*J. Mater. Environ. Sci. 6 (2) (2015) 445-454*
*ISSN : 2028-2508*
*CODEN: JMESCN*



# Comparative study of two stochastic models using the physicochemical characteristics of river sediment to predict the concentration of toxic metals

**Abdelaziz ABDALLAOUI*, Hicham EL BADAOUI**
*Department of Chemistry, Analytical Chemistry and Environment Team,*
*Moulay Ismail University, Faculty of Science, PO Box 11201, Zitoune Meknes. Morocco*

*Received 18 May 2014; Revised 11 June 2014; Accepted 16 June 2014*
** Corresponding author: E-mail: a.abdallaoui@gmail.com ; Tel: +212 6 67 33 36 66*

## Abstract
Environmental models, whether deterministic or stochastic, are based upon the relationships between environmental variables and the characteristics of the dependent variables. Several models have been constructed from linear correlation or multiple linear regression (MLR). This procedure assumes a linear relationship between the variables or their functions, which is rather rare in the case of environmental data. On the other hand, artificial neural network (ANN) systems have recently demonstrated their capacity to process non-linear relationships. Their performance has already been proven in various areas of science, even for modelling environmental processes. In this work, we have studied the predictive capacity of MLR with that of ANN for the estimation of four heavy metals (Cd, Cr, Cu and Pb) concentration from eight physicochemical variables (Organic Matter, Water Content, Fine Fraction, pH, $CaCO_3$, Carbon and Phosphorus in the sediment, and the Suspended Matter in the water column) of the Beht River Basin in Morocco. Performed with the MLR method, the determination coefficients ranged from $r^2$=0.26 for Cd to $r^2$=0.83 for Cr, with intermediate values for Cu ($r^2$=0.55) and Pb ($r^2$=0.67). Trained with the back-propagation algorithm of ANN, we obtained the determination coefficient at $r^2$=0.88 for Cd, $r^2$=0.93 for Cr, $r^2$=0.96 for Cu and $r^2$=0.80 for Pb. These results were much better than those obtained by the MLR method, and they demonstrate the higher predictive performance of ANN. The ANN can thus be a powerful alternative in comparison with the traditional techniques of modelling. Its ability to give good prediction with non-linearly related variables, as is commonly encountered in environmental sciences, opens promising horizons in this research field.

*Keywords :* Multiple linear regression, artificial neural network, Heavy metals, sediment, Beht River, sediment characteristics.

## 1. Introduction
Many studies have shown that the heavy metals determination in sediments can be an instrument of observation allowing the evaluation of the aquatic environment pollution level [1], [2]. In fact, sediments constitute a trap where heavy metals and other pollutants are accumulated [3]. It can thus be considered as a chronic pollution memory [1]. Nevertheless, these metals accumulated in river sediment, can be liberated in some conditions, and consequently be found transported by several ways until the man, to provoke its ominous repercussions [4]. These repercussions are very serious in the case of some metals that are considered as toxic, such cadmium, chromium, copper and lead. They have many biological effects on the man [5], [6], [7] and on the aquatic flora and fauna [8], [9], [10], [11], [12]. However, the heavy metals' concentration and their form in sediments depend largely on their physicochemical characteristics (organic matters, $CaCO_3$, C, N, ...) and the quantity of the suspended matter in the water column [13], [14], [15]. Thus, we have sought to predict contents in four toxic metals (Cd, Cr, Cu and Pb) of sediments of a Moroccan River from their physical and chemical characteristics and the concentration of suspended matters in the water column. Diverse multivariate techniques have been used to investigate how independent variables are related to explain the dependent variable, including several methods of ordination and canonical analysis and univariate and multivariate linear, curvilinear, and logistic regressions. These conventional techniques based notably on multiple regression are able to solve many problems, but show some times serious shortcomings. This difficulty is due to the fact that

relationships between variables in environmental sciences are often non-linear, while methods are based on linear principles. Non-linear transformations of variables (logarithmic, power or exponential functions) allow improving appreciably results, but often remaining insufficient. In general, relationships between environmental variables are non-linear or non-monotonous. The use of techniques based on correlation coefficients is often inappropriate [16]. The artificial neural network, with the error back-propagation procedure, is at the origin of an interesting methodology which could be used in the same field as regression analysis particularly with the non-linear relations [17]. Nevertheless, few applications of this new technology in ecological science were published in contrast with physical or chemical sciences. Since the 1980's there has been an explosive growth in the development and application of neural networks in different fields e.g. physical or chemical or medical domain [18], [19], [20], [21], but the use of artificial neural networks in ecology is not widely spread yet. Some applications have been published, e.g. the use of artificial neural networks for modelling of the greenhouse climate [22], identification of the major goals of underwater acoustics [23], prediction of density of brown trout reeds [24], prediction of density and biomass of trout [25], [26], prediction of the penetration of wild boar into cultivated fields [27], prediction of phytoplankton production [28], prediction of production/biomass (P/B) ratio of animal populations [29]. Artificial neural networks may be applied to different kinds of problems, e.g. pattern classification, interpretation, generalization or calibration. In this paper, neural networks are used for a multiple regression problem. The aim of this study is to analyze the level of relationships between some chemical variables of the sediment and the four metal concentrations, but also to propose the basis of the development of predictive tools of metal concentrations using neural network methodology.

