Algerian Journal of Environmental Science and Technology
September edition. Vol.8. N°3. (2022)
ISSN : 2437-1114
www.aljest.org

Quantification of salinity using artificial neural networks. Case study of the Isser River (Algeria)

K. Houari

M'hamed Bougara University, Faculty of Sciences, 35000 Boumerdes, Department of agricultural sciences, Algeria.

Corresponding author: k.houari@univ-boumerdes.dz; Tel: +213 541 141 576.

ARTICLE INFO

Article History:
Received ***: 25/07/2020***
Accepted ***: 26/01/2021***

Key Words:

Artificial intelligence; saline flow; salinity; Isser River; MLP.

ABSTRACT/RESUME

Abstract: *In this study, we examine a very complex phenomenon in the salinity of rivers. The complexity of the mineralization of rivers makes it difficult to quantify by traditional statistical models. Artificial intelligence is an interesting and fully justified alternative for modeling non-linear phenomena. The present study aims to develop a model based on a multilayer perceptron neuron network (MLP), capable of explaining the salt concentration-liquid flow relationship. The method is based on observations recorded at the Oued Isser outlet at the Lakhdaria hydrometric. The obtained results indicate that the artificial neural network model (MLP) has the best performance compared to the empirical model, with an "R2 " value for regression analysis of training, testing and validation of 80%, 75% and 78%, respectively. Further, the «Nash» value for training, testing and validation are 79%, 72% and 75%, respectively.*

I. Introduction

One of the most sensitive aspects of water resource management is the salinity of rivers. In recent decades, Isser rivers, like most other rivers, have noticed a quality deterioration. This degradation of water resources is due to the leaching of land loaded with mineral salts, during its course [1]. According to the pioneering work of Garrels and Mackenzie (1971) [2], the influence of lithology on erosion rates, both locally (watersheds less than 150 km²) and globally, has been clearly demonstrated by the work of Meybeck and Probst [3,4,5].

The analysis of small unpolluted mono-lithological basins (temperate region) has shown, in particular, that rivers draining sedimentary rocks have erosion rates (estimated from the total dissolved load or TDS = Total Dissolved Solid) five times higher than those of rivers draining crystalline or metamorphic rocks, and 2.5 times higher than rivers draining recent volcanic rocks [3]. The physical properties (texture and structure) of rocks play a major role in the process of mechanical erosion of them. They will therefore affect the proportion of particles supplied to river systems. The chemical properties of rocks have a role in the nature of the exported material. All of these properties (physical and chemical) derive from the geological context in which the rocks were formed.

Therefore, the chemical composition of the water depends on the nature of the terrain crossed and with the presence of unfavorable climatic conditions, the water during its journey carries dissolved bodies ($<10^{-5}$ mm); what is called salinity. This salinity is only a chemical contribution, its main components are mineral salts expressed by the sum of anions and cations

The salinity of streams and rivers is a complex hydrological and environmental phenomenon due to the large number of obscure parameters [6].

Moreover, the processes involved in salinization are so complex that it is difficult to establish an analytical model allowing precise quantification of the salinity of rivers [7,8]. Artificial intelligence has been successfully applied in various fields [9,10,11], particularly in the context of hydrology [12,13]. In our study, we used a new method of extrapolating saline concentrations; based on MLP neural networks.

II. Quantification models

II.1. Empirical model

This model was proposed by the National Agency for Hydraulic Resources (ANRH) in 1996 for the

quantification of rivers salinity [14]. It is established between liquid flow rates (*Ql*) and salt concentrations (C_S):

$$C_s = AQ_l^B \quad (1)$$

With:

C_S : Saline concentration (g. l^{-1}) ;
Ql: Liquid flow (m^3. s^{-1});
A: Degree of salinity;
B: Coefficient.

Or:

$$Ln(C_S) = Ln\left(A \cdot Q_L^B\right) \quad (2)$$

By correlating the series Ln (C_S) and Ln (*Ql*), it becomes:

$$Ln(C_S) = LnA + BLnQ_L \quad (3)$$

II.2. Model (MLP)

The artificial neural network (ANN) is a data processing system consisting of a large number of simple and highly interconnected processing elements, resembling a biological neural system. It has the capability of learning from an experimental or real data set to describe the nonlinear and interaction effects with great success [15,16,17]. Artificial intelligence is a viable and fully justified alternative for modeling phenomena with non-linear behavior [18,19]. There are a large number of types of neural networks, each of them has advantages and disadvantages. The network chosen in our case is a multilayer network (MLP). This choice is made for the ease and speed of its construction and, again, by the fact that our problem has a limited number of input variables (Figure 1). A multilayer perceptron consists of a succession of layers made up of neuronal units, which have a non-linear activation function. Within a layer each neuron receives signals from the previous layer, performs a calculation, and passes the result to the next layer [20]. There are no interconnections between neurons located inside the same layer: the activations of the different neurons are only propagated from the input layer to the output layer, through all the constituent neurons of the network. The input layer collects the input variables while the output layer produces the results [21].

