

DEVELOPMENT OF MATHEMATICAL MODELS FOR PREDICTING THE IRON CONCENTRATIONS OF LAKE OUBEIRA WATERS (NE ALGERIAN)

H. Alayat*¹, H. El Badaoui², A. Abdallaoui², D. Abrid², A. El Hmaidi²

¹University Chadli Bendjedid El Tarf, Laboratory agriculture and ecosystem functioning, PO Box 36000, Algeria

²Faculty of Sciences, Department of Chemistry, Analytical Chemistry and Environment Team, PO Box 11201, Zitoune Meknes, Morocco

Received: 29 April 2017 / Accepted: 21 October 2017 / Published online: 01 January 2018

ABSTRACT

Facing the increase of surface water samples contaminated by ETMs, usually from the geochemical background, the emergence of new human diseases is worrying. To solve this problem, we have developed several models based on different learning algorithms qualified by high performance, using different transfer functions. We have shown that all the Neural Models presented more or less important performance compared to the one based on multiple linear regressions.

The best revealed model ANN in the current work is a MLP type that uses the Levenberg-Marquardt algorithm as a learning algorithm, with Tansig and Purelin as transfer functions, respectively in the hidden layer and the output layer. This successful model can be considered as an important tool of great effectiveness in the context of environmental prediction and especially in anticipation of the iron contents of the Oubeira Lake water.

Keywords: Prediction, heavy metals, Linear multiple regression, artificial neural networks, Oubéira Lake.

Author Correspondence, e-mail: alayathacene@yahoo.fr

doi: <http://dx.doi.org/10.4314/jfas.v10i1.6>



I. INTRODUCTION

Since the last decade, the use of artificial neural networks (ANN) is being developed in many disciplines (environmental sciences, economics, ecology, biology, medicine, etc.). They are especially applied to solve problems of prediction [1], classification [2], of categorization, optimization, pattern recognition [3] and associative memory [4].

An artificial neural network (ANN) is qualified on data through a learning mechanism that acts on the network components to best achieve the expected task. The most widely used ANN family in recent years, as a decision support tool is the Multi-Layer Perceptron: MLP [5]. We cite as examples, a few relevant scientific investigations that showed the effectiveness of artificial neural networks basic models in environment forecasts:

Bouaoune and Dahmaniin [6] have presented an interpolation method application of neural networks. They showed nonlinear relationships between the variables.

Perez and Trier in [7] suggested predicting the NO₂ concentration and NO nitric oxide in Santiago; their results showed that the neural network is the method which performs the lower prediction error, compared to other conventional methods. Cheggaga and Ettoumien in [8] showed the feasibility and effectiveness of using neural networks to non-recurrent layers to predict the wind speed from the studied data.

In Algeria, the Mediterranean climate area, potable, industrial and agricultural water comes mainly from groundwater and incidentally by surface water. To meet the continuous growing demand, the surface water is increasingly required. However, this water usually contains ETM from the geochemical background and / or human impact. For a proper utilization of such surface water, a view on their contents is important.

Water and mud analyses of OubeiraLake, have revealed the presence of iron sometimes exceeding recommended concentrations. To monitor the evolution of iron responsible for the water quality degradation in the Oubeira Lake, it seemed important to have a predictive model most suited to the ecosystem modeling. The approach presented in this work to predict iron evolution according to the major elements (Ca²⁺, Mg²⁺, Na⁺, K⁺, Cl⁻, etc.) is the application of multiple regression for its flexibility and Artificial Neural networks (ANN) type Multilayer Perceptron to establish a better relationship between these variables. Networks artificial neurons are currently widely used for modeling nonlinear complexes and linear systems. The NAS implementation based on statistical learning is more tedious. Both methods require a database.

II. MATERIAL AND METHODS

II.1. Presentation of the study area

Oubeira Lake, extensive inland hydro system, with an area of 21.73 km² is located NE Algeria (36 ° 51' N - 8 ° 23' E), at an altitude of 23 m (Fig. 1).

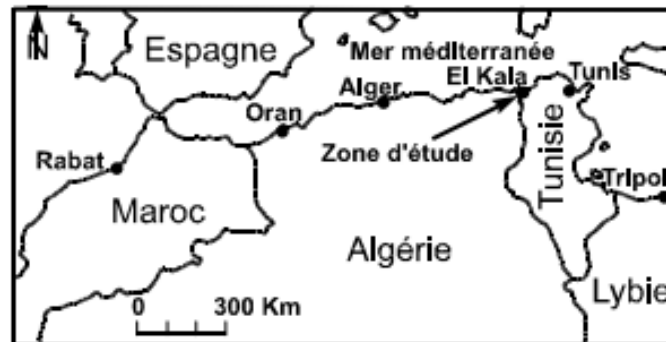


Fig.1. Location of the study zone

The bathymetric survey shows that the bottom is flat and smooth, with an increase in the depth of banks to the center. The maximum depth of Oubeira observed at the center varies between 2.40 m and 1.50 m during the wet period of closure dry [8].