## 2. Site study

The Beht River (Fig. 1) has an important role, because its waters are retained in the El Kansera impoundment. These waters are used for the irrigation and for the production of drinking water for all the region of Khemisset and Tiflet cities (Center - North of Morocco). For our study, we have chosen four sampling sites, located upstream of the El Kansera impoundment. The first site is located at Ouljat Sultan (OS). The second is located at approximately 65 km of the first, under the bridge of the Beht River on the road R.P. N°1 (PB), between Khemisset and Meknes. The third station is chosen on the D'kor River just before its confluence with the Beht River (OD). The fourth site is located on the Beht River, approximately 300 m after the confluence of this river with the D'kor River (AB).

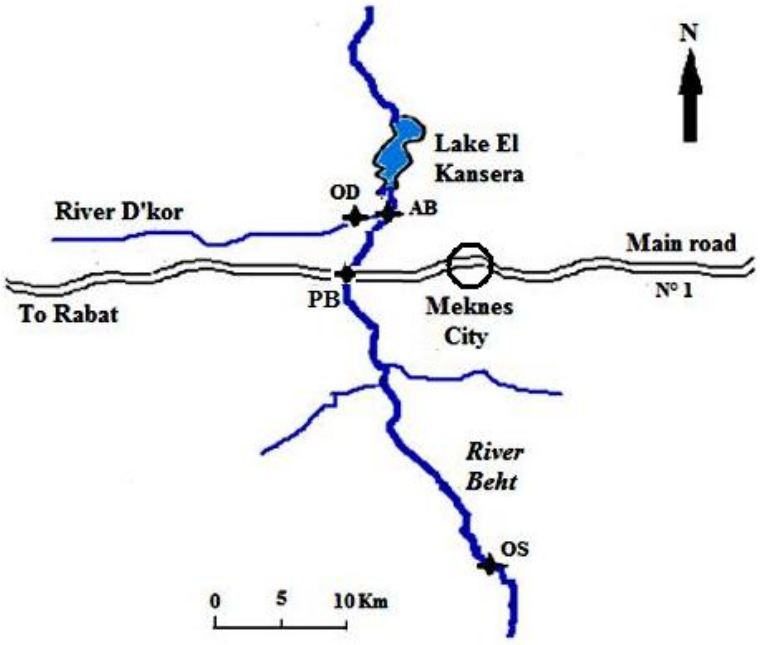


**Fig. 1.** Studied area with location of different stations

Thirteen monthly sampling sessions have been undertaken. Water and sediment samples have been collected. On these 52 samples, we have made the following analyses:

- On the one hand, we have studied the physical and chemical characteristics of sediments by determination of: pH, Water Content, Organic matters, Fine Fraction, $CaCO_3$, Carbon and Phosphorus

of sediments and Suspended Matters of the water column. These parameters will be used as independent variables.

- On the other hand, we have determined four metals (cadmium, chromium, copper and lead) concentrations in the collected sediments. These are dependent variables that will allow us to establish relationships with the previous group. These parameters' values as well as their determination methods are summarized in table 1.

**Table 1:** Studied variables table (x: independent variables, y: dependent variables).

| Variable | Type | Signification and units | Measuring method |
|---|---|---|---|
| OM | x | Organic matter (%) | weight loss on ignition at 550 °C during two hours [30] |
| WC | x | Water Content (%) | weight loss on drying at 105 °C during 24 hours [30] |
| FF | x | Fine Fraction (%) | sieved on a 40 µm sieve [30] |
| pH | x | pH | pH electrode (WTW 90) after shaking dry sediment with $H_2O$ [30] |
| SM | x | Suspended matter (mg. $l^{-1}$) | filtration on 0.45 µm Millipore filters [31] |
| C | x | Carbon (%) | Anne method [30] |
| P | x | Phosphorus (µg.$g^{-1}$) | ascorbic acid method after oxidation with potassium persulfate [32], [33] |
| $CaCO_3$ | x | Carbonate (%) | Bernard calcimeter [30] |
| Cd | y | Cadmium (µg.$g^{-1}$) | by flame atomic absorption spectrometry of aqua regia digested sediments [34], [35] |
| Cr | y | Chromium (µg.$g^{-1}$) | idem |
| Cu | y | Silver (µg.$g^{-1}$) | idem |
| Pb | y | Lead (µg.$g^{-1}$) | idem |

## 3. Methodology

*3.1. Classical statistic analysis*

Univariate, bivariate and multivariate analysis of data were performed by the SPSS Software®. The univariate analysis consists in the determination of parametric statistical parameters (mean, standard deviation and coefficient of variation) and non parametric (minimum, maximum, median and quartiles). In the bivariate analysis, we studied the correlation between variables by the Spearman coefficients (values and probabilities of significance at 5 and 1% of confidence levels). In multivariate analysis, the stepwise multiple linear regression procedures were applied. The diagnosis of the studentized residuals (normality and independence) was used to test the validity of the determination coefficients obtained with each of the models [36], [37].

*3.2. Neural network*

For modelling we used a multilayer feed forward neural network [17]. It concerns a mathematical principle of modelling imitating the mode of human neuron functioning transforming the activation into a non-linear type response. A network with back-propagation typically comprises three kinds of neurons layers: an input layer, one or several hidden layers and an output layer consisting of one or several neurons (Fig. 2). All the neurons of a given layer, except those of the last layer, emit an axon to each neuron of the layer downstream. The network is said to be entirely connected. In the majority of cases, to limit the calculation time and especially when the results are satisfactory, a network with a single hidden layer is used. The input layer comprises n neurons coding the n elements of information ($X_1$ ... $X_n$) at the network entry. The number of neurons in the hidden layer is chosen by the user according to the reliability of the results desired. Finally, the output layer comprises a single neuron calculating the value to predict. The adjustable parameters are the connection weights between neurons of two adjacent layers. The parameter adjustment is done by back-propagating the prediction error up to the input layer by using a gradient-like optimization method. Although the states of the input layer neurons are determined by variables at the entry of the network, the other neurons (of the hidden and output layer) have

to evaluate the intensity of the stimulation from the source neurons of the previous layer by the following relationship:

$$\mathbf{a_j} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{X_i}\,\mathbf{W_{ji}}$$

With $a_j$: activation of the $j^{th}$ downstream neuron;
$X_i$ : output value of the $i^{th}$ neuron of the previous layer;
$W_{ji}$: synaptic weight of the connection between the $i^{th}$ neuron of the previous layer and the $j^{th}$ neuron of the current layer.

