrah is relatively large. It covers an area of 45.77 km² with a perimeter of 32.9 km. The altitude of this area is between 1620 m and 1940 m [13]. It corresponds to a collapse caused by a series of longitudinal and transverse faults. It is mainly fed by direct precipitations (snow and rain), the emergence of water table during wet years as well as by the temporary flow during the wet season. Iffer lake (33° 36'30 "N, 4° 54'30" W) is the smallest lake in the Middle Atlas. It occupies a deep depression produced by tectonic-karst collapse. This lake has a sub-circular shape with steep edges [14], its watershed rises to 1772 m sea level and covers 388 ha. The third site under investigation is Afourgagh lake (33° 36'N, 04° 52'W). It is considered the greatest depression of the region. Its level has decreased over the past three decades [15]. The area of this lake is 6 ha. The watershed is 49 km2 with a perimeter of 37 km. This surface is divided between 1357 m and 1830 m of altitude. The geology of the region is calcareous dolomite attributed to the Lower and Middle Lias, overlying Triassic mudstones [16]. Smaller outcrops of Eocene and Miocene limestone appear mainly in the catchment of Lake Ifrah. The Quaternary formations are very small and are alluvial fans and fluvial lacustrine deposits of alluvial and colluvial slopes, oriented towards the low lake depressions.

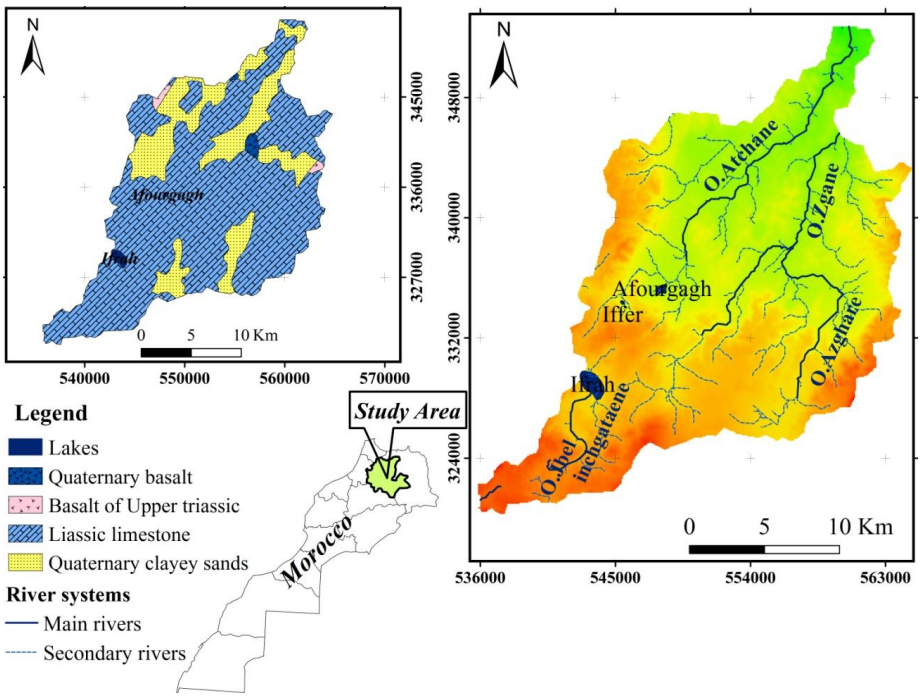

Figure 2: Location MNT and geology of the three lakes: Iffer, Ifrah and Afourgagh.

2.2. Database

The database used in this study based on the work of Etebaai [15].The database contains sixty six samples and seventeen of physic-chemical parameters which are obtained from the three lakes (Iffer, Ifrah and Afourgagh) between February and November of 2006. The database consists of 66 samples. The independent variables (explanatory) are seventeen physic-chemical parameters: pH, Temperature (T), Electrical conductivity (C.E), Chloride (Cl^-), Bicarbonate (HCO_3^-), Potassium (k^+), Turbidity (NTU), Total Hardness (THf), Sodium (Na^+), Magnesium (Mg^{2+}),

Calcium (Ca^{2+}), Suspended matter (MSE), Saturation index of calcite, Saturation Index of aragonite (IsArg), Saturation index of dolomite (IsDol), Salinity (TDS) and Dissolved oxygen (O_2). The dependent variables (to explain) are nitrates (NO_3^-) and Sulfate (SO_4^{2-}).

