American Journal of Engineering Research (AJER) 2016
American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-5, Issue-4, pp-99-105
www.ajer.org
Research Paper Open Access

Topological Maps of Kohonen Self-Organization (SOM) Applied To the Study of Sediments Contaminated With Heavy Metals

Naoual Monyr[1], Belgacem Agoubi[2], Mohamed Berrada[3], Abdelaziz Abdallaoui[1], Abdellah El Hmaidi[4]

[1)] University of Moulay Ismail, Faculty of Science, Department of Chemistry, Research Team "Analytical Chemistry and Environment", BP 11201, Zitoune, Meknes, Morocco.
[2)] University of Gabes, Higher Institute of Water Sciences and Techniques, Research Team "Hydrogeological, Geochemistry and Geostatistical Modeling", Zrig University Campus 6072, Gabes, Tunisia.
[3)]University of Moulay Ismail, ENSAM, Laboratory of Mathematical and Computational Modeling, Marjane II, BP 15290, Al Mansour, 50000, Meknes, Morocco.
[4)] University of Moulay Ismail, Faculty of Science, Department of Geology, Research Team "Water Sciences and environmental engineering", BP 11201, Zitoune, Meknes, Morocco.

ABSTRACT: *This work aims to apprehend the history of the metallic pollution in the retaining of the Sidi Chahed dam since its inception in 1997 through the sediments of 4 carrots levied the level the embouchures the main wadis (Wadi Mikkes, embouchure Wadi Mikkes, embouchure Wadi Lmallah and embouchure Wadi Jajouiyne). The classification of data by self-organizing maps Kohonen allowed understanding and visualizing the spatial and temporal distribution of samples. Principal component analysis (PCA) and hierarchical classification of SOM maps (SOM AHC) were also used for validating the obtained results.*
Correlations and relationships between the samples and the variables can be easily visualized using the viewing of planes of components of the self-organizing map. The results have highlighted the dependencies between the different metal elements and the classification of studied sediments into four classes into function of four stations coring and their pollution levels.
Keywords *- Artificial neural networks, self-organizing maps (SOM), Sediments, Carrots, Principal component analysis (PCA), hierarchical classification of SOM maps (SOM AHC), Heavy metals.*

I. INTRODUCTION

Several factors may participate in the enrichment of sediments water bodies into heavy metals. These factors, which find explanation in the relationship of heavy metals with their support, directly depend of the grain size of organic matter and hydroxides. Indeed, the heavy metals have a particular affinity for the fine fraction in which the inert organic matter can be associated with heavy metals by forming organ metallic complexes more or less stable. Also, the reducing conditions observed generally in the dam retaining are manifested the sediment-water interface by the salting out the hydroxides starting at the sediments anoxic to the hypolimnion **[1]**.

Sediment pollution by heavy metals may be imputable of a number of sources, including industrial activities, agricultural and urban, household effluent well as surface water flows. Faced with the scale of these phenomena, several studies were conducted in Moroccan aquatic ecosystems **[2,3,4,5,6]**. For the analysis and interpretation of data gathered by these studies, most of them have used statistical methods such as multiple linear regression (MLR), principal component analysis (PCA) or artificial neural networks supervised. The present work deals with the application of self-organizing maps (SOM) of Kohonen using artificial neural network based on unsupervised learning algorithms for data classification and visualization of the spatial and temporal distribution the content into heavy metal of sediments from 4 carrots levied in the retaining of Sidi Chahed dam in Morocco (Fig. 1). The expected results should lead to a definition of existing relations between the heavy metal content of the sediments studied as a function the situation of the sampling stations and sources of contributions. To confirm the results obtained by SOM, these analysis are completed by a principal component analysis (PCA) and hierarchical classification (AHC-SOM).

II. STUDY AREA

The retaining of Sidi Chahed dam is located on Wadi Mikkes, about 30 km NE of Meknes city and 30 km North West of Fez city, on the main road linking this latter to Sidi Kacem city (Fig. 1). The construction of this dam was intended primarily to feeding the city of Meknes by potable water. Its storage capacity is 170 million m^3. However, since its commissioning in February 1997, the quality of the retaining water has proven unsuitable for consumption because of its high salinity.