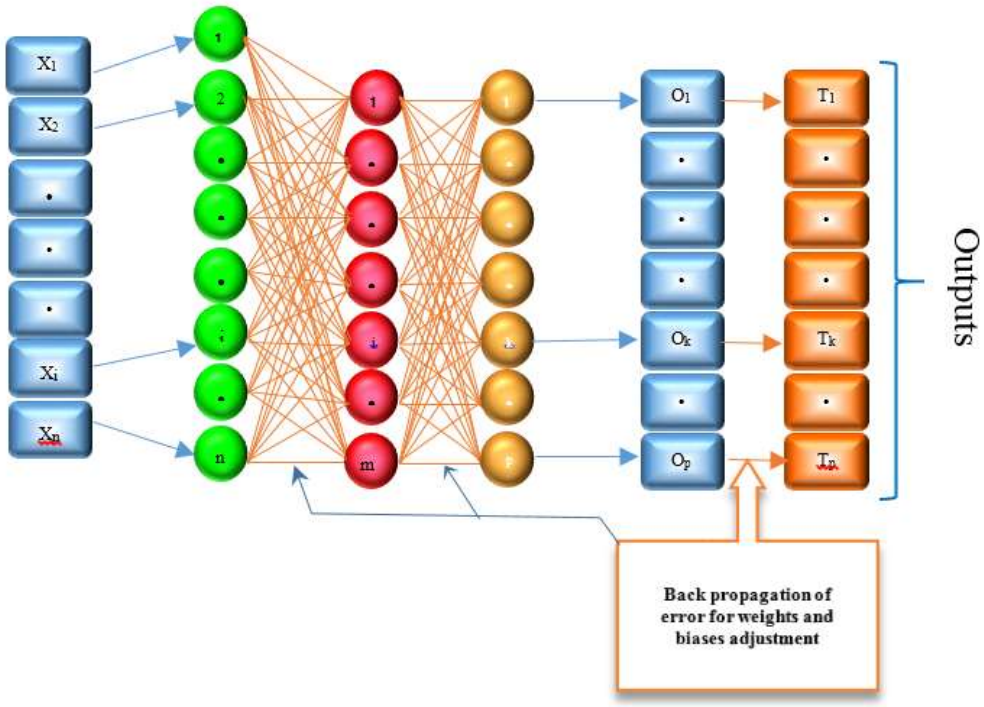

Figure 1. *Architecture of the MLP model multilayer perceptron.*

The first layer of the network is the input layer. It contains (n) neurons. The second layer, called the hidden layer, contains (m) neurons. The last layer of the network is its output layer, containing (p) neurons. Input neurons are numbered 1 to n, hidden neurons 1 to m, and output neurons 1 to p. By convention, the parameter wij relates to the connection going from neuron i (or from input i) to neuron j. Thus, the parameter wjk relates to the connection going from the hidden neuron j to the output neuron k. The states of the neurons of the first layer will be fixed by the problem treated through a vector $x = (x_1; x_2; \ldots x_n)$. The states of the first layer being fixed, the network will be able to calculate the vector $x = (x_1; x_2; \ldots x_n)$. The states of the first layer being fixed, the network will be able to calculate the states of the neurons of the other layers. In this sense,

each neuron in the hidden layer receives a sum weighted by the parameters
(wij), which are, then, often referred to as "weights" or, due to the biological inspiration of neural networks, "synaptic weights " of all entries, to which is added a constant term w_0 or" bias ":

$$j = W_0 + \sum_{i=j}^{n} W_{ij} \times x_i \quad (4)$$

The output of the neuron is a nonlinear function of its input (Aj):

$$Y_j = f(A_j) \quad (5)$$

In this study, we used the sigmoid activation function:

$$S(y) = \frac{1}{1+e^{-y}} \quad (6)$$

Each neuron in the output layer receives a sum weighted by the parameters (wjk), to which is added a constant term B_0 or "bias":

$$O_K = B_0 + \sum_{j=1}^{m} W_{jk} \times Y_j \quad (7)$$

The output of the network O (for Output) is the weights linear function of the last layer of connections (which connects the m hidden neurons to the output neurons), and is a non-linear function of the parameters of the first layer connections (which connect the n inputs of the network to the m hidden neurons). This property has very important consequences [22]. It has been shown that a neural network comprising a layer of finite-numbered hidden neurons, all having the same activation function, and a linear output neuron is a universal approximator [23, 24]. The values of the weights and the bias are changed and updated via a supervised learning algorithm. The latter consists of obtaining a set of examples, that is to say a finite set of known input / output pairs (examples which constitute the learning set). The objective of this calculation is to minimize an error function between the desired response and the response obtained from the output of the model. The most widely used error backpropagation algorithm estimates the gradient of the error function with respect to the parameters (weight and bias) of the model, performing the adaptation of these parameters successively from the output layer to the input layer. It consists of performing a gradient descent on the error criterion 'E' while minimizing a cost function, the root, generally, means square error [21, 25]

