Artificial intelligence for supervised classification purposes: Case of the surface water quality in the Moulouya River, Morocco

Imad Manssouri 1), Abdelghani Talhaoui 2), Abdellah El Hmaidi 2), Brahim Boudad 3), Bouchra Boudebbouz 1), Hassane Sahbi 4)

1) Moulay Ismail University, National School of Arts and Crafts, Laboratory of Mechanics, Mechatronics, and Command, Team of Electrical Energy, Maintenance and Innovation, Meknes, Marjane 2, BP: 298 Meknes 50050, Morocco
2) Moulay Ismail University, Faculty of Sciences, Water Sciences and Environmental Engineering team, Meknes, Morocco
3) Moulay Ismail University, Faculty of Sciences, Department of Geology, Laboratory of Geo-Engineering and Environment, Meknes, Morocco
4) Moulay Ismail University, Meknes, Morocco

Abstract: From a management perspective, water quality is determined by the desired end use. Water intended for leisure, drinking water, and the habitat of aquatic organisms requires higher levels of purity. In contrast, the quality standards of water used for hydraulic energy production are much less important.

The main objective of this work is focused on the development of an evaluation system dealing with supervised classification of the physicochemical quality of the water surface in the Moulouya River through the use of artificial intelligence. A graphical interface under Matlab 2015 is presented. The latter makes it possible to create a classification model based on artificial neural networks of the multilayer perceptron type (ANN-MLP).

Several configurations were tested during this study. The configuration [9 8 3] retained gives a coefficient of determination close to the unit with a minimum error value during the test phase.

This study highlights the capacity of the classification model based on artificial neural networks of the multilayer perceptron type (ANN-MLP) proposed for the supervised classification of the different water quality classes, determined by the calculation of the system for assessing the quality of surface water (SEQ-water) at the level of the Moulouya River catchment area, with an overall classification rate equal to 98.5% and a classification rate during the test phase equal to 100%.

Keywords: artificial intelligence, environment, supervised classification, the Moulouya River, water quality

INTRODUCTION

Water is an essential element for life and vital for countless human activities. However, the already scarce water supply is subject to a continuous increase in demand due to a rapid growth in population, the improvement in standards of living, industrial development, and the extension of irrigated agriculture. These pressures on water resources are accompanied by increasingly severe deterioration in their quality.

Therefore, it is essential to quantify and analyse the quantity and quality of surface water and find ways to manage these resources to ensure their sustainability.

The main objective is to suggest a supervised classification model, namely artificial neural networks of the multilayer
perceptron type (ANN-MLP). The model will assess the overall quality of the surface waters of twenty-two sites of the Moulouya River in the Northeast of Morocco. The first sources upstream of the wadi in the High Atlas Mountains to the mouth in the Mediterranean about 600 km in length from March to August 2014.

**MATERIAL AND METHODS**

**STUDY AREA**

The Moulouya watershed covers almost the whole of eastern Morocco with an area of 55,500 km². Geographically it is located between latitudes 32°18’ and 35°8’ North and longitudes 1°11’ and 5°37’ West and has an elongated shape of general direction ENE-WSW (Fig. 1).

This watershed is bounded to the northwest by the Mediterranean coastal basins, to the west by the Sebou River watershed, to the southwest by the Oum Er-Rbia River watershed, to the South by the Ziz River watershed, to the Southeast by the Guir River watershed and to the East by the territory of Algeria.

The Moulouya River, about 600 km long, takes its source at Alemsid at an altitude of 1170 m at the junction of the Middle Atlas and High Atlas Mountains and flows into the Mediterranean at Saidia (Ras El Ma). From upstream to downstream, the watershed is subdivided into three sub-basins: the Upper, the Middle and the Lower Moulouya. Its river receives three important tributaries: the Ansegmir in the Haute Moulouya then successively, the Melloulou River and the Zal River in the Lower Moulouya (Fig. 1).

The great extent and diversity of the reliefs of the Moulouya watershed reflects the climate’s variation from one region to another. These climate changes vary from a Mediterranean type, in the Lower Moulouya valley, to a relatively cold continental climate in the Upper Moulouya and sub-arid Saharan tendency in the high plateaus of the Middle Moulouya [RIAD 2003].

**USED DATABASE**

The data used within the framework of this study includes data relating to the physicochemical analysis of surface waters of the Moulouya River [TAYBI et al. 2016] and the calculation of the quality index by the SEQ-Water quality assessment [TALHAOUI et al. 2020]. The water samples were taken at 22 stations (Fig. 2) over six months, during three sampling campaigns carried out successively during the months of March–April, May–June and July–August 2014. The database, therefore, comprises of a total of 66 samples.

The dependent variables of the supervised classification model based on artificial neural networks of the multilayer perceptron type (ANN-MLP) include the potential of hydrogen pH, dissolved oxygen O₂-diss, electrical conductivity (EC), temperature (T), measured (in situ) in the field, sulphates (SO₄²⁻), biological oxygen demand after five days (BOD₅), orthophosphates (PO₄³⁻), ammonium (NH₄⁺) and nitrates (NO₃⁻) analysed in the laboratory [TAYBI et al. 2016]. According to the analysis methods recommended by AFNOR standards [AFNOR 1997] and RODIER et al. [1996], these parameters were determined.