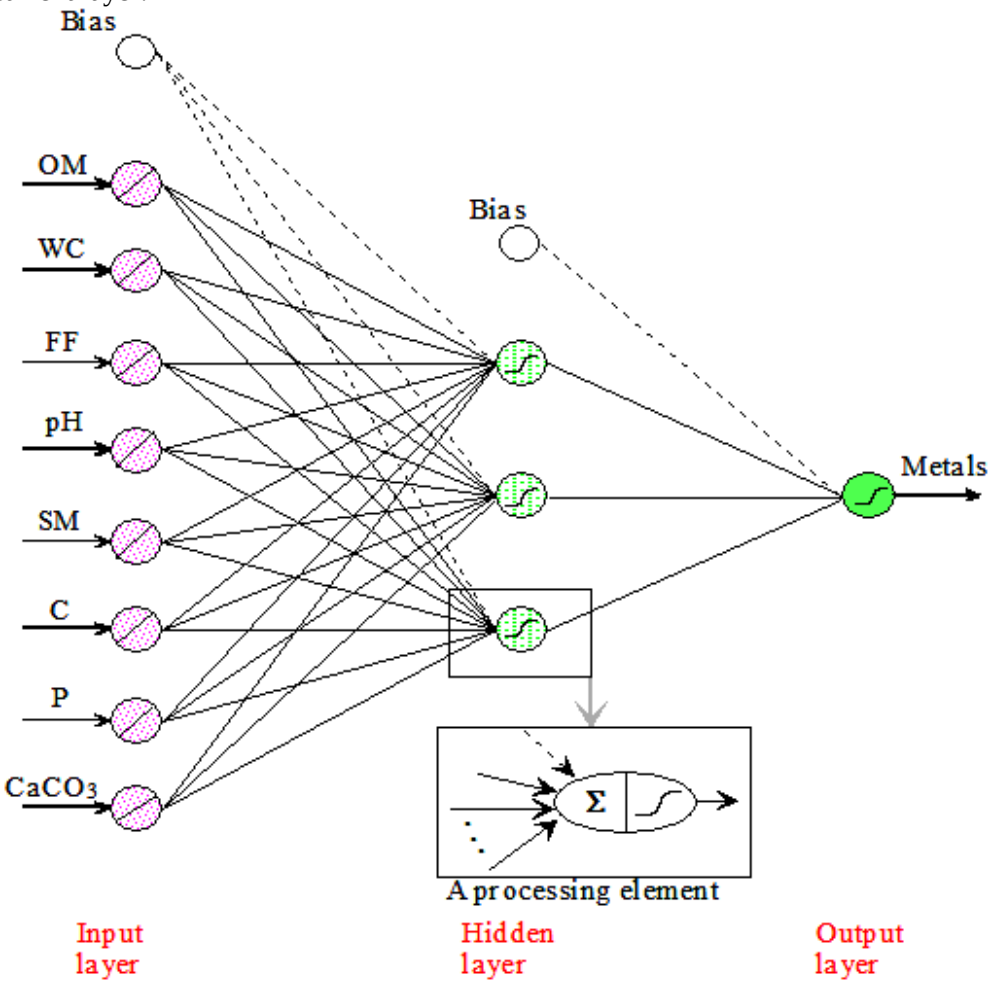


**Fig. 2.** Structure of ANN used.

The ways in which output of hidden and output neurons are calculated from their net input depend on the type of the activation function used in the ANN. Most back-propagation ANNs use the additive sigmoid function, because of its non linearity:

$$f(a_j) = \frac{1}{1 + exp(-a_j)}$$

The technique of back-propagation is related to supervised learning (to learn, the network has to know the response that it should have given). It then modifies the intensity of the connection in order to minimize the error of the considered response. The estimation of the error signal differs according to the layers considered. Many articles, notably by Rumelhart *et al*. [17], Carpenter [38] and Weigend *et al*. [39] detail the algorithms of errors back-propagation. Finally, one can use parameters such as η (learning coefficient) and α (momentum), which serve to accelerate learning while avoiding the network from falling into local minima. The learning of the network continues until minimization of the sum of the errors square (SSE) [40], [41]. Given by the relationship:

$$\mathbf{SSE} = \frac{\mathbf{1}}{\mathbf{2}} \sum_{\mathbf{j=1}}^{\mathbf{N}} \left(\mathbf{Y_j} - \widehat{\mathbf{Y}}_\mathbf{j}\right)^\mathbf{2}$$

With : $\mathbf{Y_j}$: Expected value at the output of the network ("Theoretical value" ) ;

$\widehat{\mathbf{Y}}_\mathbf{j}$: Calculated value by the network (neuron of the output layer) ; j = 1..N : number of recordings.

The computing programs were performed under Matlab® environment for Windows®, the matricial calculation software.