2.3. Artificial Neural Network (ANN) Models MLP

The Artificial neural networks provide an alternative to mathematical modeling and they are among nonparametric statistics and nonlinear models which are able to respond to issues of aid to the decision, diagnosis, prediction, etc. [17, 18]. The goal is to introduce to the input and output data of the ANN and make it learns the relationship between them by a process called learning [19]. This is to minimize the error between the model output and the desired output by adjustment of the model parameters (weights). The learning process is usually done in five steps (Fig.3):

- Weighting the entries by parameters called weight (W);
- Summation of weighted inputs A_j.

$$A_j = \sum_{i=1}^{n} w_{ij} I_i \qquad \text{Eq. (1)}$$

Where w_{ij} the synaptic weight and I_i the input values.

- 3- Calculation of response the neuron by activation function f;

$$S_j = f\left(\sum_{i=1}^{n} w_{ij} I_i + b_j \right) \qquad \text{Eq. (2)}$$

Where b_j is the bias of neuron j.

- Calculation of the error between the theoretical output and that calculated by ANN;

$$E = \frac{1}{N} \sum_{i=1}^{n} (Y_{ri} - Y_{di})^2 \qquad \text{Eq. (3)}$$

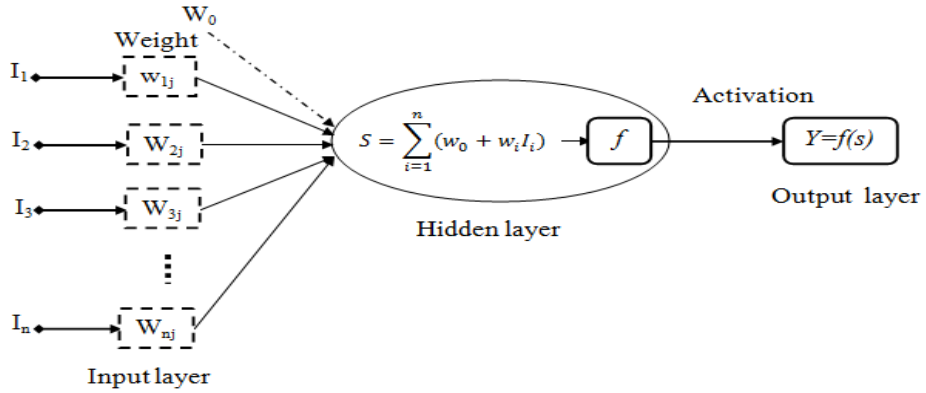

Figure 3: Learning process and functioning of the ANN.

- Modification the weights to minimize the error by a specific mathematical algorithm named learning algorithm.

2.4. Normalization and data pre-processing

The primary purpose of data transformation is to modify the distribution of input variables so that they can better match outputs. Before training and validation, we scaled the inputs and targets using a normalized equation such that the data always fall within a specified range:

$$Y_N = (y_{max} - y_{min}) \left(\frac{x_i - x_{min}}{x_{max} - x_{min}} \right) - y_{min} \qquad \text{Eq. (4)}$$

Where Y_N is the data value after normalization, x_{max} and x_{min} denote the Maximum and the minimum of the data respectively; y_{max} and y_{min} are taken as -1 and 1; x_i the Original values.

2.5. Modeling Performance Criteria

In the present study, three different criteria are used in order to evaluate the effectiveness of each network and its ability to make precise prediction [20, 21 and 22]. The coefficient of correlation (R), calculated by

$$R = \frac{\sum_{i=1}^{N}\left(x(i)-\bar{x}(i)\right)\left(y(i)-\bar{y}(i)\right)}{\sqrt{\left(x(i)-\bar{x}(i)\right)^2\left(y(i)-\bar{y}(i)\right)^2}} \qquad \text{Eq. (5)}$$

Also, the Mean Square Error (MSE), given by

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(NS_{Obs}(i) - NS_{Pred}(i)\right)^2 \qquad \text{Eq. (6)}$$

The Mean Absolute Error (MAE), determining by

$$MAE = \sum_{i=1}^{N}\left|NS_{Obs}(i) - NS_{Pred}(i)\right| \qquad \text{Eq. (7)}$$

Where $NS_{Obs}(i)$ and $NS_{Pred}(i)$ are the Observed and predicted values, respectively; $y(i)$ and $\bar{y}(i)$ are absolute and average predicted values; $x(i)$ and $\bar{x}(i)$ are absolute and average of observed values, respectively.