The bottom of the lake is covered with a thick layer of mud, which reaches its maximum at the center (that is 2 m). This siltation derived from erosion could eventually lead to a serious destabilization of the ecosystem [9].

II.2 Geology of the study area

The Numidian is widely exposed in the Oubeira Watershed (Fig.2). It is represented by a predominantly sandstone formations at the basis and the top with respectively sub-numidian clays and associated clays with supra-numidian marl [10].

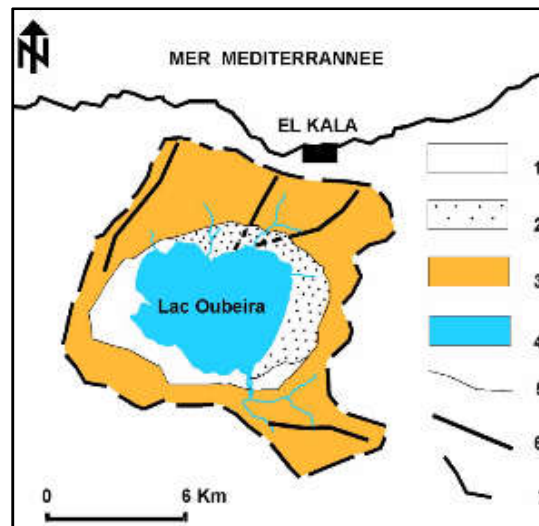


Fig.2. Geological card simplified the study zone.

1: Quaternary, 2: Pleistocene, 3: Flysch undifferentiated (especially Numidian sandstones and basic clays), 4: Bodies of water, 5 : Geological contacts, 6 : Fault, 7 : Watershed boundaries.

II.3. Description of data

Measurements taken *in situ* and water sampling for physicochemical analysis, were performed along six parallel transects, for a period from 2014 to 2015. Twenty-four measurement points have been made to better represent the spatial and temporal bathymetry variation, major elements and iron (Fig. 3). Thus, we have 132 observations.

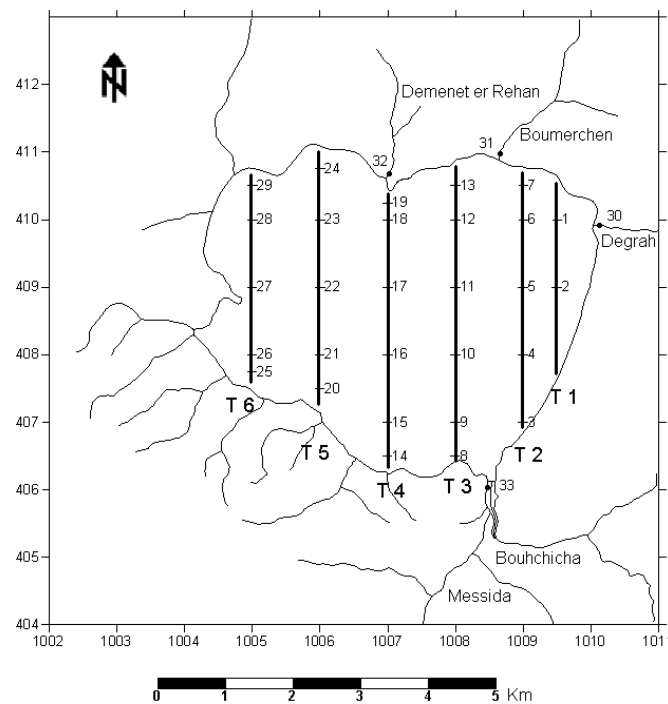


Fig.3. Measures Stations

Iron enters the colloidal lake through affluents and then disperses. The monitoring study of this element concentration, conducted between 2014 and 2015, shows a more or less cyclical (Fig. 4), and imposed by the erosion of the Oubeira watershed after perpetuation clays.