*3.3. Data scaling*

Input data have orders of magnitude that differ greatly according to the variables. In order to standardize the measurement scales, inputs are converted into reduced centered variables by the relationship:

$$\mathbf{X_s} = \frac{\mathbf{X_C} - \overline{\mathbf{X}}}{\boldsymbol{\sigma}_\mathbf{X}}$$

With $\mathbf{X_s}$: standardized values, $\mathbf{X_C}$: original values,

$\overline{\mathbf{X}}$ and $\boldsymbol{\sigma}_\mathbf{X}$ mean and standard deviation of the variable under consideration.

The dependent variables are also converted in the range [0 ; 1] to adapt it to the demands of the transfer function used (sigmoid function) by the relationship :

$$\mathbf{Y_n} = \frac{(\mathbf{Y} - \mathbf{Y_{min}})}{(\mathbf{Y_{max}} - \mathbf{Y_{min}})}$$

With $\mathbf{Y_n}$ : standardized values, $\mathbf{Y_i}$ : original values,

$\mathbf{Y_{min}}$ : Minimal value, $\mathbf{Y_{max}}$: maximal value.

*3.4. Modelling techniques*

The goal of our study is to determine the performance of artificial neural networks (ANN) method. The traditional method for the kind of problems studied here, multiple linear regression (MLR) was first undertaken as reference to see if one can obtain a good prediction performance. Modeling was carried out in two steps : firstly, application was made on the whole of the database (52 patterns) ; secondly, to assess the predictive quality of the models, the two procedures were applied to a set of 45 units randomly chosen in the total database to obtain multivariate models with regression analysis or for neural network training. After the training phase, we presented remaining patterns of 7 new data points to test the multivariate models or the neural network. To evaluate the quality of the results obtained in MLR and in ANN, the coefficient of determination ($r^2$) between observed and estimated values was used.

## 4. Results and discussion

*4.1. Preliminary processing*

Table 2 gives the statistical characteristics of the different variables involved. One can note a very great variation of SM (suspended matter) and FF (fine fraction), characterized by a large standard deviation, and therefore a high coefficient of variation (CV>100%).

**Table 2:** Statistical parameters of different variables studied (STDEV: standard deviation, $Q_1$, $Q_3$: first and third quartile, CV: coefficient of variation in percentage).

| Variable | OM | WC | FF | pH | SM | $CaCO_3$ | C | P | Cd | Cr | Cu | Pb |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean | 3.57 | 27.65 | 23.60 | 8.67 | 268.7 | 17.93 | 1.82 | 714.7 | 1.57 | 36.41 | 53.07 | 34.69 |
| Median | 2.88 | 23.98 | 12.09 | 8.76 | 78.8 | 14.63 | 1.57 | 566.5 | 1.08 | 33.50 | 32.77 | 26.3 |
| STDEV | 2.36 | 11.09 | 25.23 | 0.44 | 649.1 | 11.25 | 1.24 | 543.6 | 1.22 | 16.04 | 67.83 | 27.26 |
| Minimum | 0.32 | 14.10 | 1.35 | 7.56 | 10.0 | 1.20 | 0.29 | 332.0 | 0.25 | 19.00 | 5.50 | 2.30 |
| Maximum | 11.43 | 73.25 | 89.58 | 9.75 | 3269.5 | 42.50 | 6.15 | 4144.0 | 5.00 | 132.00 | 361.39 | 146 |
| $Q_1$ | 2.05 | 21.50 | 5.35 | 8.31 | 50.0 | 8.87 | 0.86 | 469.2 | 0.82 | 29.25 | 21.00 | 19.40 |
| $Q_3$ | 4.44 | 28.52 | 33.97 | 8.97 | 161.9 | 25.46 | 2.39 | 770.2 | 1.76 | 40.50 | 61.66 | 39.15 |
| CV | 66 | 40 | 107 | 5 | 242 | 63 | 68 | 76 | 77 | 44 | 127 | 78 |

The other independent variables fluctuate less, especially the pH whose coefficient of variation is only 5%. As for dependent variables, we observe that Cd is weakly represented with respect to the other metals. Cu is the

most concentrated in our samples with the largest variations (CV=127 %). The study of relationships between these metals taken two by two showed a very strong correlation for Pb and Cu (r = 0.92, n=52, p<0.001). Concerning the other correlations, values are rather weak (|r|<0.21, n=52, p>0.129). Cu was negatively correlated with the two other metals as indicated in Fig. 3 adjusted with the method of Lowes.
By comparison with independent variables, the correlation with the four studied metals is weak (Table 3). It is not significant for most of them, except between Pb and $CaCO_3$ (r=-0.52, n=52, p<0.001).