III. SAMPLING

Four carrots were levied at the edge of the retaining Sidi Chahed dam at the level mouths of the main wadis. The carrots C1 and C2 respectively belong to the wadi Mikkes and its mouth. C3 belongs to the mouth of the wadi Lmellah and C4 at the mouth of Wadi Jajouiyne (Fig. 1). The carrots C1 and C3 have a length of 30 cm and the carrots C2 and C4 have a length of 35 cm. The four carrots were cut out into slices of sediment in 5 cm in thickness forming a database of 26 samples **[7]**.

- Carrot C1: contains 6 samples numbered from 1 to 6;
- Carrot C2: contains 7 samples numbered from 7 to 13;
- Carrot C3: contains 6 samples numbered 14-19;
- Carrot C4: contains 7 samples numbered 20-26;

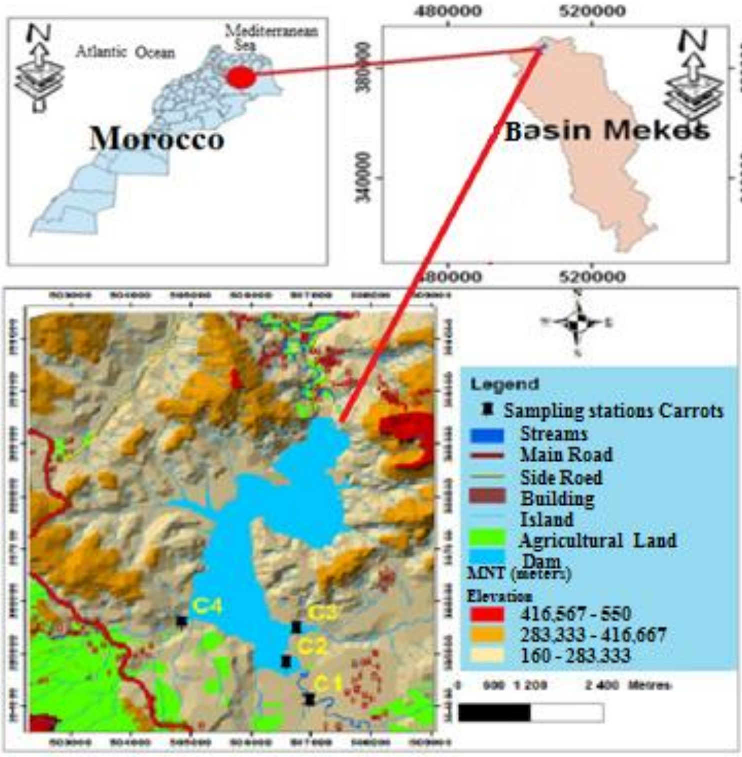

Figure 1. Geographic situation of the retaining Sidi Chahed dam and position of sampling stations.

IV. TOOLAND STATISTICAL DATA BASE

The Artificial Neural Network (ANN) operates with the same model as the human brain composed of neurons. Each network is an algebraic function **[8]** and each weight connection W between two cells plays the role of a synapse which is the main element of interaction between neurons (Fig. 2).

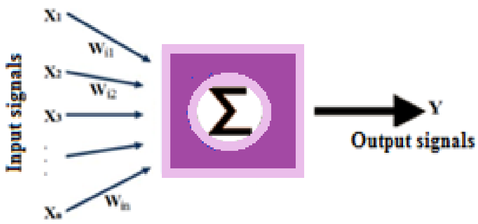

Figure 2. Artificial Neuron multiple input.

Neurons are the elementary cells of the system, capable of transmitting electrical and chemical signals. Hey are organized in networks where they are linked to each other, closely by closely, through synapses. Furthermore, artificial neural networks at self-organization, and in particular self-organizing maps Kohonen are an important group of neural networks. They have received special importance since the work of Von der Malsburg **[9]** and Kohonen **[10]**.