II.2.1. Learning neural networks

Learning is the most interesting phase of neural networks. The training phase consists of determining the weights of the connections between neurons on the sample in order to minimize the error between the desired output and the output calculated by the network [26, 27]. In this study, a supervisor (or human expert) provides an output value or vector (called a target or desired output), which the neural network must associate with an input vector. In this case, learning consists of adjusting the network parameters in order to minimize the error between the desired output and the actual output of the network (Figure 2)

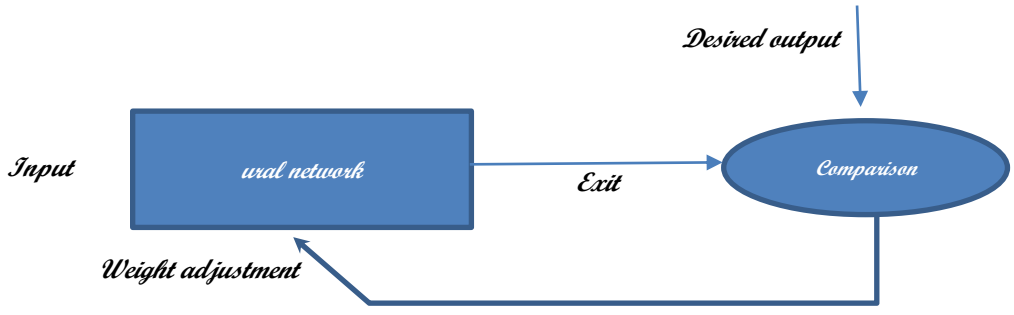

Figure 2. *Supervised learning.*

II.3. Validation of models

Validation makes it possible to judge the ability of the model to reproduce the modeled variables in the learning, validation and testing phases. In this study, we have based on the coefficient of determination (R^2) and the Nash-Sutcliffe coefficient of efficiency Nash and Sutcliffe (Nash) [28]. These parameters are given by the following relations:

$$E = 1 - \frac{\sum_{i=1}^{n}(Qt_i - \hat{Q}t_i)^2}{\sum_{i=1}^{n}(Qt_i - \overline{Q}t_i)^2} \quad (8)$$

$$R^2 = \left(\frac{\sum_{i=1}^{n}(\hat{Q}t_i - \overline{\hat{Q}t_i})(Qt_i - \overline{Q}t_i)}{\left(\sum_{i=1}^{n}(\hat{Q}t_i - \overline{\hat{Q}t_i})^2\right)\left(\sum_{i=1}^{n}(Qt_i - \overline{Q}t_i)^2\right)} \right)^2 \quad (9)$$

Where:

Qt_i is the measured value of the flow;

$\hat{Q}t_i$ is the flow calculated by the model;

$\overline{Q}t_i$ is the average measured flow and 'n' is the number of data.

The following table (Table 1) indicates the values of the criteria Nash and R², corresponding to the different degrees of performance of the model at the daily time step.

Table 1. *Quality of the models according to the values of the various criteria for a daily time step.*

	Nash (%)	**R² (%)**
Very good	**75 < Nash ≤ 100**	75 < R² ≤ 100
Good	**65 < Nash ≤ 75**	65 < R² ≤ 75
Satisfactory	**50 < Nash ≤ 65**	50 < R² ≤ 65
Unsatisfactory	**Nash ≤ 50**	R² ≤ 50

III. Study area

The Oued Isser watershed is located in the North of Algeria. It is limited to the North-East by the daïra of Drâa El Mizane, to the South-East by the wilaya of Bouira, to the South by the daïra of Ain Boucif to the South-West by the daïra of Ksar El Boukhari and the wilaya of Médéa, to the south. North-West by the daïras of Tablat and Larbâa and in the North by the Mediterranean Sea (Figure 3). The area of the basin is 4,126 km². The Isser watershed is located approximately 70 km south-east of Algiers. It is substantially in the form of a quadrilateral with a Southwest / Northeast orientation. From an administrative point of view, this basin straddles several wilayas: Médéa, Bouira, Tizi Ouzou and Boumerdes. The geographical framework in which the watershed develops is made up of the Tellian Algiers Atlas in the North which culminates at 1130 m at Jebel Tamesguida and the chain of Bibans in the South which culminates at 1810 m at Jebel Dira. These two chains are separated by the Aribs Plain at an altitude of 550 m. The Oued Isser watershed, whose crest lines are between 90 and 1810 m at an average altitude of 710 m [29]. The sparse vegetation cover is located in the center and represents 20% of the total area; while the rest of the area, ie 80%, is occupied by crops mainly cereals and fodder. At the bottom of Isser, we mainly find annual crops [30].