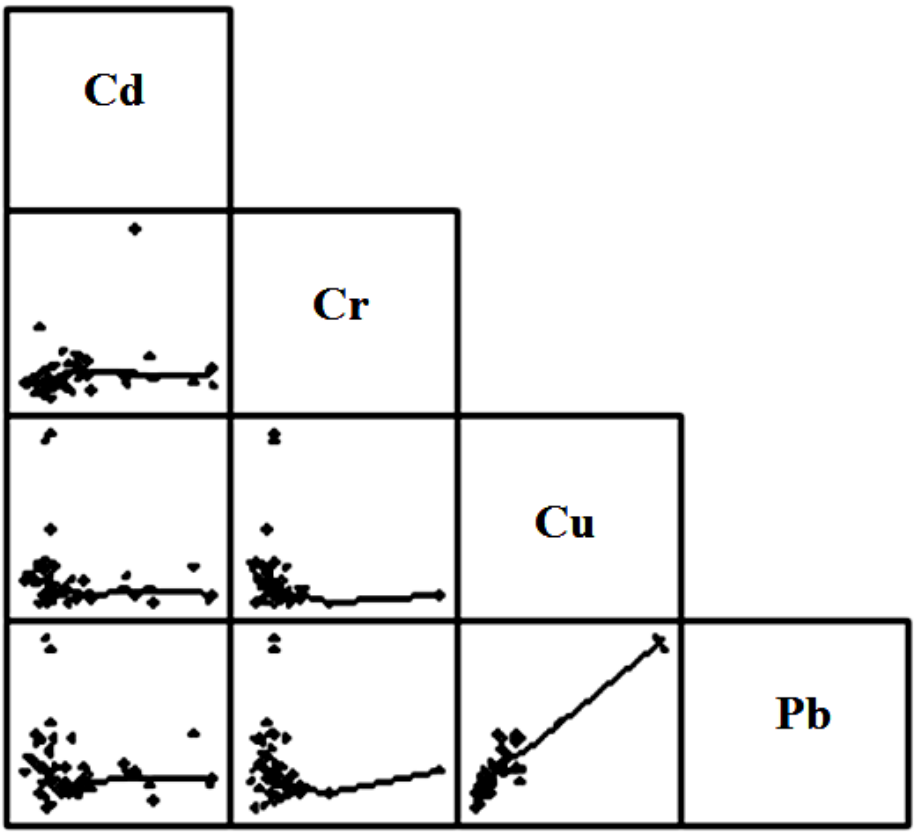


**Fig. 3.** Sploom with Lowes

**Table 3:** Pearson coefficient of correlation between sediment characteristics and the studied four metals concentration.

| Variable | Cd | Cr | Cu | Pb |
|---|---|---|---|---|
| OM | 0.12 | 0.33 | -0.13 | 0.02 |
| WC | 0.06 | 0.46 | -0.18 | -0.06 |
| FF | 0.22 | 0.15 | -0.09 | -0.01 |
| pH | 0.13 | -0.18 | -0.37 | -0.34 |
| SM | -0.15 | -0.20 | 0.08 | -0.07 |
| C | -0.08 | 0.34 | 0.12 | 0.21 |
| P | 0.27 | 0.87 | -0.18 | -0.13 |
| $CaCO_3$ | 0.27 | 0.38 | -0.39 | -0.52 |

*4.2. MLR*

Multiple linear regression modelling was performed in two steps. First, a stepwise regression was carried out to determine the significant variables selected by the models. All the available variables were then entered into the models in order to show the impact of the remaining variables. The following results were obtained for the studied four metals:

**Cadmium**: 2 variables were selected by the model ($CaCO_3$, and FF) with a determination coefficient $r^2$=0.15, F-test=4.57 (df=2, 49), p=0.015. With all requested variables entered, we obtained $r^2$=0.26, F-test=1.9 (df=8.43) ; p=0.084. So, the regression is not significant at the 5% level.
Stepwise regression **$Cd = 0{,}295\ FF + 0{,}341\ CaCO_3$**

All variables **Cd = 0,242 MO - 0,447 TE + 0,414 FF + 0,075 pH - 0,034 MES + 0,244 $CaCO_3$ - 0,147 C + 0,271 P**

**Chromium**: only 3 variables (C, FF and P) were retained by the stepwise model with the determination coefficient $r^2$=0.81 (F-test=61.04 (df=3, 48), p<0.001), whereas the models with 8 variables gave the determination coefficient $r^2$=0.83, F-test=23.9 (df=8, 43), p<0.001. The MLR model is thus highly significant.

Stepwise regression **Cr = -0,159 FF + 0,194 C + 0,864 P**

All variables **Cr = 0,243 MO + 0,001 TE - 0,280 FF - 0,040 pH - 0,067 MES + 0,158 $CaCO_3$ + 0,156 C + 0,742 P**

**Copper**: 3 variables ($CaCO_3$, OM and pH) were selected by the stepwise regression with the determination coefficient $r^2$=0.40, F-test=10.84 (df=3.48), p<0.001. With all the 8 requested variables entered, we obtained $r^2$=0.55, F-test=6.12 (df=8. 43), p<0.001. The MLR model is thus highly significant.

Stepwise regression **Cu = -0,447 MO - 0,419 pH - 0,492 $CaCO_3$**

All variables **Cu = -0,633 MO - 0,579 TE + 0,393 FF - 0,408 pH - 0,023 MES - 0,741 $CaCO_3$ + 0,214 C + 0,428 P**

**Lead**: all of 8 variables were retained by the stepwise regression giving a model that explains 66% of variances ($r^2$=0.66, F-test=8.03, (df=8.43), p<0.001). The MLR model is thus highly significant.

Stepwise regression **Pb = -0,577 MO - 0,483 TE + 0,291 FF - 0,310 pH - 0,176 MES - 0,908 $CaCO_3$ + 0,271 C + 0,467 P**

Except for cadmium, highly significant coefficients of multiple correlation were found for the studied heavy metals, with values ranging from $r^2$=0.40 for copper to $r^2$=0.81 for chromium (p<0.001). The variance explained by the models ranged thus from 40% for copper to 81 % for chromium, with intermediate results for lead (66 %). Compared to the stepwise regression, the models with all the variables (8) improved only slightly the correlation [$r^2$ ranged from 0.55 (55 % of explained variance) for copper to 0.83 (83 % of explained variance) for chromium]. Variables in addition to those selected by the stepwise regression contributed only very weakly to the improvement of the linear model.