These networks are composed of a grid of neurons (also called nodes); each unit of the grid is connected to the input vector through the N synapse weight W_{ij}. In fact, at each unit is associated with a vector of dimension N that contains the weight W_{ij} **[11,12]**.This assemblage of the neuron in grid presents the self-organizing map of Kohonen (Kohonen Self-Organizing Maps, SOM) which is a network of unsupervised artificial neurons (Fig. 3).

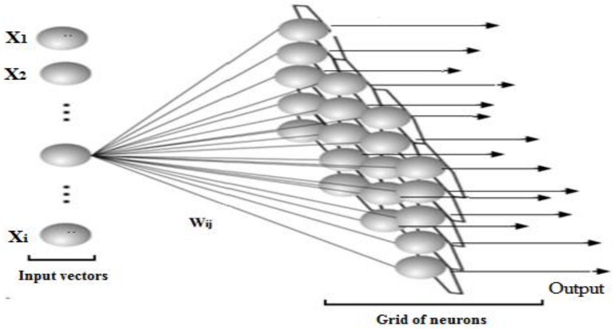

Figure 3. Scheme of a map Kohonen

The algorithm self-organizing map of the Kohonen is one of the oldest models neural. It seeks to project high dimensional data in a low-dimensional space. The principle of this algorithm is to perform a nonlinear classification of complex databases, by identifying similar groups **[13]**. For an unsupervised learning of an artificial neural network, the different steps of the algorithm are:

- Initialize the network with random values of weight W of the input neurons X

At each iteration t:

- Presentation of a learning example X (t), chosen at random, at the entrance of the map
- Determining the winner neuron k whose the weight is the closest to the input data X

$$d_N(X(t), W_k(t)) = \min_i d_N(X(t), W_i(t)) \quad (1)$$

With d_N is the distance in the input space.

- Assessment the neighborhood of the winning neuron in the map

$$g_k(i,t) = g(d(i,k),t) \quad (2)$$

With d is the distance in the card, and g function neighborhood.

- The winner neuron k, the nearest is called "Best Matching Unit (BMU)"; and all the neurons in the neighborhood of the BMU are modified (putting in order and adjustment)

$$W_i(t+1) = W_i(t) + \Delta W_i(t) \quad (3)$$

$$\Delta W_i(t) = \varepsilon(t).g_k(i,t).\big(X(t) - W_i(t)\big) \quad (4)$$

With ε the learning step which decreases in time to allow better adjustment of the weights **[11,14]**.

This algorithm iteratively searches the similarities among the observed data and represents the on output map (Kohonen card), which allows the hierarchy and data visualization **[10,13,15]**.

The normalization of variables before making the SOM is of particular importance. There are three normalization methods that were considered in this work:

- "range", which balance the values variable between 0 and 1 with a linear transformation.
- "var", which normalizes the variance of each variable to the unit and its average to zero,
- "log", which is a logarithmic transformation.

In this study, the input layer (input vector) is constituted of 8 neurons represented by 8 metal tracer element that are Fe, Mn, As, Cu, Zn, Pb, Cd and Cr. These neurons are connected to each 26 neurons that represent the 26 samples studied sediments.

All values in the database have been normalized in the range [0, 1] to adapt to the requirements of the transfer function used by the neural networks. This normalization was carried out according to the relationship:

$$X_n = \frac{(X - X_{min})}{(X_{max} - X_{min})} \quad (5)$$

With:

X_n : Normalized values,
X : Original Values,
X_{min}: Minimum value,
X_{max}: Maximum value.

V. RESULTS AND DISCUSSION

V. 1. Choice of the output card

Learning the Kohonen map uses a sequential algorithm with the Gaussian neighborhood function (Fig. 4) and with different card sizes. The SOM quality obtained with each size and each normalization method is evaluated using two measures as criteria:

- The quantization error (QE) which is the mean to the distances between each input vector and its BMU (Best Matching Unit) or the winner neuron, and consequently, it measures the resolution of the Kohonen map.
- The topographic error (TE) which represents the proportion of all data vectors for which the first and second BMU are not adjacent **[8,10]**.