Algerian Journal of Environmental Science and Technology
September edition. Vol.8. N°3. (2022)
ISSN : 2437-1114
www.aljest.org

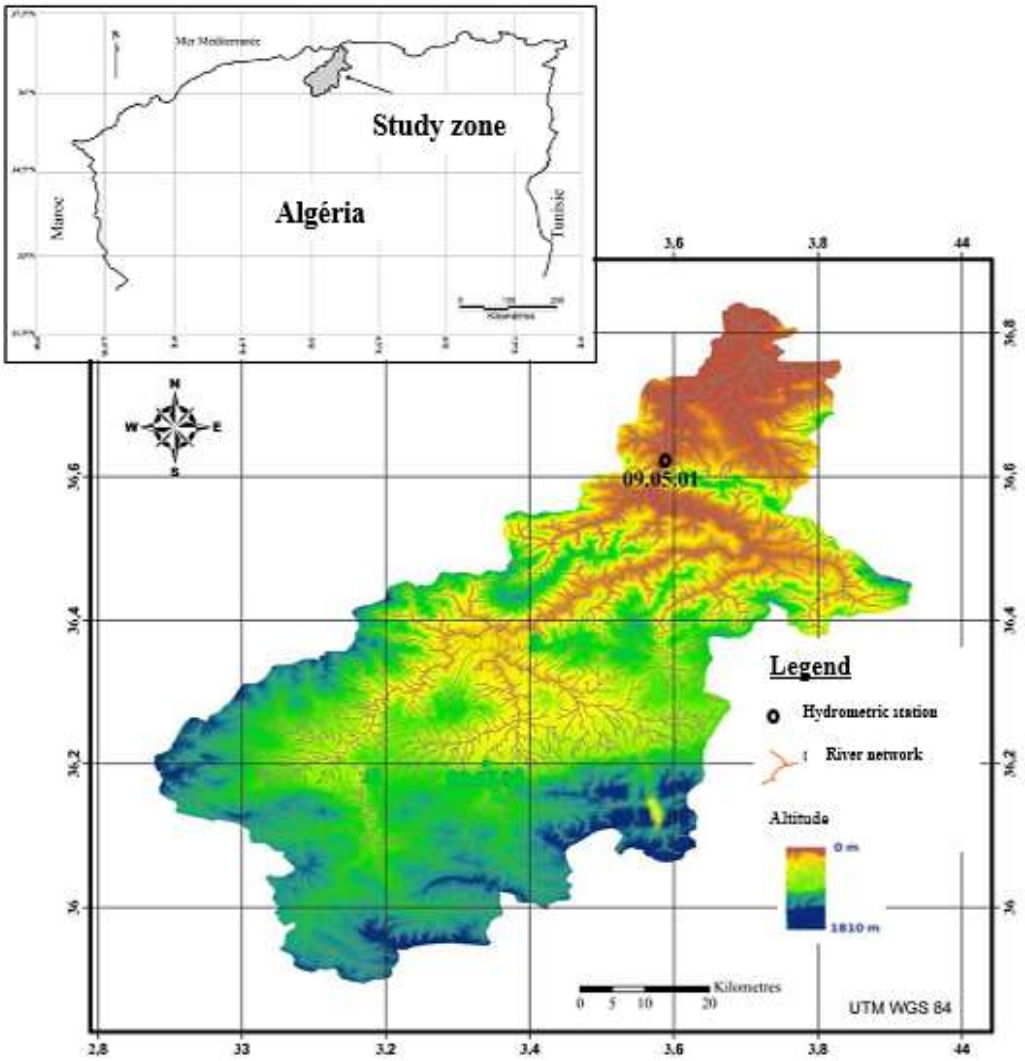

Figure 3. *Location of the study area [28].*

IV. Data and method

The choice of the appropriate methodology for the calculation and prediction of salinity depends on the nature of the hydrometric data available. We entered the baseline sample containing the following information:

- Name of the hydrometric station.
- Date and time of collection.
- Liquid flows *Ql* ($m^3. s^{-1}$).
- Water height H (cm) as a function of time; taken daily $H = F(T)$. To translate the data from heights of water in (cm) to liquid flows ($m^3. s^{-1}$), we used the calibration curves $Ql = f(H)$ simplified by the calibration scales which express the liquid flows from a water height that varies from 5 to 10 cm. By interpolation, we obtained the liquid flow corresponding to the desired height of water.
- C_S salt concentrations ($g.l^{-1}$) which express mineralization as a function of time corresponding to similar scales of water heights (H).