### *4.3. ANN*

The comparison between MLR predictive power and ANN is not quite fair, unless the number of parameters (coefficients) of MLR model is almost the same. A MLR was performed here to see if a significant correlation could be obtained with this classical linear method. In order to see the difference of performances, we established ANN models with the variables selected by stepwise MLR, but also with all the available variables. To avoid risks of overfitting, results were examined after 200 iterations and we used a network with 3 hidden neurons for all models. The number of parameters is higher in ANN models compared to those used in the MLR models. In the ANNs with 3 hidden neurons, and in the case where all input variables were used, i.e. the 8 $\rightarrow$3$\rightarrow$1 neural network, we have a total of 27 parameters (8x3+3). The results of ANN models that used the same variables as those selected by stepwise regression gave determination coefficients that were significantly higher than those from MLR, especially concerning Cd (with 2 independent variables retained) for which $r^2$ passed from 0.15 to 0.89, i.e. an improvement of 74 % of explained variance (Table 4). The improvements for the other metals were 7, 55 and 14 % for Cr, Cu and Pb respectively. On the other hand, with all the available variables (8), results obtained by ANN were clearly better than those from MLR (the determination coefficients ranged from 0.80 for Pb to 0.96 for Cu). Cd and Cr had a value of $r^2$ equal to 0.88 and 0.93 respectively). Compared to the MLR models with the same number of variables, improvements with ANN results were 61, 10, 41 and 14 % for Cd, Cr, Cu and Pb respectively. These improvements testify the existence of non-linear relationships between the studied variables. The scatter plots between observed and predicted values are illustrated in Fig. 4. They clearly show that the ANN give, in general, a good fit of all the range of data. Residuals are independent of estimated values for all models: a plot of residuals against predicted values show a random horizontal band (around the horizontal line through 0), which indicates that the models fit well and no obvious outliners exist (Fig. 5).