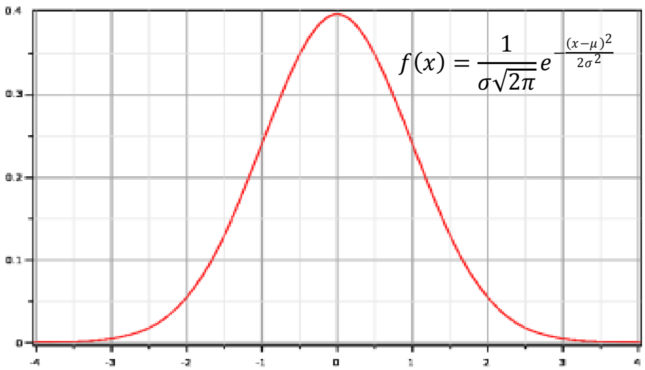

Figure 4. Gaussian function

The application of the three methods of normalization with test of many different card sizes SOM has identified several values for QE and TE parameters (Table I). The lowest values of QE and TE (TE = 0.000 and QE = 0.248) were clearly obtained by using the normalization "range" and correspond to two-dimensional output layer composed of 28 neurons (7 rows × 4 columns).

Table I. Performance Comparison of three methods of normalization with different card sizes SOM.

Size of the SOM card	Type of Normalization					
	Normalization range		Normalization var		Normalization log	
	QE	TE	QE	TE	QE	TE
7x4=28	**0.248**	**0.000**	0.963	0.000	0.981	0.000
8x3=24	0.263	0.038	1.023	0.000	0.978	0.000
5x5=25	0.274	0.000	1.102	0.000	1.027	0.000
9x3=27	0.264	0.038	1.026	0.000	0.940	0.978

V. 2. Classification with SOM card

The plans SOM components of the dataset are shown in Figure 5. The identical colour patterns between the variables correspond to a positive correlation; this can be considered among the variables Cr, Pb, Cu and the variables As, Cd and Fe. The negative correlation can be seen between the two Cd and Pb variables. The other variables are neither positive nor negative correlations in particular those relating to the two metals Mn and Zn that vary independently of the others (Fig. 5).

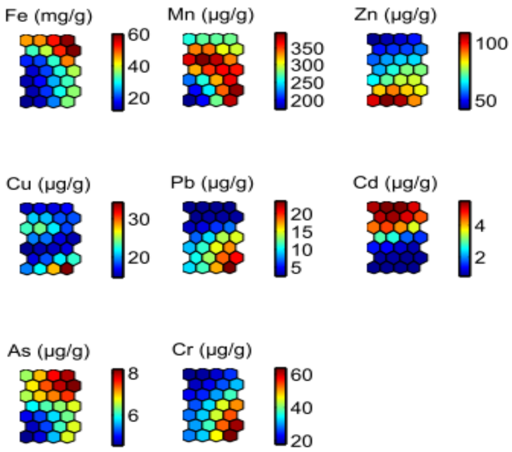

Figure 5. SOM component planes for the eight input variables.

V. 3. Principal component analysis (PCA)

The result of PCA shows the route of score formed by the two components PC1 and PC2, which are considered the most informative since they account for the largest proportion of the variance. In our case, PC1 and PC2 respectively explain 39.8% and 30.7% of the total variance. Figure 6 shows the circle of correlations between variables on the factorial design (PC1 X PC2).

The circle of correlations between variables on the factorial design (PC1 X PC2) shows that the As and Cd elements correlate positively with the PC1 axis with respective coefficients of 0.49, 0.52 and 0.90. However the elements Cr and Pb correlate positively and are negatively associated with the PC1 axis, with respective coefficients of -0.70, -0.86 and -0.32. Thus the Cd and Pb elements correlate negatively.