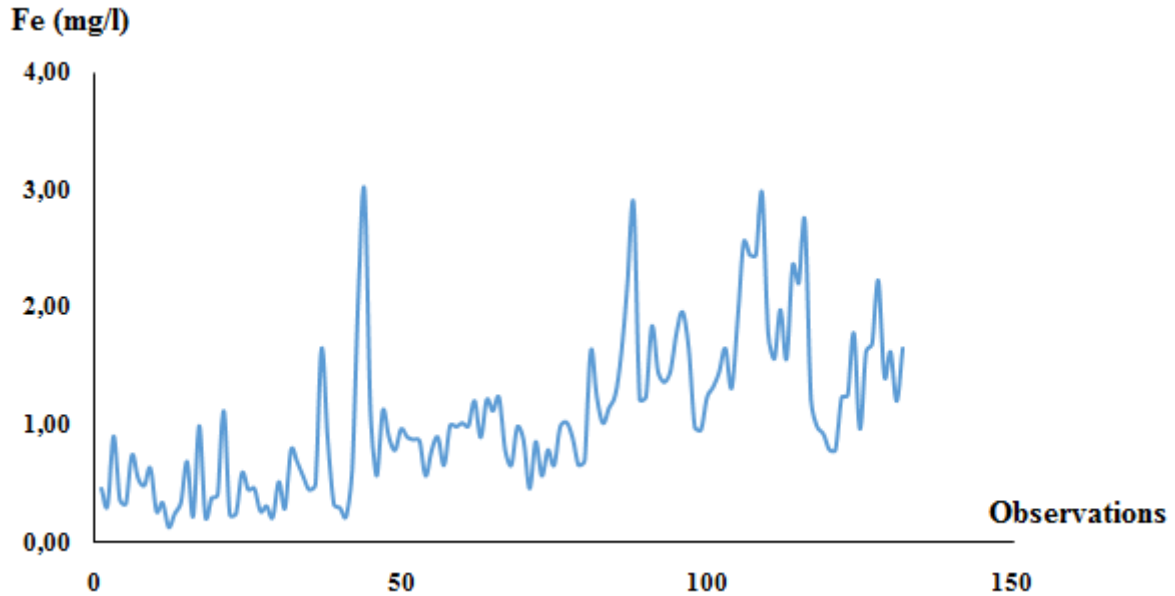


Fig.4. Variation in iron content of the lake Oubeira (Observation Period: 2014-2015).

II.4 Multiple Linear Regression

Multiple linear regression is a powerful statistical tool which models the values of a random variable noted (Y) based on several explanatory variables noted (X_1 , X_2 , ... X_n) [11]. The model is generally used to predict new values. The general form of the theoretical model of multiple linear regression is written as follows [12,13]:

$$Y = A_0 + A_1X_1 + A_2X_2 + \dots + A_nX_n + \varepsilon$$

With:

Y: Dependent variable (explained);

X_1, \dots, X_n : Independent variable (predictor);

A_0, \dots, A_n : Model parameters estimated using the least squares method;

ε : Residue.

The quality of fit is assessed using the correlation coefficient R.

II.5. Artificial neural networks type MLP

A forward-network of MLP type is graphically represented by a set of interconnected neurons; information flowing from inputs to outputs without backtracking [14,15]. The diagram of a feedforward network is acyclic (Fig. 5).

In fact, if we move in this type of network from any given neuron following connections, we cannot come back to the same initial neuron. Neurons that are running the latest calculation of functions composition is the output neurons. Those who make intermediate calculations are hidden neurons.

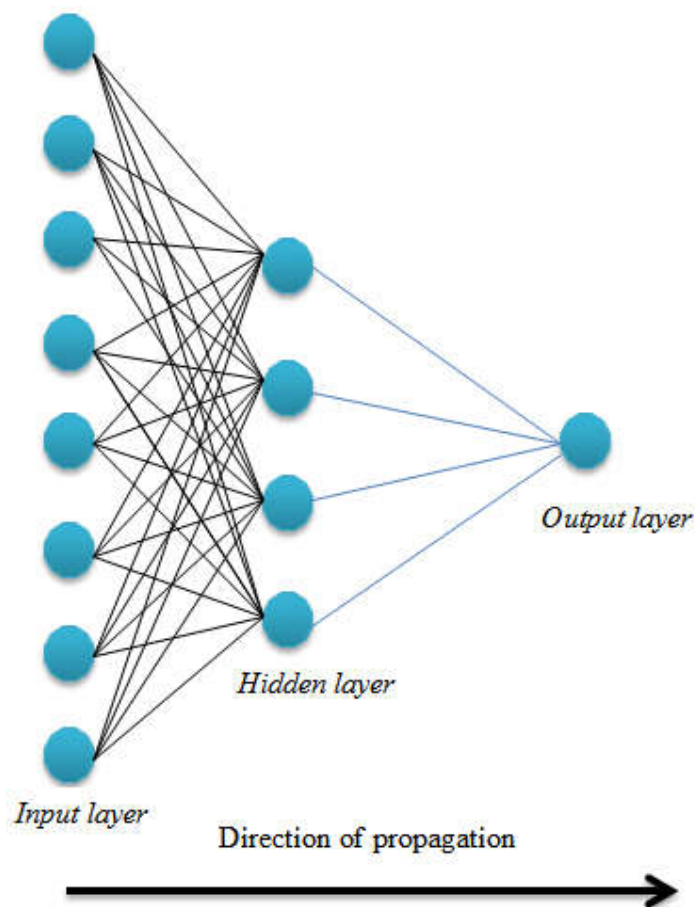


Fig.5. Schematic of an artificial neural network MLP

II.6. Modeling Techniques

Modeling approaches that have been developed to establish the best relationship between iron and major elements of the lake are based on two methods:

- Multiple Linear Regression
- MLP type Networks of artificial neurons.

