Journal of New Technology and Materials(JNTM)
J. New Technol. Mater.
Vol. 07, N°02 (2017)38-46

OEB Univ. Publish. Co.

Self organizing map of artificial neural network for groundwater quality classification in the F'kirina plain (Oum El Bouaghi province-NE of Algeria)

M. Ouanes [1], B. Houha [1], A. Elhmaidi[2], M. Berrada [3], H. Ousmana[2] and A. Ghedjmis [1]

[1]. *Department of ecology, University of Abbes Laghror, Khenchela 40004, Algéria,*
[2]. *Department of Geology, University of Moulay Ismaïl, Meknès 11201, Maroc.*
[3]. *LM2ILaboratory of Mathematics Modelisation and Informatics, University of Moulay Ismail, 15290, Maroc*

Corresponding author: email: ouanes.miyada@gmai.com

Received date: June 15, 2017; revised date: Dec 06, 2017; accepted date: Dec 16, 2017

Abstract

The topological Self-Organizing Maps of Kohonen and other methods of artificial intelligence are effective tools for modeling and solving environmental problems. In this study, we propose an approach to classify the annual physico-chemical parameters of subterranean waters in the F'kirina plain based on the artificial neural network type. The results obtained demonstrate the presence of 4 classes and make it possible to clearly understand and visualize the spatial and temporal distribution of the physicochemical quality of subterranean waters. Class 1 shows high concentrations for all parameters, whereas class 3 is represented by very low concentrations.

Keywords: Clustering, Self organizing map, ANN, Physico-chemical parameters, Water quality, F'kirina plain.

1. Introduction

The topological Self-Organizing Maps (SOM) are classification algorithms developed by Teuvo Kohonen since 1982 [1]. They are vectorial quantification methods based on unsupervised learning algorithms and grouping information into classes, respecting the topology of the space of observations [2]. The objective of this study is to exploit data related to the physicochemical parameters of subterranean waters by the SOM method to visualize and understand the spatial and temporal distribution of water samples and their physicochemical parameters at the level of the FKirina plain (NE of Algeria). Kohonen's classification was chosen because it provided well-developed interpretation tools to improve the prediction analysis potential, which results in a map that makes it easy to identify neighboring classes and to detect the spatial and temporal variations of studied parameters. In this study, the self-organizing maps of Kohonen are realised from the MATLAB software with a program realized by [3].

2. Studied area

The F'Kirina plain belongs to the Garaet and Taref sub-watershed which represents a part of the Constantine High Plains and mountains of Mellegue, Harectas and Nememchas watershed. It is 900 m above sea level and covers an area of 650 km2.

It is surrounded by mountains whose summits rarely exceed the 1100m, thus clearly dominating the plain and thus giving it the appearance of a basin (Fig.1). It is a plain of immense interest for studies of anthropogenic and climatic effects on the quality of groundwater [4]. It is located at a distance of 55 km to the southeast of the Chief, place of Oum El Bouaghi province, 40 km to the north east of Khenchela and to 18 km of the seat of the sub province of Ain Al Beidha. It is considered fundamentally as a mainly rural agro-pastoral sector [4].

The F'kirina region is drained by two important rivers in the southeast of the township. These are Isfert River and Oulmen River. The geological structure is characterized by calcareous mountains whose tectonic movements are very strong. The climate of this region is semi-arid of cold and rainy continental type in winter, hot and dry in summer with short spring and autumn periods [4].

Figure 1: Location of the F'kirina plain in relation to Algeria map (image from google maps, 2017).

3. Materials and methods

3.1. Data Acquisition

In this work, 97 water points were selected throughout the whole plain to monitoring the evolution of groundwater characteristics during the years 2003, 2008 and 2016. The samples were subjected to 8 physico-chemical analysis.