We have chosen Lakhdaria station (090501) as a reference station for the validation of the models proposed on a daily scale. This station allows us to follow the evolution of the salt concentration downstream of Oued Isser. For this, we took the daily average values of 300 samples of liquid flows ($m^3. s^{-1}$) and saline concentrations ($g.l^{-1}$), collected over a 10-year observation period (1970 to 1973 and 1979 to 1984) (Figure 4).

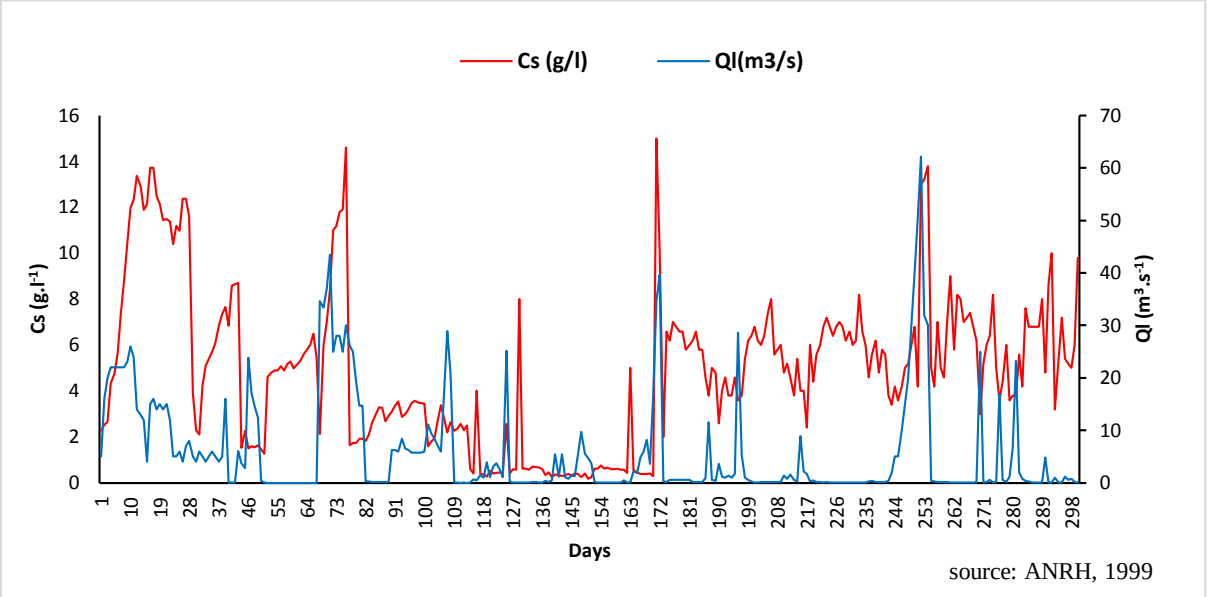

Figure 4. *Daily average values of liquid flows and saline concentrations observed over 10 years (1970 to 1973 and 1979 to 1984).*

The data set (300 samples) was divided into three subsets:

- Training.
- Testing.
- Validation.

For each model, after several scenarios and several tests, we set a percentage for each phase:

- Training phase: We used two hundred and fifty-two pairs of liquid flow rates and salt concentration, i.e. 60% of the population; during the periods extending from 04/09/1970 to 18/09/1973 and 18/01/1979 to 26/01/1980.
- Test phase: This is another observation period (18/05/1980 to 05/01/1982) of eighty-four samples, representing 20% of the population.
- Validation phase: Eighty-four samples, i.e. 20% of the population, are used for this validation phase (from 05/02/1982 to 03/12/1984).

V. Results and discussion

The validity criteria of the two models are shown in Table 2 for the three phases (learning, testing and validation).

The comparison graphs of the simulated and observed values for the two models are shown in Figure 5. The neural network model performed better, compared to the empirical model, with an "R^2 " value for regression analysis of training, testing and validation of 80%, 75% and 78%, respectively. Further, the «Nash» value for training, testing and validation are 79%, 72% and 75%, respectively.