Regression was applied to 132 observations, while artificial neural networks observations were split randomly between learning (70%), validation (15%) and the test (15%). 30% of the sample for the validation and performance testing of the prediction model involved in training.

III. RESULTS AND DISCUSSION

III.1. Multiple Linear Regression

The linear model provides a relation of the type:

$$[\text{Fe}] = -0.63112 + 0,03337 * [\text{Ca}^{2+}] + 0,08661 * [\text{Mg}^{2+}] + 0,00073 * [\text{Na}^+] - 0,15393 * [\text{K}^+] \\ + 0,00690 * [\text{Cl}^-] - 0,01556 * [\text{SO}_4^{2-}] - 0,00338 * [\text{HCO}_3^-] + 0,07095 * [\text{NO}_3^-]$$

The correlation coefficient or empirical explanation of the multiple regression model is 0.63. Variance analysis shows that the Fischer test calculated is higher than the theoretical Fischer test (about 1%). This model is highly significant at $P < 0.01$ (ie $P < 0.0001$), however only the magnesium and nitrates (variables) are significant at $P < 0.05$; the others are insignificant.

Observed levels scatter in iron based contents estimated by this model may fit a type of line $Y = a.X$. However, few high values of iron deviate from the confidence interval at 95% (Fig. 6). The model developed by the multiple linear regression method is limited, it explains only 63% of the variance.

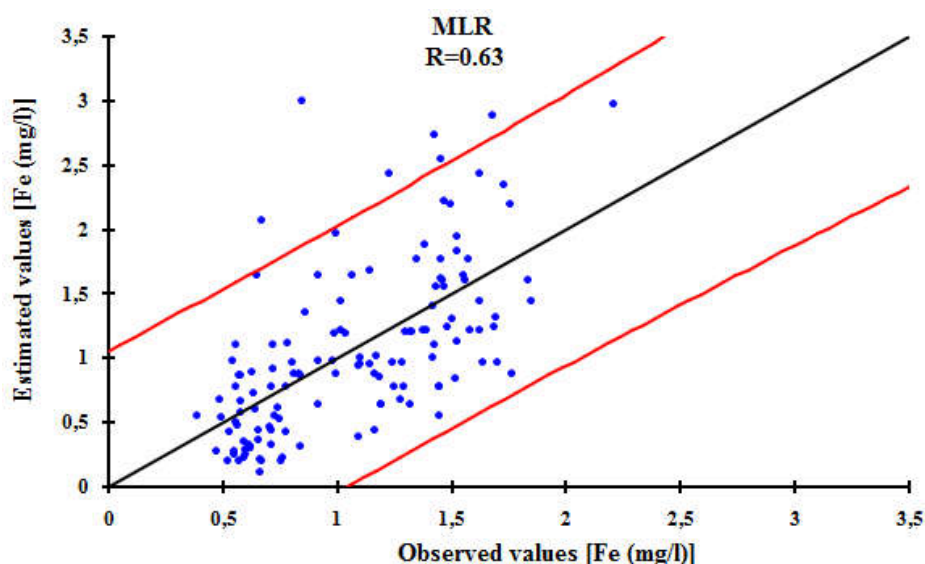


Fig.6. Iron levels observed as a function of levels estimated by the model.

The correlation coefficient obtained by the MLR model is $R = 0.63$. This shows that the levels of iron are not correlated very significantly in a linear way, with all the other physico-chemical parameters.

From a practical point of view, for environmental prediction of physicochemical parameters, nonlinear components of the studied systems and the number of variables become more important. Consequently, it is best to explore nonlinear models such as artificial neural networks.

III.2. Artificial Neural Network

To improve the performance of a model established by the MLP type neural networks (multilayer Perceptron), we must change the network architecture, playing mostly on the number of hidden layers, or on the hidden neurons and / or the number of iterations [16,17].

We used Matlab 2013 software [18]. The database is divided randomly into three groups:

- 70% to drive the network.
- 15% to validate the network.
- 15% to test the predictive quality of the model.

To identify the most suitable architecture, several tests with a hidden layer consisting of 1 to 20 neurons, using the sigmoid function as a hidden layer activation function and the linear transfer function for the output layer. Supervised learning helps retain the best model because it iterates until the minimum middle squared error noted (MSE) between calculated and observed values. Figure 7 depicts the training of the network in the case of iron.