3. Results and discussion

3.1. Selection and development of models of artificial neural network MLP

Eleven training algorithms were examined, namely train GD, train GDA, train GDX, train RP, train CGF, train CGP, train CGB, train SCG, train BFG, train OSS and train LM with different numbers of neurons in the hidden layer and different transfer functions. The optimal artificial neural network used in this work is the Multi-Layer Perceptron (MLP) based on the Levenberg-Marquart algorithm (LM) with a hyperbolic tangent transfer function (Tansig) (Eq. (8)) at the hidden layer and a linear transfer function (Purelin) (Eq. (9)) at output layer [23, 24].

$$f(x) = \frac{2}{\left(1+e^{-2x}\right)-1} \qquad \text{Eq. (8)}$$

$$f(x) = x \qquad \text{Eq. (9)}$$

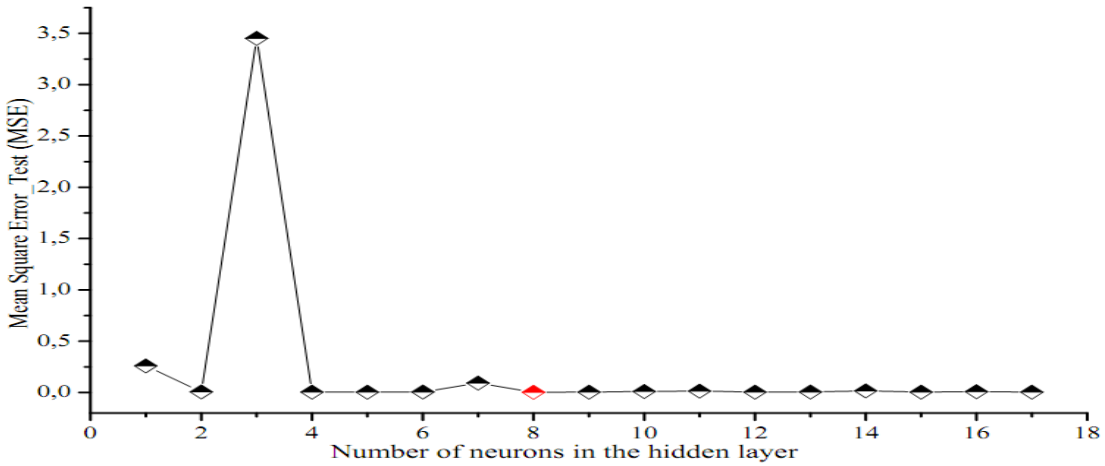

Figure 4: Robustness test: Mean Square Error.

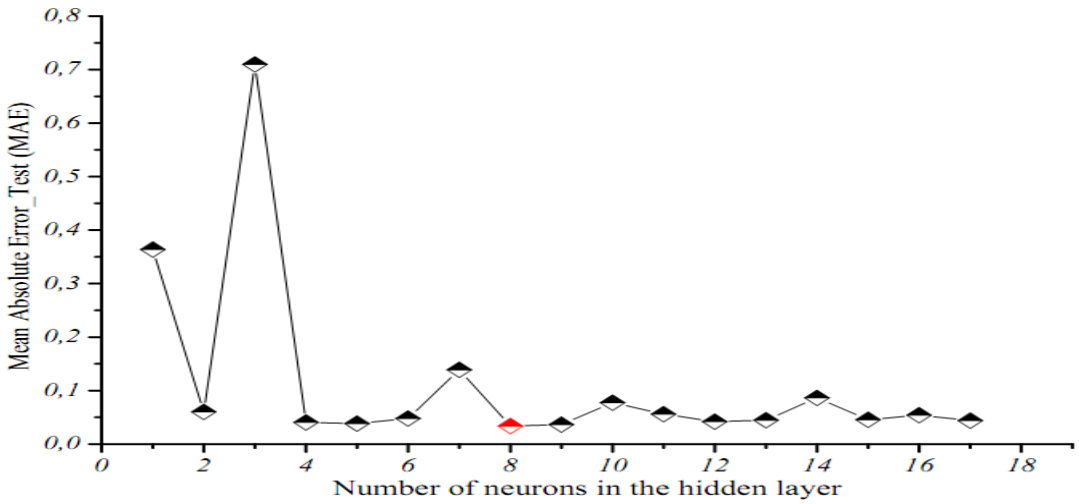

Figure 5: Robustness test: Mean Absolute Error.