Table 2. *Daily evaluation criteria for the MLP and empirical models.*

Model	Phase	Nash (%)	R^2 (%)
MLP	***Training***	79.00	80.00
	Testing	72.00	75.00
	Validation	75.00	78.00
Empiriqual $C_s = aQ_L^B$	***Training***	70.00	69.00
	Testing	66.00	60.00
	Validation	68.00	66.00

Algerian Journal of Environmental Science and Technology
September edition. Vol.8. Nº3. (2022)
ISSN : 2437-1114
www.aljest.org

Figure 5 shows the time series of observed and expected salt flows during training, validation and testing; respectively for the MLP and empirical models. The MLP model explains, with precision, the non-linear model of saline flows, during the formation phase and then produces a good generalization during the other two phases.

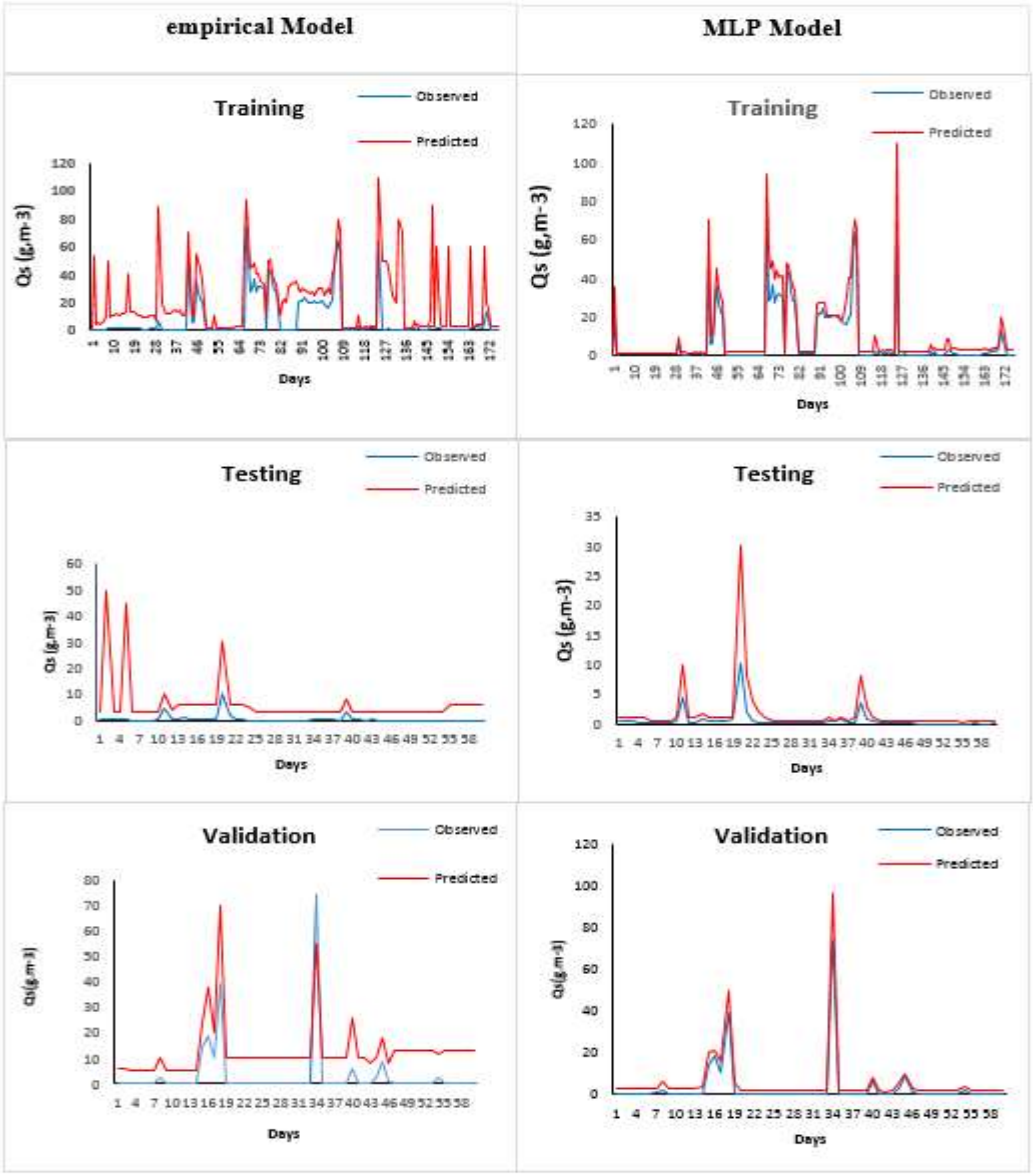

Figure 5. *Comparison of predicted and observed values for MLP and empirical models.*

On the basis of these results, we quantified the saline inputs (Table 3 and Figure 6). In this study area, the water became increasingly saline over time. Saline inputs from Oued Isser are increasing significantly. Nevertheless, this quantity, although, important remains negligible in the studies of quantification of the solid transport. This automatically affects the sizing calculations of the dike as well as the life and site of dams and hill reservoirs.