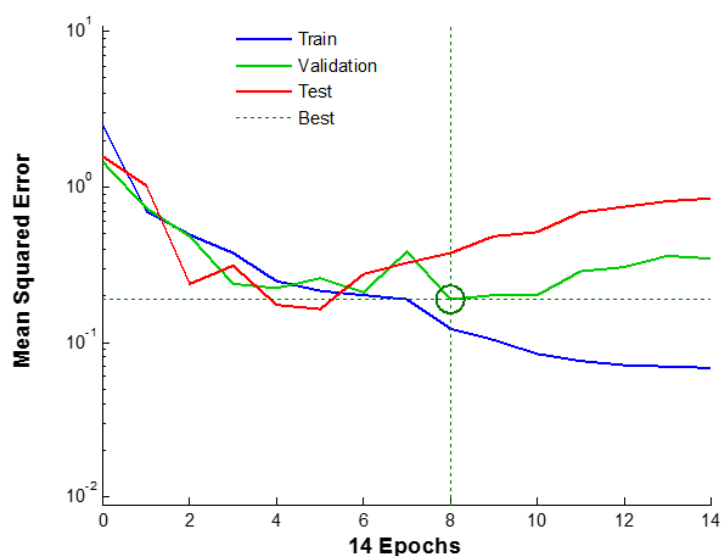


Fig.7. Evolution of the mean squared error in the case of iron with 9 neurons in the hidden layer

After eight iterations, the desired result is achieved with 9 neurons in the hidden layer (Fig. 8). The three curves (learning error, validation and test) properly converge on the optimum value of the middle square error.

Figure 8 shows the architecture of MLP type of artificial neural network. It is a network topology [8-9-1], which contains three layers:

- ✚ Eight neurons in the input layer represent the physical and chemical variables;
- ✚ Nine neurons in the hidden layer, are neurons that we have already chosen by trial and the performance indicators calculations;
- ✚ One neuron of the output layer, represents the iron contents.

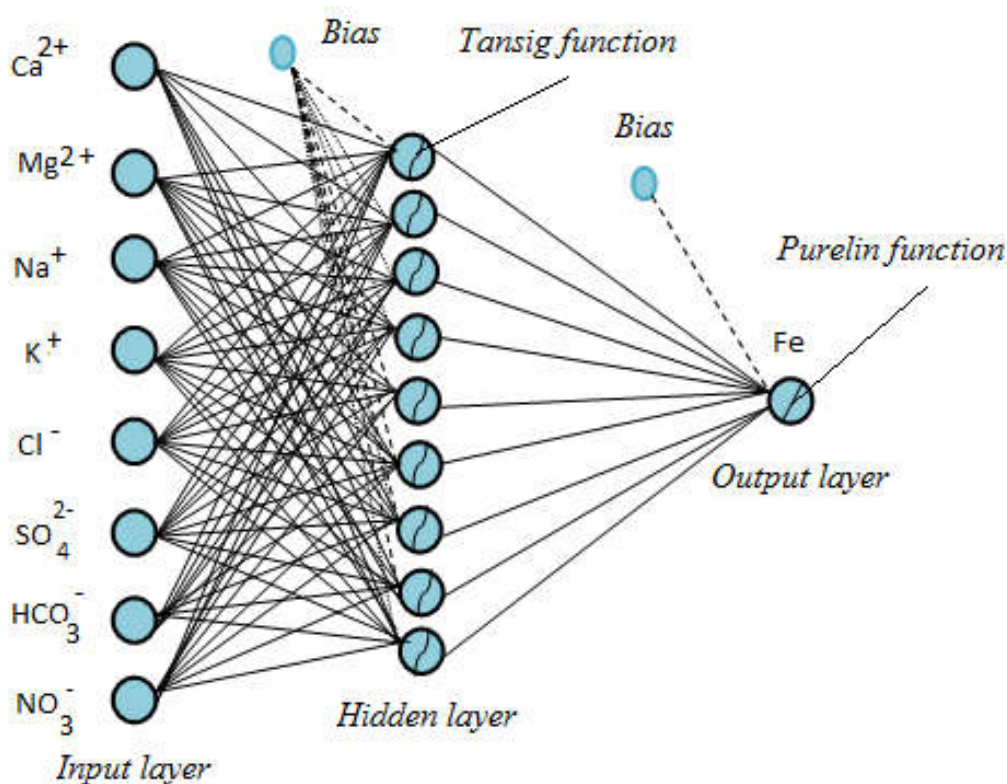


Fig.8. Architecture of the neural network to three layers of configuration [8-9-1] developed in this study.