3.2. Self-Organizing maps (SOM)

The self-organizing map method [5] is a data classification tool, allowing clustering a data space in M classes. This partition is represented by a map (grid often in 2D) [6], containing M nodes (called neurons) connected to each other by edges. The assignment of data to a neuron must ensure homogeneity, but also retain the topology of the map [7]. That means, two neurons that are neighbors on the map must correspond to nearby data on the data space. The map allows to view the data of high dimension (here the dimension is 8) on a map in dimension 2. The SOM algorithm can be seen as a nonlinear projection of a data space on a map.

The SOM map consists of an input layer, representing the presence-absence matrix of individuals (measurement points). The input layer is directly connected to a two dimensional output layer called the map (the output layer) (Fig 2).

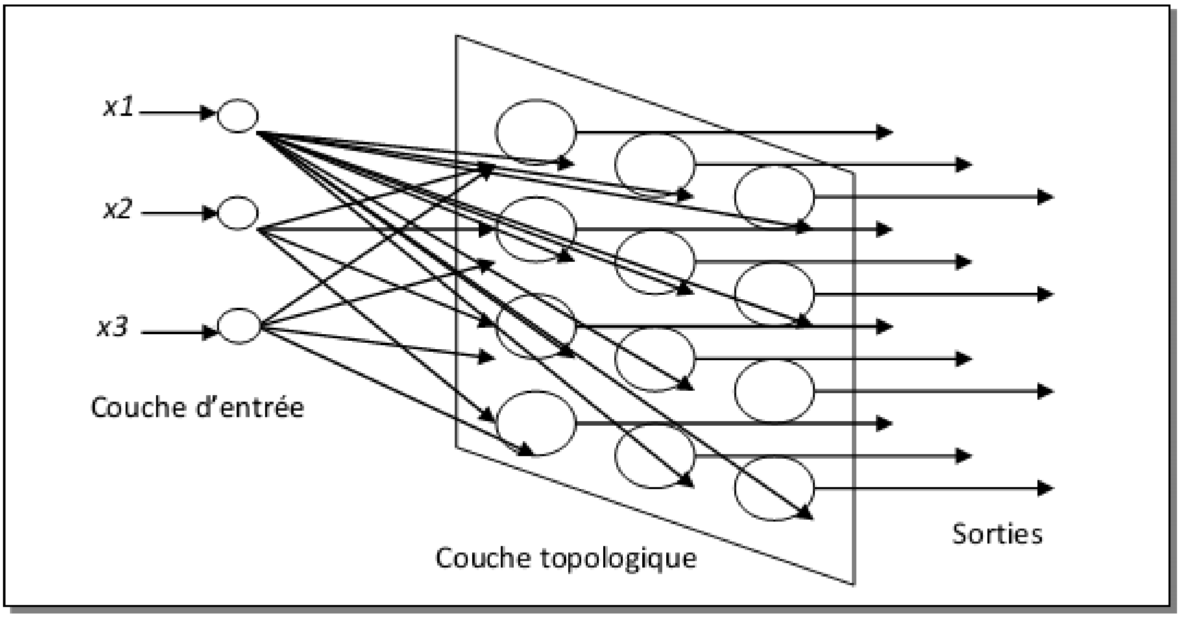

Figure 2: Structure of a topological map.

The SOM map estimates the connection intensities between the input and output layers, using an unsupervised competitive procedure [1]. This procedure iteratively looks for similarities among the observed data and represents them in the output map by neurons. Each neuron of the map is represented by a weight vector in the data space. Finally, the SOM map produces an output matrix with the final weight vectors [5].

The SOM learning is done with the batch algorithm for different sizes and architectures. The optimal size is chosen minimizing the so-called quantization error (QE) and topological error (TE). To validate the SOM classification, the two criteria TE and QE are generally used. According to Kohonen [5], the QE that measures the resolution of the map is the average of the distances between each input vector and its Best Matching Unit (BMU) or neuron winner [8]. The smaller is the value of Qe, the better is the algorithm. The Te describes how well the SOM preserves the topology of the studied data set [9]. It's the proportion of all data vectors for which first and second BMUs are not adjacent neurons (i.e. are not connected with a topological connection) [10, 11, 12]. A small value of Te is more desirable. Unlike the quantization error, it considers the structure of the map.