**METHODS OF THE SUPERVISED CLASSIFICATION**

Over the past decade, the use of artificial neural networks (ANN) has grown in many disciplines (economics, ecology and environment, biology and medicine, industry, etc.). They are notably applied to solve classification problems [BERRADA et al. 2016; MANSsouri 2009; OUSsANA et al. 2016; TALHAOUI et al. 2016], prediction [AGuiRe-BASuREKO et al. 2006; BOUDAD et al. 2014; LASuLe et al. 2017; LUK et al. 2001; MANSsouri et al. 2014; OUSsANA et al. 2018], for troubleshooting [HARTERT 2010; MANSsouri et al. 2008; Su et al. 2011] optimisation, character recognition [PREVoST 2007]
as well as modelling [Altman et al. 1994; Boudebouz et al. 2014; 2015; Kanagou 2007; Voyant 2011].

STRUCTURE OF AN ARTIFICIAL NEURON

Each artificial neuron is an elementary processor. This processor receives a variable number of inputs from upstream neurons [Uncini et al. 1998]. Each of these inputs is associated with a weight, where $w$ symbolises weight representative of the strength of the connection. In addition, each elementary processor has a unique output, which branches out to supply a variable number of downstream neurons. Thus, a weight is associated with each connection, as illustrated in Figure 3.

Regarding the behaviour of the artificial neuron, there are two phases: The first is usually the calculation of the weighted sum of the inputs ($A_j$) according to the following expression:

$$A_j = \sum_{i=1}^{n} W_{ij} x_i \tag{1}$$

where: $W_{ij}$ is the synaptic weight and $x_i$ the input values.

It is the weighted activation sum that converges to neuron $j$ [El Tarach et al. 2007].

The second phase is from the $A_j$ value, a transfer function $f$ that calculates the value of the state of the neuron, called activation $O_j$, which will be transmitted to downstream neurons. It is equal to:

$$O_j = f \left( \sum_{i=1}^{n} W_{ij} x_i + b_j \right) \tag{2}$$

where: $b_j$ is the neuron bias $j$.

The bias allows flexibility to be added to the networks by adjusting the weight during learning and allowing the threshold of the neuron trigger. It is used in several types of activation functions.

MULTILAYER PERCEPTRON

ARTIFICIAL NEURAL NETWORKS

It is an extension of the perceptron by Rosenblatt [1960], with one or more layers hidden between the inlet and the outlet (Fig. 4). The first layer corresponds to the vector composed of the input data and the last layer to the output vector composed of the values that one seeks to obtain [Cornet 2003]. Each neuron in a layer is connected to all the neurons in the previous and the following layers (except for the input and output layers). There are no connections between cells in the same layer. The activation functions used in this type of network are mainly threshold or sigmoid. The latter also follows supervised learning according to the error correction rule.

Figure 4 illustrates the multilayer perceptron in the case of a single outlet. It should be noted that there is a constant input called bias performing as an affine shift. This bias allows a horizontal shift of the sigmoid of the hidden layer and a vertical shift of the output. Thus, the exit from the network is not necessarily zero when the entries are all zero [Kong 2011].

RESULTS AND DISCUSSION

Our problem will be represented by a black box (Fig. 5). The box has nine inputs representing the independent variables, namely the hydrogen potential ($pH$), the temperature ($T$), the electrical conductivity ($EC$), the dissolved oxygen ($DO$), ammonium ($NH_4^+$), nitrates ($NO_3^-$), sulphates ($SO_4^{2-}$), orthophosphates ($PO_4^{3-}$), and the biological oxygen demand after five days ($BOD_5$); and a single output. The output variable reflects the water quality class, namely the surface water quality evaluation system (SEQ-Water).

It is a question of distributing the results of 66 samples over three clusters which should correspond to:

- $1^{st}$ cluster equivalent to the excellent quality water class containing 20 samples;
- $2^{nd}$ cluster equivalent to the class of good quality water containing six samples;
- $3^{rd}$ cluster equivalent to the average, poor, very poor and non-potable water quality classes containing 40 samples.

The network chosen in our case is a multilayer perceptron type (MLP) network with a single hidden layer. The supervised classification model based on artificial neural networks of the multilayer perceptron type (ANN-MLP) used in this work has been developed using Matlab version 8.5 (R2015a) and represented by a graphical interface (Fig. 6) on an I3 machine PC, 2.4GHz and 3GB of RAM.

In order to determine the architecture of the network to be used, preliminary tests were undertaken. These tests showed that, in order to improve the performance of a model established by neural networks like MLP (multilayer perceptron), it is necessary to modify the architecture of the network, playing mainly on the number of neurons in the hidden layer. For this purpose, we have chosen to work with the Levenberg–Marquardt algorithm as a supervised
learning algorithm and by the non-linear function (sigmoid) as the activation function of the neurons in the hidden layer.