Table 3. Annual variation of salinity.

Years	Salinity (10^3 t.year^{-1})	Years	Salinity (10^3 t.year^{-1})
1970	110,00	1978	150,00
1971	158,10	1979	527,62
1972	384,81	1980	196,64
1973	352,00	1981	832,78
1974	209,21	1982	641,10
1975	166,93	1983	999,19
1976	483,15	1984	712,33
1977	156,38		

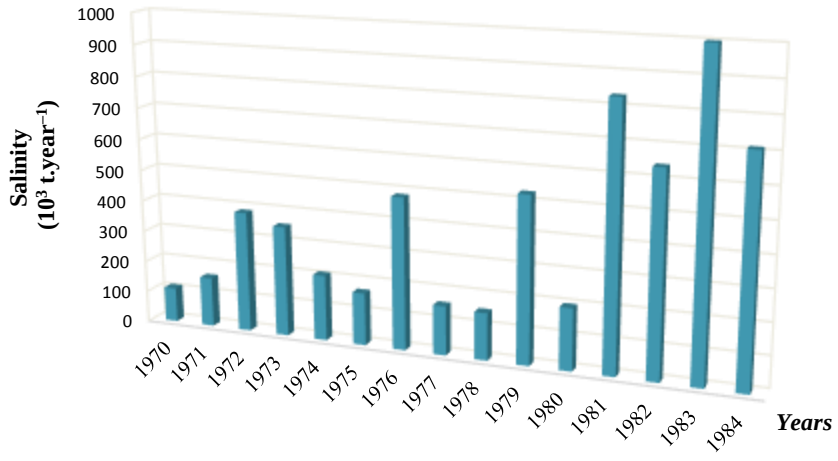

Figure 6. *Annual variation of salinity.*

VI. Conclusion

We have tested two models, one empirical and the other based on neural networks. These two models were used to explain the relationship between salt concentrations and liquid flows and to assess the degree of salinization. In this study, we used the time series of the daily mean liquid flows and the daily mean salt concentrations from the hydrometric station (090501).

The model based on neural networks gave efficient results on a daily scale and on the scale of well observed floods, the criteria "Nesh" and "R²" are generally high, whether in calibration, validation or in testing, justifying the predictive power of this model with an R^2 equaling 80% for training, 72% for tests and 75% for validation, and an E equaling 79% for training, 75% for tests, and 75% for validation. In addition, artificial intelligence constitutes an interesting and fully justified alternative for modeling non-linear phenomena. In this study, the neural networks were used to explain the relationship between salt concentrations and liquid flows, and to assess the degree of salinization.

The development of a methodology for predicting salinity from liquid flows by developing a model based on neural networks has made it possible to predict the salinity of rivers inputs at the time series data of average daily liquid flow, saline concentrations and better explained the nonlinear relationship between salt concentrations and liquid flow.

Saline inputs from Oued Isser are increasing significantly and the water has become increasingly saline over time. The quantities of salts transported by Oued Isser are significant with an average of 400,000 t / year.

These encouraging results open up a number of perspectives, where it would be interesting to try to apply MLP models on a larger scale in hydrology and the environment. This approach should be tested on a large scale for the whole country and, if successful, integrated into early warning systems in the event of Water Supply and Sanitation water quality degradation.