The choice of the artificial neural network is based on the ease and the speed of construction of its function [25] and on the fact that our problem has a limited number of input variables. According to the figures 4 and 5 of MSE and MAE, the error stabilizes even with the increased number of neurons. This is reflected on the network by a relatively stable prediction from 8 neurons in the hidden layer. From these statistical indicators of robustness, the best values are obtained when the number of neurons in the hidden layer NNC = 8. The best ANN architecture was chosen with eight numbers of neurons in hidden layer. The best structure is based on the lowest error (MSE and MAE) and highest correlation (R). Moreover, the architecture [17-8-2] by taking into consideration its good predictive ability. This configuration has three layers: an input layer containing seventeen neurons, a hidden layer containing eight neurons and an output layer comprising two neurons.

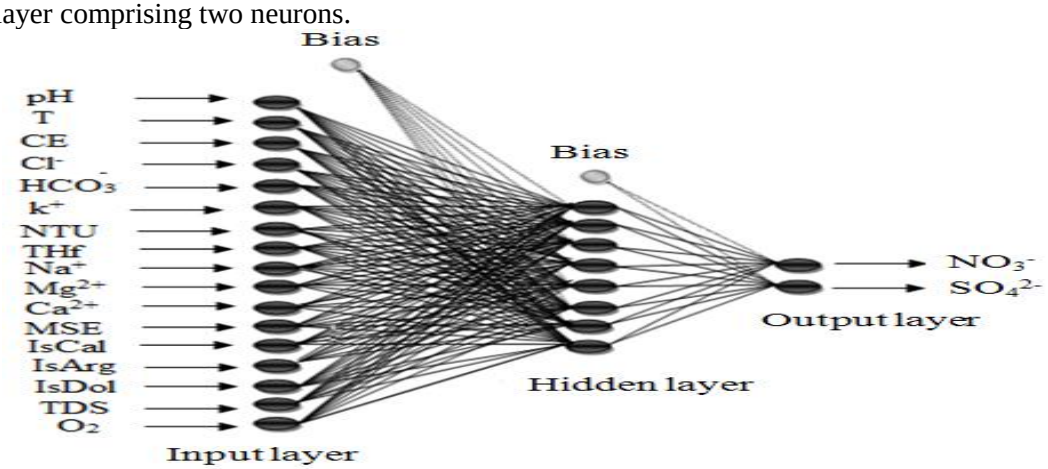

Figure 6: Architecture of the neural network with three layers configuration [17-8-2].

3.2. Learning and validation

The network development followed several steps:

- Step 1: Setting of the number of hidden layers, neurons, training algorithm (backpropagation in our case), initial connection weights, and neuron biases and the activation function for each neuron.
- Step 2: Network training and validation. Sixty per cent of the time, samples introduced to the ANN and the connection weights and biases values were determined within the network. The next 20 per cent was used for validation.

- Step 3: Estimation of the predicted output. The input values for the last 20 per cent of the observations were then used by the trained ANN to generate output.
- Step 4: Evaluation of the forecast performance of the ANN, and comparison of it to other models.
- Step 5: Steps 1–4 were repeated if the error goal was not reached.

This paper determined that, in the case of predicting the sulfates, the correlation between the measured values and those predicted by the model is very good with a correlation coefficient R = 0.9987 for the learning base and R = 0.9983 for the test (Fig. 7), which is equivalent to a mean square error of 0.00219 and 0.03378 for MAE. Also, In the case of predicting the nitrates, the correlation between the observed values and those predicted by the model is also very good with a correlation coefficient R = 0.9998 for the learning base and R = 0.9952 for the test (Fig. 8), which is equivalent to a mean square error of 0.0022 and 0.03378 for MAE.