The performance of the models thus developed was evaluated using the study of the mean square error (MSE) and the classification rate.

The results of these tests, represented in Figures 7 and 8, showed that the minimum of the mean square error and the maximum overall classification rate are reached when the number of neurons of the hidden layer $NNHL = 8$. So, we can choose eight neurons for the hidden network layer. After 54 iterations, the desired result is achieved with eight neurons in the hidden layer (Figs. 7 and 8).

Then, twelve learning algorithms, the best known in the literature, were applied:

- resilient backpropagation,
- Bayesian regularisation backpropagation,
- variable learning rate backpropagation,
- Fletcher–Reeves updates conjugate gradient,
- conjugate gradient with Powell/Beale restarts,
- Polak–Ribiere conjugate gradient,
- Levenberg–Marquardt backpropagation,
- scaled conjugate gradient backpropagation,
- BFGS quasi-Newton backpropagation,
- one-step secant backpropagation.

For each of these learning algorithms, the number of neurons in the hidden layer was fixed ($NNHL = 8$), as well as the non-linear activation function of neurons in the hidden layer (sigmoid). The performance was assessed by studying the mean square error and the overall classification rate. The results show the variation of the mean square error as a function of the 12 learning algorithms and the variation of the overall classification rate as a function of the 12 learning algorithms (Figs. 9, 10).

Furthermore, they show that with architecture $[9 \ 8 \ 3]$, the minimum of the mean squared error and the maximum of the overall classification rate are reached when the Levenberg–Marquardt backpropagation algorithm is used (Figs. 9, 10).

These results allow for the conclusion that the best ANN model of MLP type is of architecture $[9 \ 8 \ 3]$. It uses the Levenberg–Marquardt algorithm as the learning algorithm, the sigmoid function, and the linear transfer function as activation functions of the hidden layer and the output layer, respectively.
This network comprises three layers: an input layer made up of nine neurons representing the explanatory variables, a hidden layer where all the calculations for optimising the parameters of the neural networks are made, and an output layer of the network, which represents the water quality classes. Figure 11 shows the architecture of the neural network model thus developed.

Therefore, the validation of the neural architecture [9 8 3] consists of judging its classification capacity using the weights and biases calculated during the learning phase to apply them to another validation/test database.

Figure 12 translates the classification results during the learning phase, which is defined at 70% of all samples (46 samples). It is important to note from the confusion matrix that the first class contains 13 samples (water of excellent quality), the second class includes four samples (water of good quality), and the third class contains 28 samples (grouping together, the medium, bad, very bad and non-potable water).

The classification rate during the learning phase (number of correct classification / total number of elements) of the ANN-MLP model [9 8 3] is 97.8% with reference to the surface water quality assessment system (SEQ-Water). Figure 13 shows 15% of all samples (10 samples) for data validation.

The first class contains five samples of the class of water of excellent quality, the second class includes a sample of the class of
water of good quality, and the third class contains four samples of the class which groups together the qualities of average water, bad, very bad and non-potable.

The classification rate during the validation phase of the ANN-MLP model [9 8 3] is 100% with reference to the surface water quality assessment system (SEQ-Water). Figure 14 shows 15% of all samples (10 samples) for the model test phase.

The first class contains two samples of the class of water of excellent quality. The second class does not include any samples, and the third class contains eight samples of the class which groups together the qualities of average, bad, very bad and non-potable water. Thus, the classification rate during the test phase of the ANN-MLP model [9 8 3] is 100% by referring to the surface water quality evaluation system (SEQ-Water).

The performances of this ANN-MLP supervised classification model are more relevant compared to those relating to the models established by the two unsupervised classification methods $K$-means / $C$-means and the Self-Organising Maps (SOM) [MANSOURI et al. 2020] – see Table 1.

The results of this study open up the prospects of exploiting the ANN-MLP model [9 8 3] as a model for supervised classification of the quality of surface water in the same Moulouya River during future sampling campaigns. It could even be generalised to the level of all Moroccan rivers, provided that the same physico-chemical input parameters are used. Furthermore, this application will avoid the repetitive use of different formulas and the calculation of various water quality indices such as the system for evaluating surface water quality (SEQ-Water), the water quality index ($WQI$), etc.

CONCLUSIONS

This study made it possible to highlight the capacity of the classification model of artificial neural networks of the multilayer perceptron type (ANN-MLP) in the supervised classification of the different water quality classes determined by the calculation of the evaluation system of the quality of surface water (SEQ-Water).
REFERENCES


RIAD S. 2003. Typologie et analyse hydrologique des eaux superficielles à partir de quelques bassins versants représentatifs du Maroc [Typology and hydrological analysis of surface water from a few representative watersheds of Morocco]. Lille. Universite des Sciences et Technologies de Lille, Universite Ibnou Zohr d’Agadir pp. 147. [Access 15.03.2020]. Available at: https://ori-nuxeo.univ-lille1.fr/nuxeo/site/esupversions/e5d351a6-ce4c-4b64-b891-84d85f3d8f02