The other variables have neither positive nor negative correlations and are poorly represented in the circle especially the metals Cu, Fe, Mn and Zn, which vary independently.

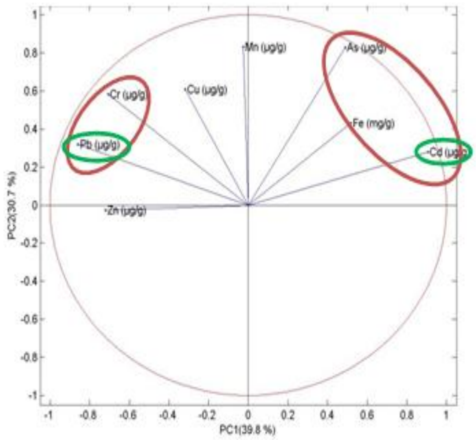

Figure 6. PCA circles of correlation (PC1 X PC2)

V. 4. SOM hierarchical clustering

The dendrogram obtained by the hierarchical classification (CH-SOM), allowed classifying the samples of similar sediments into four classes according to their heavy metal content (Fig. 7 and Fig. 8). Thus, the projection for the samples relative to the four carrots studied on the plan factorial PC1 X PC2 which explaining 70.5% of the variation of the data set shows the presence of four classes (Fig. 9).

The plans components for the eight of input variables (Fig. 5) were a great help in interpreting obtained classes.

In Table II we presented the basic statistical parameters (minimum, mean, maximum) of all data and those of the different classes, relatively to the relevant heavy metals.

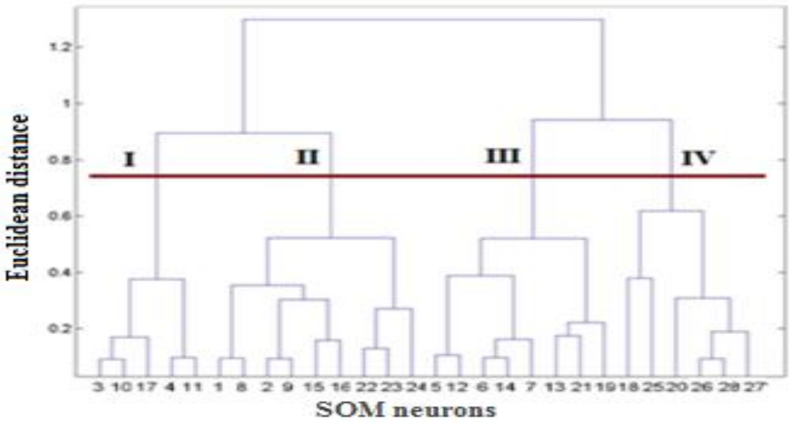

Figure 7. Dendrogram obtained with SOM hierarchical clustering

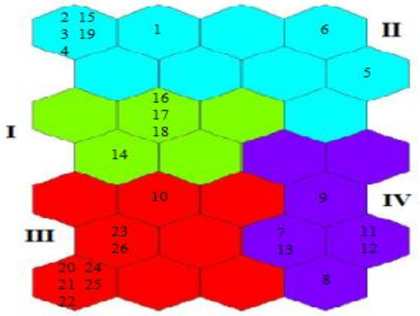

Figure 8. Illustration of the four clusters (I, II, III, IV) obtained by SOM hierarchical clustering.

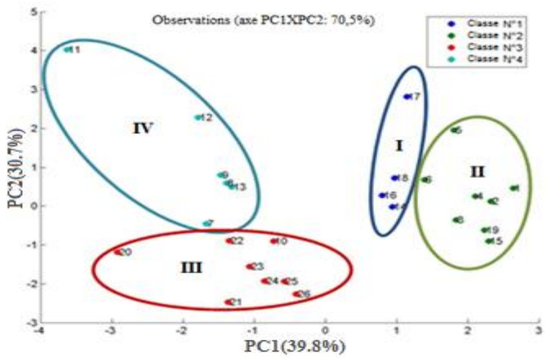

Figure 9. Projection of the samples relative to the four carrots studied on the plan factorial PC1X PC2

Class I: it contains 4 samples, representing most of the stations of the C3 carrot at Wadi Lmellah is mouth. The main feature of this class is seen in the very low values of Pb, Fe and Cr and homogeneous moderately high uniform values of As and Zn. It presents also the highest levels of Cu, Mn and Cd.