VII. References

1. Tessier, L. Transport and characterization of suspended matter in the watershed of the Seine: identification of natural and anthropogenic signatures, *Doctoral Thesis* (2003), Ecole nationale des Pont et Chaussées France.
2. Garrels, R.M.; Mackenzie, F.T. Evolution of sedimentary rocks, *W.W. Norton and Company Inc*., (1971), New York, 397 p.
3. Meybeck, M. Chemical composition of unpolluted French streams. *Geological Sciences* (1986).
4. Meybeck, M. Global chemical weathering of surfical rocks estimated from river dissolved loads. *American Journal of Science*, 5:287 (1987) 401-428.
5. Probst, J.L. Geochemistry and hydrology of continental mechanical erosion, current global balance, fluctuations over the last 500 million years, *Doctoral thesis*, University of Strasbourg, (1990) 323 p.
6. Touat, S. Methodology for the Study of the Salinity of Rivers, Report No. 5 (1992), 52-63. ANRN.
7. Wenrui, H,; Simon, F. Neural network modeling of salinity variation in Apalachicola River, *Water Research*, 36 (2002) 356-362.
8. Houari, K.; Hartani, T.; Remini, B.; Lefkir, A.; Abda L.; Heddam S. A hybrid model for modelling the salinity of the Tafna River in Algeria, *Journal of Water and Land Development*. 1:3 (2019) 127-135.
9. Benatiallah, D. ; Benatiallah A. ; Bouchouicha K. ; Nasri B. Predicting hourly solar radiation using artificial neuron networks, *Algerian J. Env. Sc. Technology*, 6 :1 (2020) 1236-1245.
10. Nemili, Z.; Kalla, M. Modeling of Bridge Scour by Artificial Neural Networks based on PCR, *Algerian J. Env. Sc. Technology*, 5:3 (2019) 1036 -1040.
11. Jang, J.S.R.; Suni, C.T.; Mizutani, E. Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence. *IEEE transactions on automatic control*, 42:10 (1997) 1482-1484.
12. Bouzeria, H.; Ghenim A.N.; Khanchoul, K. Using artificial neural network (ANN) for prediction of sediment loads, application to the Mellah catchment, northeast Algeria. *Journal of Water and Land Development*, 47:55 (2017) 47-55.
13. Tachi, S.E.; Ouerdachi L.; Remaoun, M.; Derdous, O.; Boutaghane H. Forecasting suspended sediment load using regularized neural network: Case study of the Isser River (Algeria), *Journal of Water and Land Development*, 75:81 (2016) 75-81.
14. ANRH. Application of a salinity calculation methodology to the Tafna watershed, (1996) report 1.
15. Terfous, A. ; Megnounif, A.; Bouanani, A. Study of solid transport in suspension in the Oued Mouilah (North-West Algeria), *Journal of Water Science*, 14:2 (2001).
16. Kerem, H.C.; Kisi, Ö. Methods to improve the neural network performance in suspended sediment estimation, *Journal of Hydrology*, 3:17 (2006) 221-238.
17. Erol, Keskin, M.; Dilek, T.; Özlem, T. Adaptive neural-based fuzzy inference system (ANFIS) approach for modelling hydrological time series, *Hydrological Sciences Journal*, 51:4 (2006) 588-598.
18. Mellita, A.; Kalogirou, S. Artificial intelligence techniques for photovoltaic applications: A review, *Progress in Energy and Combustion Science*, 34:5 (2008) 574-632.
19. Chen, S.H.; Jakeman, A.J.; Norton, J.P. Artificial intelligence techniques: An introduction to their use for modelling environmental systems, *Mathematics and Computers in Simulation*, 2:3 (2008).
20. El Badaoui, H.; Abdallaoui, A. ; Chabaa, S. Perceptron Multilayers and radial basis function grating for moisture prediction, *International Journal of Innovation and Scientific Research*, 5:1 (2014) 55-67.
21. Vourkas, M.; Papadourakis, G. Effects of segmentation on the discrimination of three mental stages using ANN and different EEG signal representations, *Proc. 4th Int. Con. Neural Networks and Expert Syst*, (2001) 65-68.
22. Heddam, S.; Bermad, A.; Dechemi, N. Modeling of the coagulant dose by fuzzy inference-based systems (ANFIS) application to the Boudouaou water treatment station (Algeria), *Journal of Water Science*, 1:17 (2012).
23. Dreyfus, G.; Samuelides, M.; Martinez, J.; Gordon, M.; Badran, F.; Thiria, S.; Herault, L. Neural Networks-Methodologies and applications. *Editions Eyrolles*, *Paris*, *France*, (2004).
24. Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators, *Neural Networks* (1989) 359-366.
25. Hornik, K.; Stinchcombe, M.; White, H. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, *Neural Networks*, 3:5 (1990) 551-560.
26. Shafaei, M.; Adamowski, J.; Fakheri, F.A.; Dinpashoh, Y.; Adamowski, K. A wavelet-SARIMA-ANN hybrid model for precipitation forecasting, *Journal of Water and Land Development*, Nº. 28 (2016) 27-36.
27. Rajaee, T. Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models, *Science of the Total Environment*, 407 (2009) 4916-4927.
28. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual model. Part 1. A discussion of principles, *Journal of Hydrology*, (1970) 282-290.
29. Zeggane, H. Study of the hydrological behavior of rivers in Algeria, case of the Isser watershed, Algeria. Doctoral thesis, Kasdi-Merbah Ouargla University (2017), 186 p.
30. Larfi, B.; Boualem, B. Solid transport in the Oued Isser watershed impacts the silting up of the Beni Amrane dam (Algeria), Larhyss Journal, n° 05 (2006) 63-73.

Please cite this Article as:
Houari, K., Quantification of salinity using artificial neural networks. Case study of Isser River (Algeria), ***Algerian Journal of Environmental Science and Technology***, **8:3(2022) 2587- 2595**