The MLP type developed network model has a correlation coefficient for driving 87%. For the validation period, this coefficient is about 67 %, while for the test period, it was 84 %. By contrast, the correlation coefficients for this model to all three phases (training, validation and test) are about 79% (Fig. 9). This has been applied on a series of 132 observations.

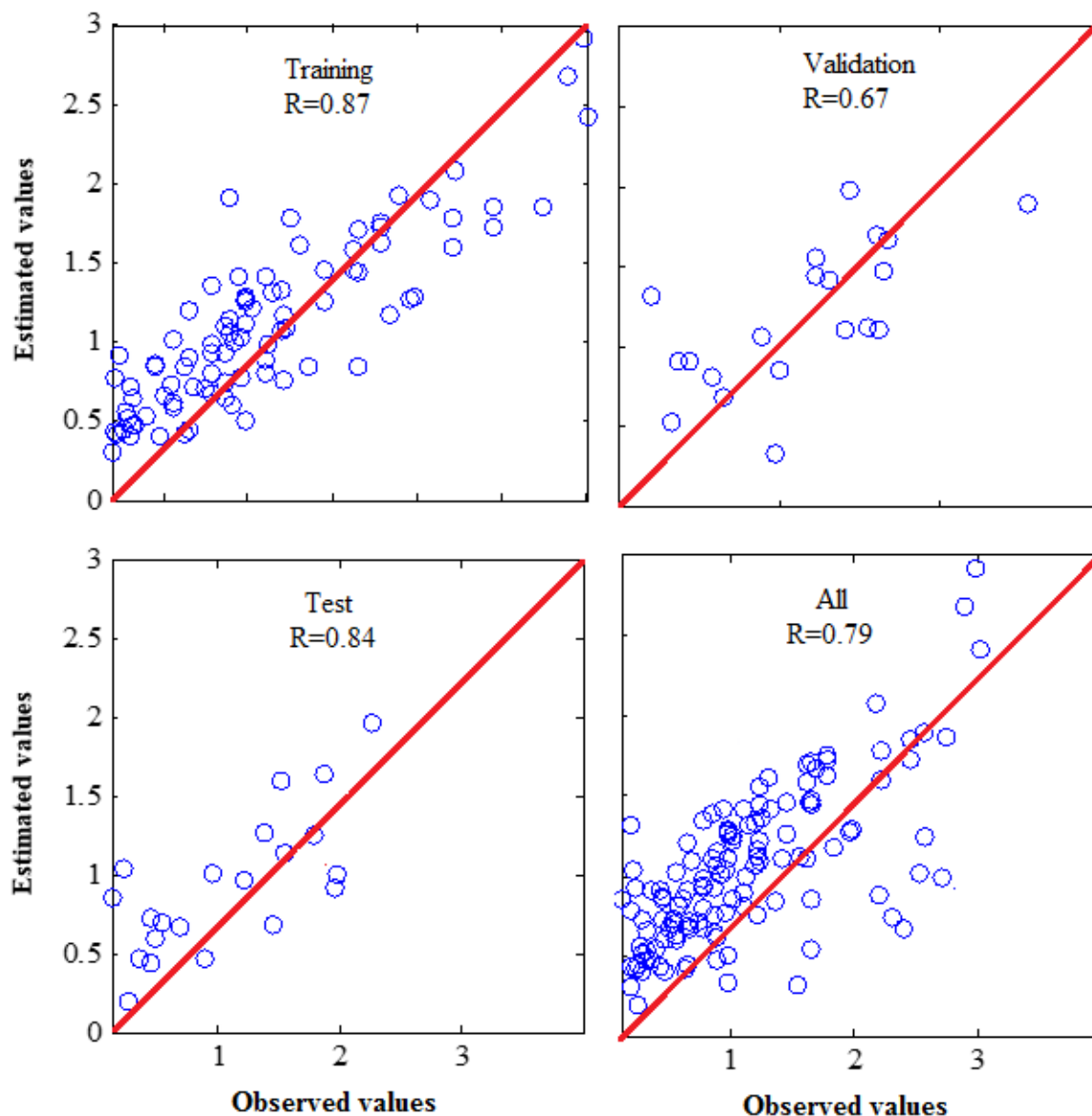


Fig.9. Relations between observed concentrations and those estimated with the model established by the model of artificial neural networks for the iron

III.3. Study of residues

The residue represents the difference between observed values and those estimated for both MLR models and ANN (Fig. 10). The model established by the neural network represents less dispersed residues comparing to those of the multiple linear regression.

Figure 10 shows the variation of residuals of the estimated values for the two models developed in this study. They show that the residues obtained from the MLP type developed model are significantly less dispersed. Generally between -0.5 mg/l and $+0.5 \text{ mg/l}$, compared with those established by the multiple linear regression model (-0.5 mg/l and $+2.5 \text{ mg/l}$).