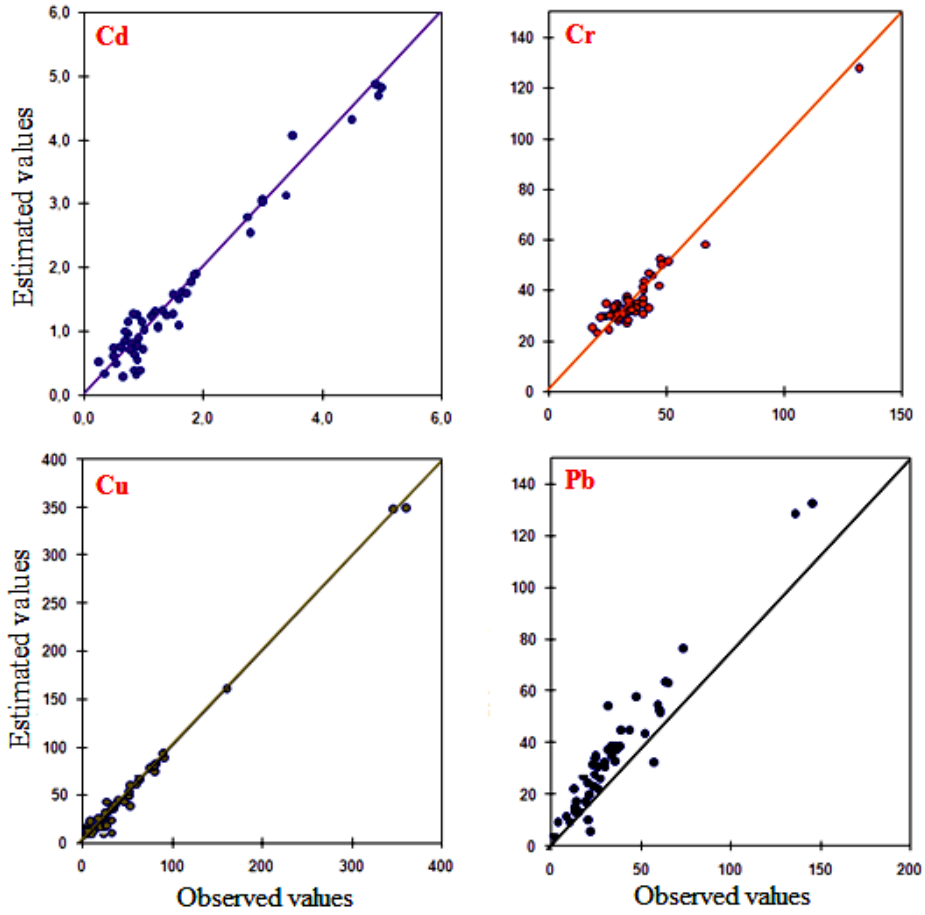


**Fig. 4.** Correlation graph between observed and estimated values of the four variables studied by ANN models.

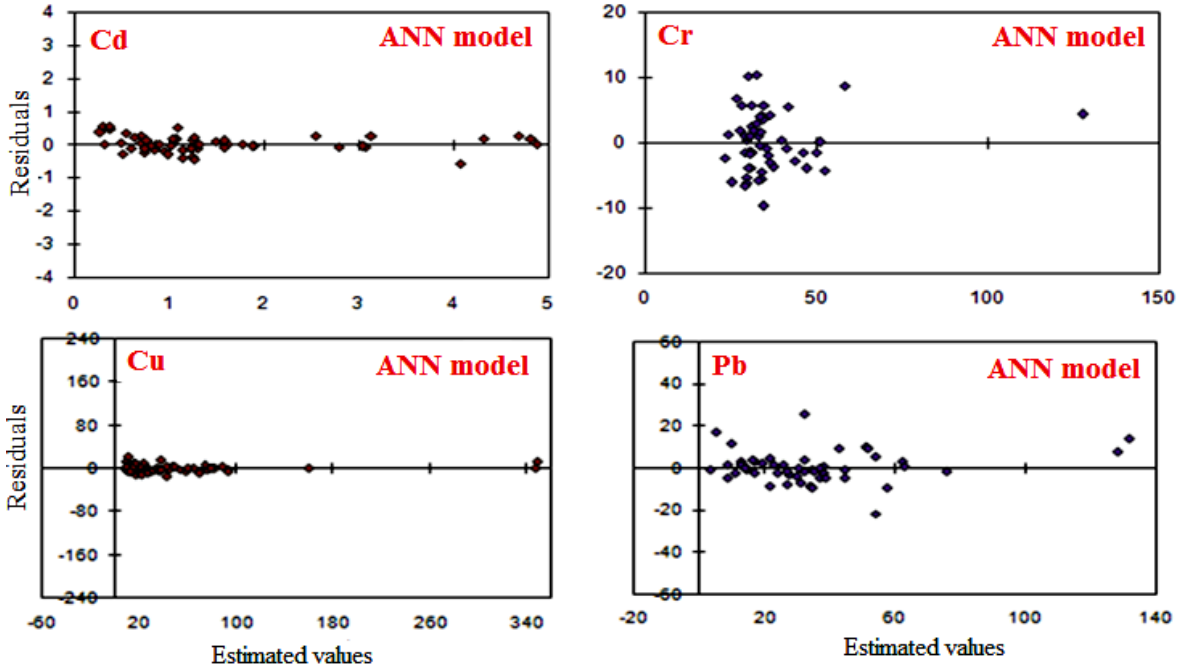


**Fig. 5.** Relationship between the residuals and estimated values of the four metals by ANN models.

### *4.4. Testing the models*

The predictive power of a model can only be judged from new sets of data that did not contribute to its establishment. Therefore, we isolated a set of 45 observations by random selection to set up the model and then tested the predictive performance with the 7 remaining observations. The results obtained from the set of 7 observations allowed us to judge the models quality.

**Table 4:** Results of the MLR and ANN (determination coefficients) on training set and test set for the four metals studied. (ANN1: ANN with the same independent variables that MLR stepwise; ANN2 : ANN with all of the available variables).

| | Cd | | Cr | | Cu | | Pb | |
|---|---|---|---|---|---|---|---|---|
| Method | Training | Testing | Training | Testing | Training | Testing | Training | Testing |
| MLR (Stepwise) | 0.154 | 0.424 | 0.808 | 0.145 | 0.404 | 0.742 | 0.665 | 0.294 |
| MLR all (variables) | 0.263 | 0.146 | 0.834 | 0.141 | 0.547 | 0.579 | 0.665 | 0.294 |
| ANN1 | 0.894 | 0.803 | 0.874 | 0.8721 | 0.954 | 0.922 | 0.800 | 0.846 |
| ANN2 | 0.876 | 0.810 | 0.933 | 0.858 | 0.960 | 0.966 | 0.800 | 0.846 |

It was seen that ANN models had a clearly higher prediction performance on the new sets of data as compared to the MLR. The determination coefficients were indeed clearly higher in ANN (Table 4). High values of $r^2$ in ANN result from better adjustments of the values over the whole range of all the studied variables. In MLR, although the prediction is good for copper ($r^2$=0.74), it is clearly less good for cadmium ($r^2$=0.42) and frankly bad for lead ($r^2$=0.29) and chromium ($r^2$=0.14). These poor results are especially the consequence of the negative predictions of certain values, causing thus large errors. In ANN with the transfer function used (sigmoid function), this type of problem cannot occur; the prediction is very good for all metals ($r^2 > 0.80$).

## Conclusion

Through this study of relationships on four metals and some chemical environmental variables, ANN can be seen as a powerful predictive alternative in comparison to the traditional multiple linear regression. It appears to be promising for metal concentration prediction problems in rivers sediments. Multiple linear regression is a predictive modelling processes commonly used in the determination of empirical relationships between various interacting factors in environmental sciences and in other scientific areas. It is simple to implement if the relationships between the variables or functions of the variables are linear. However, if they are non-linear, a preliminary transformation operation on the variables (often complex and circuitous) is necessary. It aims to minimize the non-linearity in the process. Variables can also be combined or eliminated some of them in order to reach a better predictive model. Despite all these transformations, results are sometimes non-optimum; MLR often supplies negative value predictions, and the residuals are not independent of the variable to be predicted. The modelling process in back-propagation ANN constitutes a new and alternative approach in environmental sciences. It is more direct (no necessity to specify a mathematical relationship between the input and the output variable) and able to work with non-linearly related variables. ANN can be effective for analysing a system containing a number of variables, to establish patterns and characteristics that were not previously apparent. Moreover, it can generalize correct responses that only broadly gather the data in the training set. During training, low connection weights are assigned to irrelevant input variables. In ANN, quantitative as well as qualitative information can be considered. In fact, they do not set constraints on the variables (e.g. normality and/or non-linear relationships), better still, they can be more efficient working with non-transformed data [25]. With actual computer is power, they are also relatively easy to perform. Unlike MLR, ANN do not provide simple equations for users but it is possible to easily quantify the contribution of each variable over its ranges [24], [25], [42]. As new data become available, the neural network model can be readily updated by retraining with patterns which include these new data. ANN models can be readily implemented in software on microcomputers via any high-level programming language, or using one of the increasingly available and inexpensive neural network toolkits in Internet.