3.3. Hierarchical Ascending Classification by SOM (SOM-HAC)

Under the hypothesis of a good topological order of the SOM map, it is very likely that two neighboring neurons on the map represent data of the same class. A hierarchical Ascending Classification (HAC) allows find them. The first iteration of the HAC algorithm consists in considering the partition formed by the singletons: each node is then assigned to a distinct subset. In the iteration k and for a given partition, the two closest subsets are sought in the sense of the chosen similarity criterion and are merged to form a single subset.

This process is repeated until a partition containing a single subset is obtained. The result of the algorithm is a tree of classes, called the dendrogram, which shows how the classes are connected [13] (Fig 3).

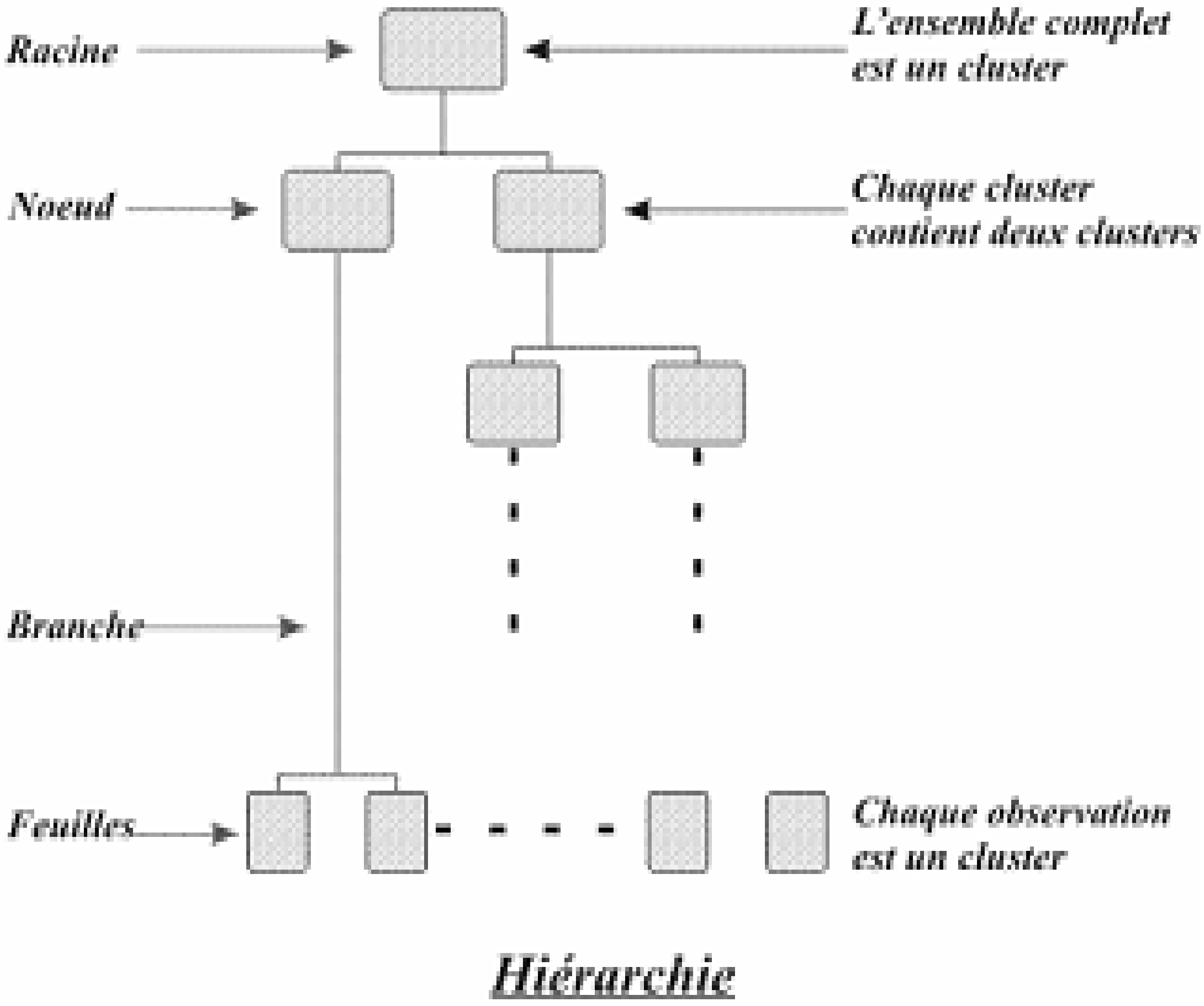

Figure.3: Hierarchy of an unsupervised classification (Clustering) [13].

4. Result and discussion

4.1. Classification of subterranean water samples by SOM

The principle of the SOM algorithm consists in performing a nonlinear classification of complex databases, identifying similar groups [14].

In this work, a topological map of 99 cells (11 lines x 9 columns) was retained for this predetermined analysis, using the Park formula with quantification errors QE = 0.113 and topography TE = 0,01. Each map represents a variable. The distribution maps or component maps from the Kohonen map are shown in Figure 4. They allow to visualise the distribution of groundwater in the Fkirina plain on the basis of their physico-chemical characteristics. The dark red cells represent high values, while the blue (pale) cells represent low values. Consequently, the map allows study the correlations between variables. The values in Fig. 4 are represented with a logarithmic scale.