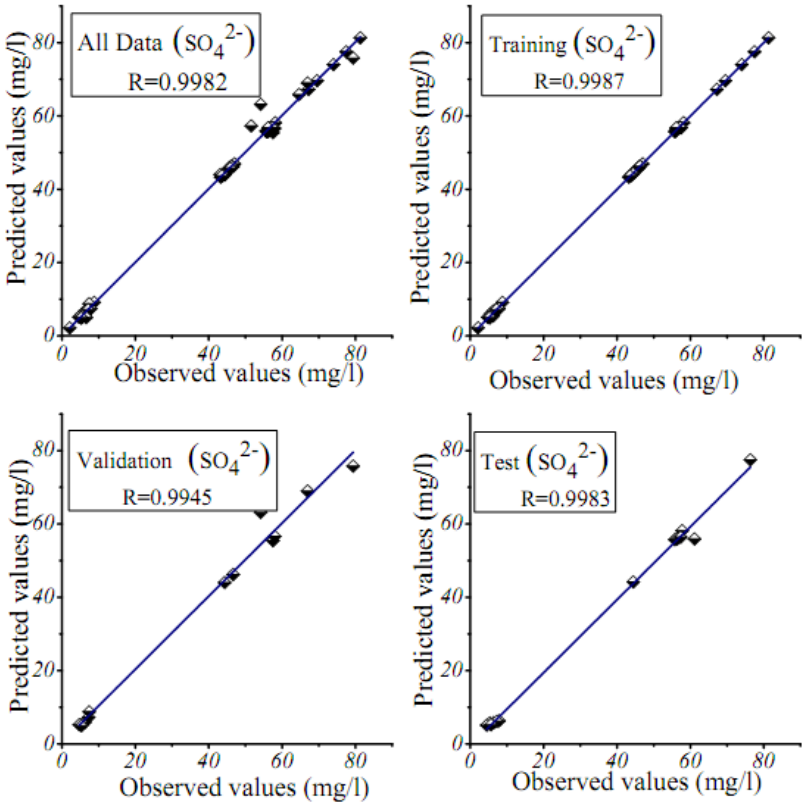

Figure 7: Observed and predicted sulphates using MLP-ANN model.

These results highlight the generalization capabilities of the ANN model adopted. This model predicts the concentrations of sulfates and nitrates in the waters of the three lakes studied under investigation with acceptable accuracy. Fig.9 to fig.11 confirms again the performance of the ANN configuration [17-8-2] with a very good superposition of the curves experimentally measured values and those estimated by the model ANN as well for the nitrates than for sulfates. The results obtained show a very good agreement explained by a high correlation coefficient and the two (MAE and MSE) low statistical indicators for the learning phase, the test phase and the validation phase. This indicates that these results are very satisfactory for nitrates and sulfates