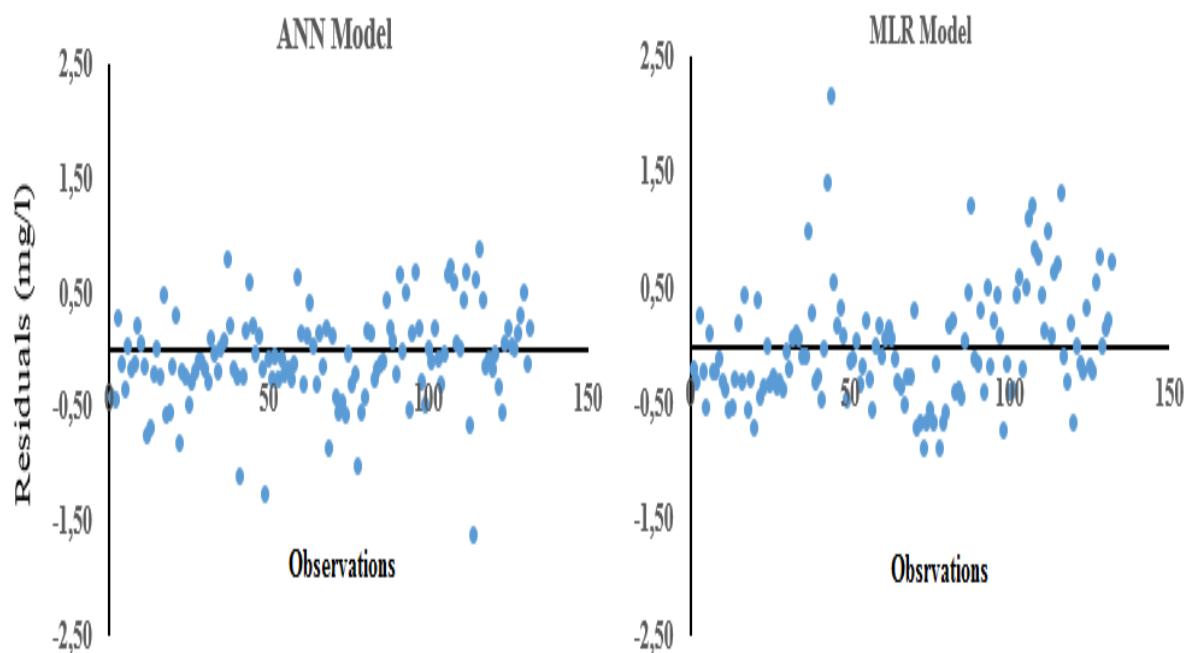


Fig.10. Relations between the estimated values of iron and their residues with the models established by the ANN and MLR

The MLP type model of artificial neural network developed in this study for the prediction of iron contents is better than the one developed by multiple linear regression. This performance was also reported in many recent works concerning studies of the heavy metals relationships with physico-chemical parameters of river sediments [19-22], and the prediction of organic carbon from the physico-chemical parameters in marine sediments [23].

IV. CONCLUSION

Iron contents of the studied aquatic ecosystem derive from the geochemical background of the area. These contents are subject to considerable variations from the $\mu\text{g} / \text{l}$ to mg / l , which confer a difficult and unpredictable track. Also, for a better understanding of this pollutant, a modeling project has been undertaken to predict the iron contents of the Oubeira Lake water. Our results reveal the neural network method to be more effective than the multiple linear regression. In fact, the artificial neural networks allow an improvement of about 20% of the explanation of the variance with respect to that of the multiple linear regression. Iron concentrations in the ecosystem are related to major elements by non-linear relationships, which is common in aquatic ecosystems.

These results are encouraging to undertake further modeling approaches provided by artificial neural networks to improve the iron prediction model and possibly other ETM on one hand and enrich the database, on the other hand.