## References

1. Lesouef A., Belamie R., Montiel A., *Journal Français d'Hydrologie*, 10 (1972) 165-172.
2. Mantei E. J., Coonrod D. D., *Environ Geol Water Sci*, 13 (1989) 51-58.
3. Mantei E. J., Foster M. V. D., *Environ Geol Water Sci,.*, 18 (1991) 95-104.
4. Charles M. J., Hites R. A., *American Chemical Society*, Washington, DC, (1987) 558.
5. O.M.S., *Organisation Mondiale de la Santé, Rapports et études EURO 61*, Copenhague, (1980) 88.
6. Amiel J., - In : Pascal P., *Masson CIE*, Ed. Paris, (1959) 33-413.
7. Haguenoer J. M., Furon D., *Technique et documentation Ed*., Paris, (1982).

8. Abdallaoui A., *Éditions Edilivre*, (2013) 114.
9. Levesque L., *A.F.E.E., Ed. Rapport* n° 4, (1978).
10. Levesque L., *A.F.E.E., Ed., Rapport* n° 3, (1985).
11. Poirel A., *Mémoire D.E.A.*, Ecole Nationale Supérieure de Géologie, Nancy, (1983) 39.
12. Franchini R., Mutti A., Cavatorta E., Pedroni C., Borghetti A., *C.E.C. Ed*., (1984) 35-51.
13. Abdallaoui A., *Editions Universitaires Européennes*, (2013) 341.
14. Horowitz A. J., Elrick K. A., *Heavy Metal Environment, Inter. Conf*., 2, 4 (1987) 227-229.
15. Guillemet M., *Thèse 3 ème Cycle*, Toulouse, (1988).
16. Prohic E., Juracic M., *Environ. Geol. Sci*., 13, 2 (1989) 145-151.
17. Ter Braak C. J. F., *Ecology*, 67 (1986) 1167-1179.
18. Rumelhart D. E, Hinton G. E., Williams R. J., *Nature*, 323 (1986) 533-536.
19. Smits J. R. M., Breedveld L. W., Derksen M. W. J, Katermam G., Balfoort H. W., Snoek J., Hofstraat J. W., *Analytica Chimica Acta*, 258 (1992) 11-25.
20. Lerner B., Guterman H., Dinstein I., Romem Y., *Proceedings of the I. E .E. E. International Conference on Neural Networks, I. E. E. E., Orlando*, (1994) 3540-3545.
21. El Badaoui H., Abdallaoui A., Mansouri I., Lancelot L., *Journal of Hydrocarbons Mines and Environmental Research*, 3 (2012) 31-36,.
22. El Badaoui H., Abdallaoui A., Mansouri I., Lancelot L., *International Journal of Computational Engineering Research*, 3 (2013) 75-81.
23. Seginer I., Boulard T., Bailey B.J., *J. agric. Engng Res*, 59 (1994) 203-216.
24. Casselman F. L., Freeman D. F., Kerrigan D. A., Lane S. C., Magley D. M., Millstrom N. H., Roy C. R., *Proceedings of the I. E. E. E. International Conference on Neural Networks, I. E. E. E., Orlando*, (1994) 3409-3414.
25. Lek S., Belaud A., Baran P., Dimopoulos I., Delacoste M., *Aquat Living Resour*, 9 (1996) 23-29.
26. Lek S., Delacoste M., Baran P., Dimopoulos I., Lauga J., Aulagnier S., *Ecological Modelling*, 90 (1996) 39-52.
27. Baran P., Lek S., Delacoste M., Belaud A., *Hydrobiologia*, 337 (1996) 1-9.
28. Spitz F., Lek S., Dimopoulos I., *J. Biol. Systems*, 4 (1996) 433-444.
29. Scardi M., *Marine Ecol. Progress series*, 139 (1996) 289-299.
30. Brey T., Jarre-Teichmann A., Borlich O., *Mar, Ecol, Prog; Ser*., 140 (1996) 251-256.
31. Rofes G., *Centre Technique du Génie Rural des Eaux et des Forêts*, (1980) 50.
32. AFNOR, *Association Française de Normalisation*, NF T-90-105 (1978).
33. Murphy J., Riley J. P., *Anal. Chim. Acta*, 27 (1962) 31-36.
34. Bonzongo J. C. J., *Thèse d'Université, U.E.R. Sciences de la Vie et Environnement*, Rennes, (1990), 197.
35. Pinta M., *Masson, Ed*, Tome II, (1979) 696.
36. Arnoux A., Nienchewski L. P., Tatossian J., *Journal Français d'Hydrologie*, 12 (1981) 29-48.
37. Weisberg S., *John Wiley and Sons*, New York, (1980).
38. Tomassone R., Lesquoy E., Millier C., *INRA actualités scientifiques et agronomiques*, Masson, Paris, 13 (1983).
39. Carpenter G. A., *Neural Networks*, 2 (1989) 243-257.
40. Weigend A. S., Huberman B. A., Rumelhart D. E., *In M. Casdagli and S. Eubank (Eds), Nonlinear Modelling and Forecasting, SFI Studies in the Sciences of Complexity, Proc. XII*, Addison-Wesley, (1992) 395-432.
41. El Badaoui H., Abdallaoui A., Manssouri I., Ousmana H., *Journal of Computer Engineering*, 11 (2013) 66-74.
42. EL Hmaidi A., El Badaoui H., Abdallaoui A., EL Moumni B., *Journal European of Scientific Research*, 107 (2013) 400-413.
43. Goh A. T. C., *Artificial Intelligence in Engineering*, 9 (1995) 143-151.