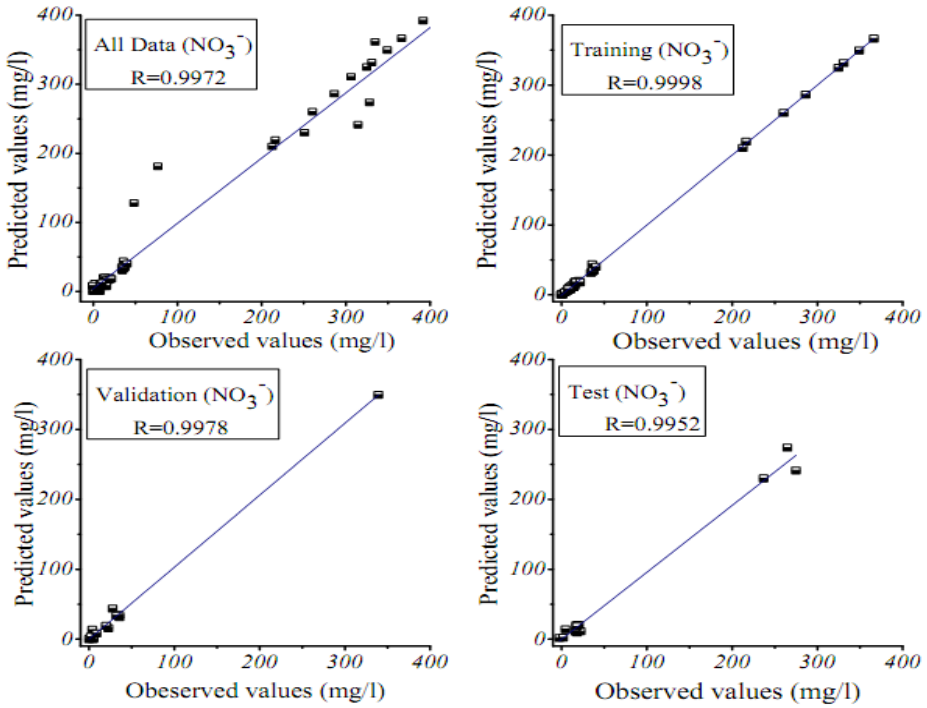

Figure 8: Observed and predicted nitrates using MLP-ANN model.

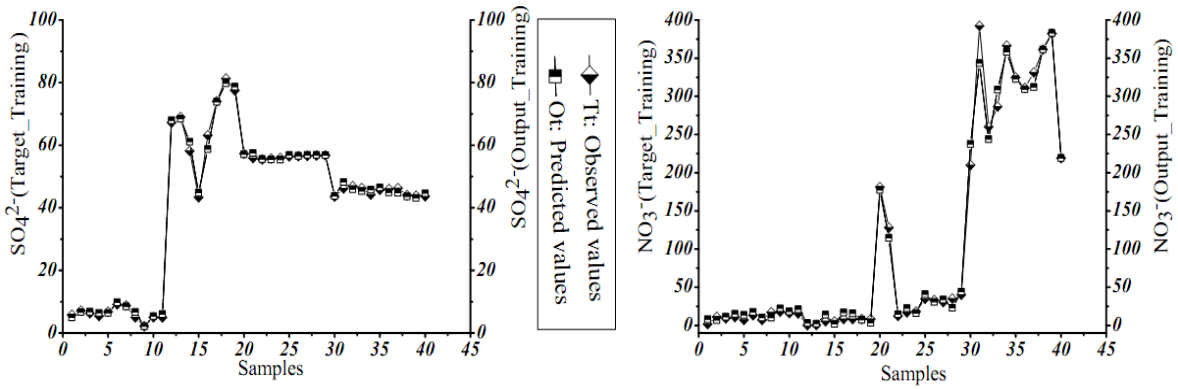

Figure 9: The relationship between the observed values and the predicted values of sulfates and nitrates by ANN-MLP: Training data set.

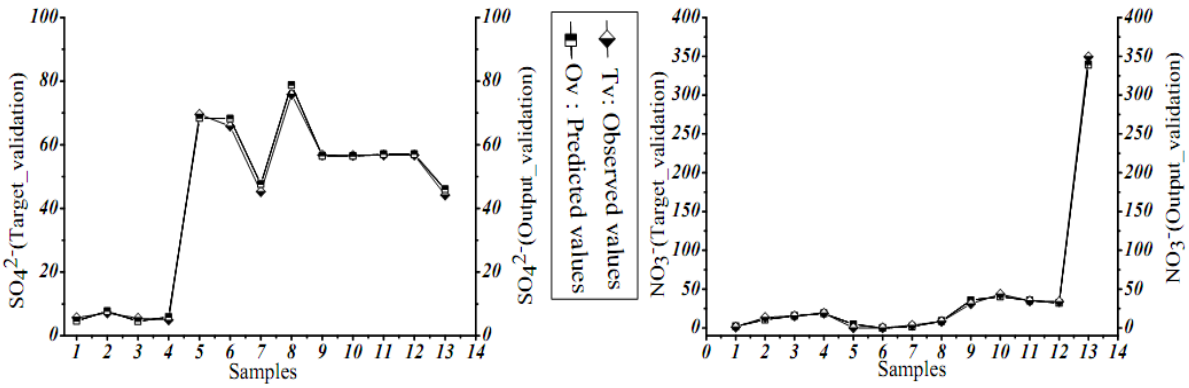

Figure 10: The relationship between the observed values and the predicted values of sulfates and nitrates by ANN-MLP: validation data set.