REFERENCES

- [1] El Badaoui, H. Abdallaoui, A. Simulation numérique par l'utilisation des techniques d'intelligences artificielles pour la modélisation des données météorologiques. Editions Universitaires Européennes ISBN : 978-3-639-50372-2, Germany , 2016 : 216
- [2] Burel, G. Réseaux de neurones en traitement d'images : Des modèles théoriques aux applications industrielles. Thèse de Doctorat, Université de Bretagne Occidentale, 1991 : 320
- [3] Prevost, L. Reconnaissance des formes : Apprentissage et fusion d'information, HDR, 2007 : 94
- [4] Drew, P.J. Monson, J.R.T. Artificial neural networks, *Surgery* 2000, 127 : 3-11
- [5] Schmitt A., Le Blanc B., Corsini M.M, Lafond C., Bruzek J. les réseaux de neurones artificiels : Un outil de traitement de données prometteur pour l'anthropologie. *Bull. et Mém. de la Société d'Anthropologie de Paris*, 2001, 13 (2) : 143-150
- [6] Bouaoune D., Dahmani, M. Reconstitution de données climatiques pour l'Algérie du Nord : Application des réseaux neuronaux. *Comptes Rendus Géoscience*, 2010, 342 : 815–822
- [7] Perez, P. Trier, A. Prediction of NO and NO₂ concentrations near a street with heavy traffic in Santiago. *Atmospheric Environment*, 2001 : 1783-1789
- [8] Alayat, H. El Khattabi, J. Lamouroux, C. Evolution spatiale des caractéristiques physico-chimiques des eaux du lac Oubéira imposé par les conditions sévères de la sécheresse (Extrême NE algérien). *European Scientific Journal*. 2013 : 564-579
- [9] Alayat, H. Lamouroux, C. Evolution spatiale de l'envasement du lac Oubéira imposé par l'érosion (extrême NE algérien). Second symposium on the geological resources in the Tethys realm. Aswan University, Egypt, 5-8 january, 2013 : 51-56
- [10] Alayat, H. Mouissi, S. Satouh, R. Contribution to the knowledge of the bathymetry and the spatial distribution of N, P, K of the mud of lake Oubeira (Extreme NE Algerian). *Lakes, reservoirs and ponds*, 2012, 6 (1) : 75-90
- [11] Falissard, B. Comprendre et utiliser les statistiques dans les sciences de la vie, 3^{ème} édition Masson, 2005 :101-372

-
- [12] Goulon-Sigwalt-Abram A. Une nouvelle méthode d'apprentissage de données structurées : Applications à l'aide à la découverte de médicaments. Thèse de Doctorat, Université Paris VI, 2008 : 135
- [13] Colliez, J. Une approche basée sur la régression par les machines à vecteurs supports : Application au suivi d'objets en mouvement dans les séquences d'images. Thèse de Doctorat, Université Littoral Cote d'Opale, 2008 : 141
- [14] Touzet, C. Les réseaux de neurones artificiels. Introduction au connexionisme, EC2 Ed., Nanterre, 1992 : 150
- [15] Touzet, C. Les réseaux de neurones artificiels – Introduction au connexionisme– cours, exercices et travaux pratiques, 1992 : 126
- [16] Cheggaga, N. Youcef Ettoumi, F. Estimation du potentiel éolien. *Revue des Energies Renouvelables*, 2010 : 99–105
- [17] Chabaa, S. Identification des systèmes non linéairement utilisant les techniques d'intelligence artificielles et les bases de fonctions de la guerre pour la modélisation des données du trafic dans les réseaux internet. Thèse de Doctorat, Université Cadi Ayad, Faculté des Sciences Semlalia-Marrakech, 2011 : 187
- [18] Logiciel Matlab. *Neuronal Network Toolbox*, 2015: 680
- [19] Abdallaoui A. Le phosphore et les métaux lourds contenus dans les sédiments. Leur influence sur les phénomènes d'eutrophisation et de pollution. Editions Universitaires Européennes, Germany, 2013 ; 348
- [20] El Badaoui, H. Abdallaoui, A. Manssouri I. and Lancelot, L. Elaboration de modèles mathématiques stochastiques pour la prédiction des teneurs en métaux lourds des eaux superficielles en utilisant les réseaux de neurones artificiels et la régression linéaire multiple. *Journal of Hydrocarbons Mines and Environmental Research*, 2012, 3 : 31-36
- [21] El Badaoui, H. Abdallaoui, A. Manssouri, I. Lancelot, L. Application of the Artificial Neural Networks of MLP Type for the Prediction of the Levels of Heavy Metals in Moroccan Aquatic Sediments. *International Journal of Computational Engineering Research*, 2013, 3 : 75-81
- [22] Abdallaoui A., El Badaoui H. Prédiction des teneurs en métaux lourds des sédiments à partir de leurs caractéristiques physico-chimiques. *Journal Physical and Chemical News*, 2011, 58 : 90-97
- [23] EL Hmaidi A., El Badaoui H., Abdallaoui A., EL Moumni B. Application des Réseaux de Neurones Artificiels de Type PMC Pour la Prédiction des Teneurs en Carbone Organique dans les Dépôts du Quaternaire Terminal de la Mer d'Alboran. *Journal European of Scientific Research*, 2013, 107 : 400-413

How to cite this article:

Alayat H, El Badaoui H, Abdallaoui A, Abrid D, El Hmaidi A. Development of mathematical models for predicting the iron concentrations of Lake Oubeira waters (NE ALGERIAN). *J. Fundam. Appl. Sci.*, 2018, *10(1)*, 83-96.