Class II: it contains 8 samples carrot C1 taking in Wadi Mikkes and two samples of carrot C3 located in the Wadi Lmellah is mouth. It is characterized by low contents of As, Cu, Cd and Fe and homogeneous moderately elevated Pb, Cr and Mn and has the highest contents of Zn.

Class III: it is composed of 8 samples belonging to the C4 carrot levied at the mouth of Wadi Jajouiyne and a sample coming from the lower tranche of the C2 carrot. It shows generally the low contents of Pb, Cr, and Zn and high of Cd and As, and above all the high contents of Fe are assigned to this class.

Class IV: it contains 6 samples the C2 carrot (mouth of Wadi Mikkes). It is characterized by high contents of Cr, Mn, Cu and Pb and homogeneous moderately low in As, Fe and Zn.

From Table II, we can see that the groups I and II contain the highest concentrations of the metals. Fe, Mn, Cd and As. We can also see that Pb, Cr and Zn show the lowest values in the group II.

Furthermore, group III and IV contain the lowest values of most Fe tracers, Mn, Cu, Cd and As, and the highest concentrations of metals: Zn, Cu, Pb and Cr.

Table II. Basic statistical parameters (Min, Mean, Max) for heavy metals of all data of different classes

			Fe		Mn	Zn	Cu	Pb	Cd	As	Cr
			(mg/g)		**(µg/g)**						
total samples	Min		7.39		120.2	32.9	8.87	2.03	0.44	3.75	11.22
	Mean		31.69		294.7	70.69	19.59	9.08	2.85	6.32	33.25
	Max		67.37		591.6	192.4	63.59	34.31	6.65	9.61	78.82
Class I	Min		11.34		370.75	50.62	24.27	2.31	4.39	6.23	19.18
	Mean		12.65		455.18	62.77	28.99	2.51	5.24	7.26	23.29
	Max		15.35		**591.6**	78.62	36.5	2.81	6.46	9.27	29.48
Class II	Min		48.36		181.86	**32.9**	10.79	**2.03**	4.27	5.87	**11.22**
	Mean		57.21		255.32	46.97	15.65	2.52	5.29	7.37	22.61
	Max		**67.37**		333.68	65.64	21.5	2.93	**6.65**	**9.61**	37.74
Class III	Min		**7.39**		**120.2**	47.06	**8.87**	6.19	0.54	**3.75**	20.9
	Mean		13.24		187.51	88.58	16.96	10.05	0.75	4.69	29.57
	Max		24.10		279.3	**192.42**	29.54	13.89	1.02	5.46	33.99
Class IV	Min		25.04		339.39	64.02	11.42	16.51	**0.44**	4.87	38.05
	Mean		34.96		383.22	83.78	22.07	20.91	0.82	6.47	58.97
	Max		43.23		440.07	96.85	**63.59**	**34.31**	1.14	7.71	**78.82**

VI. CONCLUSION

The self-organizing map SOM was used to study the distribution of trace metals of sediments of the carrots levied from the retaining dam Sidi Chahed at the mouths of wadis.

The comparison with the ascending hierarchical classification (ACH) and Principal component analysis (PCA), showed, therefore that the results obtained with the self-organizing map SOM were generally similar to those of these statistical methods. But, the SOM method has also provided more detailed classification of heavy metals from sediments of the dam Sidi Chahed.

From this study we concluded that the dam Sidi Chahed has had an uneven distribution in space and time of the contents of heavy metals in sediments.

REFERENCES

[1]. A. Abdallaoui, Le phosphore et les métaux lourds contenus dans les sédiments. Leur influence sur les phénomènes d'eutrophisation et de pollution. Éditions universitaires européennes, 2013, 349.