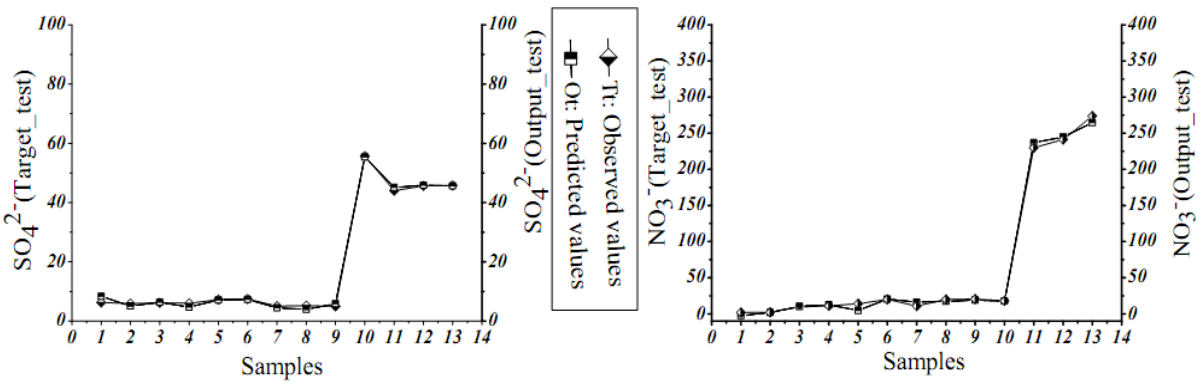

Figure 11: The relationship between the observed values and the predicted values of sulfates and nitrates by ANN-MLP: testing data set.

3.3. Comparison of MLR and ANN for predicting concentrations in nitrates and in sulfates

To evaluate the performance of the model established MLP-ANN [17-8-2], a comparison with other more classic models essentially the model of Multiple Linear Regressions (MLR) was used. The results depicted in Table 1 show the comparison of correlation coefficients and statistical indicators obtained by two models of the prediction, ANN and MLR. On the one hand the correlation coefficient calculated by the MLP-ANN [17-8-2] is significantly higher with (R = 99.875%), however, the correlation coefficients calculated by the MLR are more or less low (R=85.55%). On the other hand, the mean square error established by ANN are very low in difference to those established by the MLR model (MSE= 3.965) for sulfates and (MSE = 741, 134) for nitrates, this is compatible with scientific results as [26] and [27].

Table 1: Correlation coefficients and statistical indicators obtained by MLR and MLP - ANN [17-8-2] for sulfates and nitrates.

Methods	**Sulfates (SO_4^{2-})**			**Nitrates (NO_3^-)**		
	R	MSE	AIC	R	MSE	AIC
MLR	0,875	3,965	104,59	0,851	741, 134	447,37
ANN	0,999	0,0021	12,017	0,989	0,0021	12,017

4. Conclusion

In the present study, the prediction of the concentrations in nitrates and in sulfates in the water of Ifrah, Iffer and Afourgagh lakes was obtained using artificial neural networks Multi-Layer Perceptron (MLP) with a supervised learning, involving algorithm "Levenberg-Marquardt " that gives better results in terms of rapidity, convergence and generating performance. The results showed a high capacity for learning and prediction for the concentrations of nitrates and the sulfates with a very high correlation coefficient of 99.98% and a very low mean square error (0.0022) for the database. In addition to that, it show a better choice of the network architecture [17-8-2] achieved by the application of statistical indicators of robustness. Moreover, for the Multiple Linear Regressions (MRL), the results are less important, with a correlation coefficient between 0.85 and 0.87 which corresponds to a mean square error of 3.965 for sulfates and 71.41 for nitrates. Furthermore, the forecasting by artificial neural networks shows a good correlation between the observed and estimated values which means that the ANN model has better predictive power compared to the classic model. This shows that these parameters are related to other physic-chemical parameters with a nonlinear relationship.

References

[1] I. Etebaai, B. Damnati, H. Benhardouze., O. Benhardouze, H. Reddad, M. Maâtouk, M. Taieb, *African Geosciences Review*, 15 (2008) 303-317.

[2] L. Hinaje, L. Ait Brahim, *Comunicações Do Instituto Geológico e Mineiro*, 89 (2002) 283-294.

[3] L. M. L. He, Z. L. He, *Water Research*, 42 (2008) 2563-2573.

[4] W. C. Liu, W. B. Chen, *Computers & Geosciences*, 45 (2012) 13-25.

[5] A. Clementking, C. J. Venkateswaran, Prediction of Water Quality Attributes Variations Using Back Propagation Neural Network (BPNN) Model. *International Conference on Technology and Business Management* (ICTBM-15), American Univrsity in the Emirates, (2015) 128-138.

[6] M. Heydari, E. Olyaie, H. Mohebzadeh, Ö. Kisi, *Middle-East Journal of Scientific Research*, 13 (2013) 1367-1376.

[7] H. Banejad, E. Olyaie, *Journal of American Science*, 7 (2011) 60-65.

[8] S. S. Panda, V. Garg, I. Chaubey, *Journal of Environmental Informatics*, 4 (2004) 65-74.

[9] H. Z. Abyaneh, *Journal of Environmental Health Science and Engineering*, 12 (2014) 1-8.

[10] M. J. Diamantopoulou, D. M. Papamichail, V. Z. Antonopoulos, *Operational Research*, 5 (2005) 115-125.

[11] J. P. Suen, J. W. Eheart, *Journal of water resources planning and management*, 129 (2003) 505-510.

[12] A. Beucher, R. Siemssen, S. Fröjdö, P. Österholm, A. Martinkauppi, P. Edén, southwestern Finland. *Geoderma*, 247 (2015) 38-50.

[13] B. Damnati, I. Etebaai, H. Reddad, H. Benhardouz, O. Benhardouz, H. Miche, M. Taieb, Recent environmental changes and human impact since mid-20th century in Mediterranean lakes: Ifrah, Iffer and Afourgagh, Middle Atlas Morocco. *Quaternary International*, 262 (2012) 44-55.

[14] A. Rhoujjati, Changements paléoenvironnementaux et paléoclimatiques depuis 21.000 ans 14C dans le Moyen Atlas marocain: les lacs Ifrah et Iffer. *Thèse de Doctorat d'état,* Université Chouaib Doukkali, Faculté des Sciences El Jadida, (2007) 206p.

[15] I. Etebaai, L'environnement actuel et le fonctionnement hydroclimatique de quelques systèmes lacustres dans le Moyen Atlas marocain : Cas des lacs Ifrah, Iffer et Afourgagh. *Thèse de Doctorat National*, Université Abdelmalek Essaadi, Faculté des Sciences et Techniques Tanger, (2009) 344p.

[16] G. Colo, Contribution à l'étude du Jurassique du Moyen Atlas septentrional. *Notes et Mémoires, Service Géologique*, Maroc 139 (1961) 226p.

[17] S. Huo, Z. He, J. Su, B. Xi, C. Zhu, *Procedia Environmental Sciences*, 18 (2013) 310-316.

[18] N. D. Kaushika, R. K. Tomar, S. C. Kaushik, *Solar Energy*, 103 (2014) 327-342.

[19] N. Samani, M. Gohari-Moghadam, A. A. Safavi, *Journal of Hydrology*, 340 (2007) 1-11.

[20] I. Manssouri, M. Manssouri, B. El Kihel, *Journal of Information, Intelligence and Knowledge*, 3 (2011) 201.

[21] I. Manssouri, A. El Hmaidi, T-E. Manssouri, B. El Moumni, *Journal of Computer Engineering*, 16 (2014) 117-123.

[22] O. R. Dolling, E. A. Varas, *Journal of Hydraulic Research*, 4 (2002) 547-554.

[23] M. Marzouk, M. Elkadi, *Journal of Cleaner Production*, 112 (2016) 4540-4549.

[24] M. Laïdi, S. Hanini, *Revue des Energies Renouvelables*, 15 (2012) 513-520.

[25] N. Sarani, J. Soltani, S. Sarani, A. Moasheri,. *International Journal of Biological, Ecological and Environmental Sciences*, 1 (2012) 130-134.

[26] H. Ghorbani, H. Kashi, N. Hafezi Moghadas, S. Emamgholizadeh, *Communications in Soil Science and Plant Analysis*, 46 (2015) 763-780.

[27] S. I. V. Sousa, F. G. Martins, M. C. M. Alvim-Ferraz, M. C. Pereira, *Environmental Modelling & Software*, 22 (2007) 97-103.