[2]. H. Taouil, S. Ebn Ahmed, N. Hajjaji, and A. Srhiri, Contribution à l'évaluation de la pollution métalliques des sédiments de l'Oued Tislit-Talsint, bassin versant de Guir ((Maroc orietal), Science Lib, Editions Mersenne, 3, 111109 ISSN 2111, 2011, 4706.

[3]. M. Berrada, A. El Hmaidi, N. Monyr, D. Abrid, A. Abdallaoui, A. Essahlaoui, and A. El Ouali, Self-organizing map for the detection of seasonal variations in Sidi Chahed Dam sediments (Northern Morocco), Hydrological Sciences Journal, HSJ-2014-0140, 2016, 2150-3435.

[4]. D. Abrid, A. El Hmaidi, A. Abdallaoui, O. Fassi Fihri, and A. Essahlaoui, Impact de la pollution sur les eaux de l'oued Boufekrane (Meknès-Maroc): Etude physicochimique et bactériologique Physic. Chemical, Journal Physical and Chemical News, 58, 2011, 98-104.

[5]. A. El Hmaidi, D. Abrid, A. Abdallaoui, and A. El Ouali, Caractérisation sédimentologique et physico-chimique des sédiments de carottes de la retenue du barrage Sidi Chahed, NE de Meknès, Maroc, Journal of Hydrocarbons Mines and Environnent Research, 3(2), 2012, 91-96.

[6]. A. Abdallaoui, and H. El Badaoui, Prediction of toxic levels of heavy metals in sediments from the river Beht physico-chemical parameters, Phys. Chem. News, 58, 2011, 90-97.

[7]. D. Abrid, S. Laafou, A. Abdallaoui, A. El Hmaidi, and A. Essahlaoui, Les métaux lourds dans les sédiments de la retenue de barrage Sidi Chahed (région de Meknès, Maroc), International Journal of Innovative Research in Advanced Engineering (IJIRAE), 1(5), 2014, 1-7.

[8]. G. Dreyfus,J.-M. Martinez, M. Samuelides, M. B. Gordon, F. Badran, S. Thiria, and L. Herault, Réseaux de neurones: cartes auto-organisatrices et classification automatique, Eyrolles, 2, 2002 287-335.

[9]. V. Malsburg, Self-organization of orientation sensitive cells in the striate cortex (Kybernetics, 1973).

[10]. Kohonen, T.,Self-organizing maps (3rd ed. Berlin: Springer, 2001).

[11]. S. Khedairia, Contribution à la classification non supervisée : application aux données environnementales, Thèse de doctorat en Informatique, Universite Badji Mokhtar-Annaba, 2014.

[12]. K. Bamory, M. Adama, S. Luc, D. Amidou, B. Hamadoun and G. Droh Lancine, Différenciation hydrogéochimique entre les nappes superficielles des alterites et profondes du socle fissure dans le bassin versant de Kolondieba (sud du Mali): approche statistique par la méthode SOM des réseaux de neurones, Global Change: Facing Risks and Threats to Water Ressources (pr oc. of the Sixth World FRIEND Conference, Fez, Morocco), IAHS Pub, 2010, 340.

[13]. L. Peeters, F. Bacão, V. Lobo, and A. Dassargues, Exploratory data analysis and clustering of multivariate spatial hydrogeological data by means of GEO3DSOM. A variant of Kohonen Self-Organizing Maps, Hydrol. Earth Syst. Sci. Discuss, 3, 2006, 1487-1516.

[14]. Z. H. Bohari, A. Sharin , M. F. Baharom, M. MohdNa'im, M. H. Jali, Feature Analysis of Numerical Calculated Data from Sweep Frequency Analysis (SFRA) Traces Using Self Organizing Maps, Journal Technology, 2014, 37-42.

[15]. Kohonen T., Essentials of the self-organizing map, Journal ELSEVIER, Neural Networks, Finland, 37, 2013, 52–65.