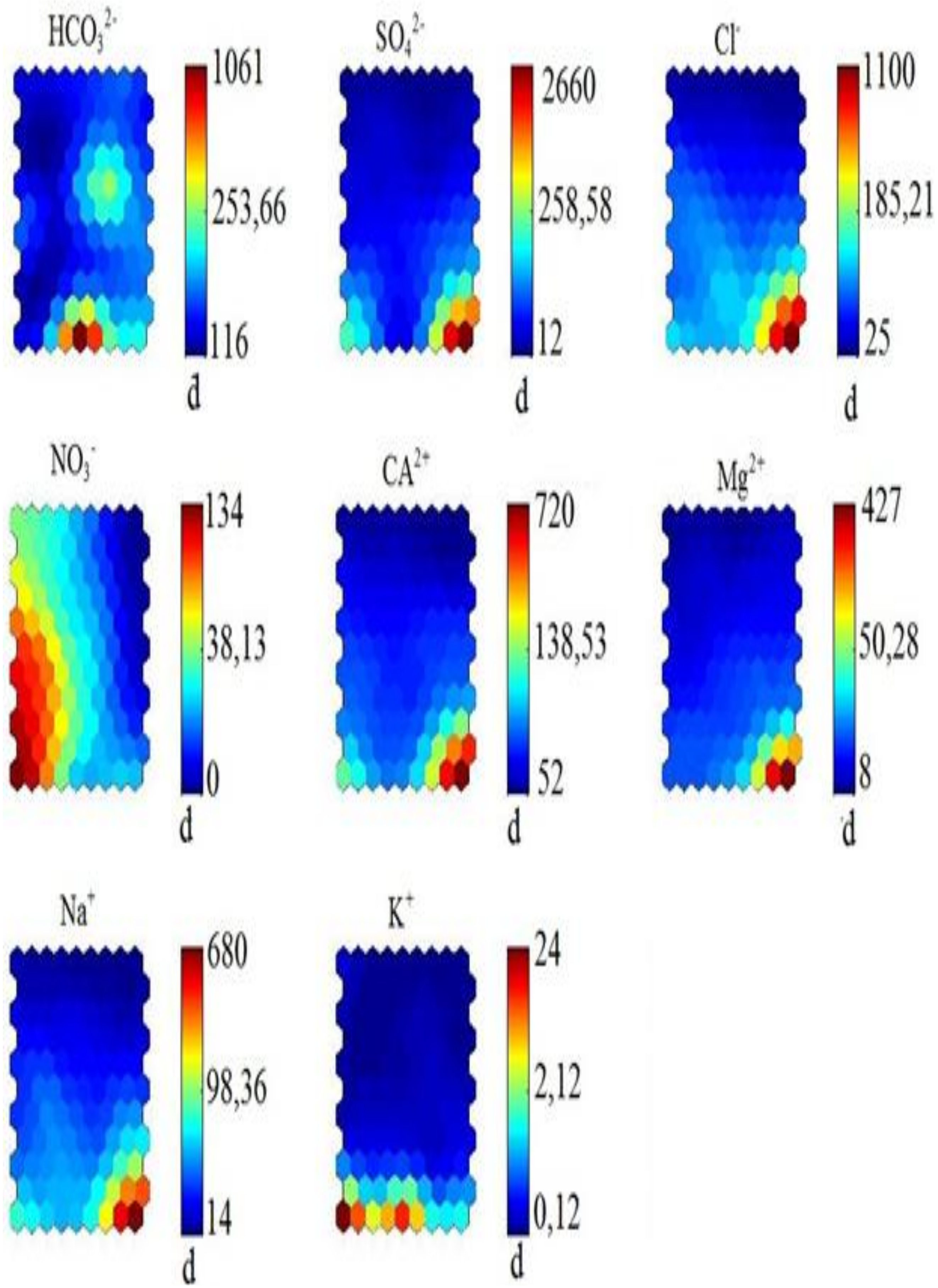

Figure 4: Gradient of values of physicochemical parameters on the Kohonen map.

Table 1: Correlation matrix.

Variables	$HCO3^-$	$SO4^{--}$	Cl^-	$NO3^-$	Ca^{++}	Mg^{++}	Na^+	K^+
$HCO3^-$	1.00							
$SO4^{--}$	-0.01	1.00						
Cl^-	0.08	0.78	1.00					
$NO3^-$	-0.22	0.22	0.29	1.00				
Ca^{++}	0.18	0.89	0.91	0.29	1.00			
Mg^{++}	0.19	0.90	0.74	0.08	0.78	1.00		
Na^+	0.12	0.88	0.92	0.29	0.88	0.81	1.00	
K^+	0.47	0.41	0.36	0.22	0.49	0.36	0.48	1.00

The maps of the different variables show the existence of a good positive correlation between the SO4, Cl, Ca, Mg and Na parameters. Parameter K can be joined to these but with a moderately positive correlation. The remaining two parameters NO3 and HCO3 show no correlation with the others. These results are confirmed by the correlation matrix (Tab. 1).

The correlation matrix expresses the different correlations between the variables analyzed (Tab.1). The correlations observed are totally strong. Indeed, |over 52.94% of the correlation coefficients are greater than 0.74. These strong noticed correlations relate to variables of the same class (Tab.1), namely:

- Sodium Na+ (mg /l) shows positive correlations with significant R coefficients respectively with Cl- (R = 0.92), Ca2+ (R = 0.88), SO42- (R = 0.88) and Mg2+ (R = 0.81).
- Cl- (mg / l) chlorides are positively correlated with sulfates SO42- (R = 0.78).
- Calcium is positively correlated with chlorides Cl- (R = 0.91) and sulfates SO42-R = (0.89).
- Magnesium Mg2+ is positively correlated with sulfate SO42- (R = 0.90) and calcium Ca2+ (R = 0.78) and to a lesser degree with chloride Cl- (R = 0.74).

SOM-HAC classification

Ward's hierarchical ascending classification [15] was then used to group the cells into groups of the samples (Fig.5). The number of classes is then determined on the basis of the chosen Euclidean distance. Thus, a distance of 0.6 leads to the distribution of 6 classes, a distance of 0.8 leads to 4 classes and a distance of 1 makes it possible to distinguish 3 classes of stations.

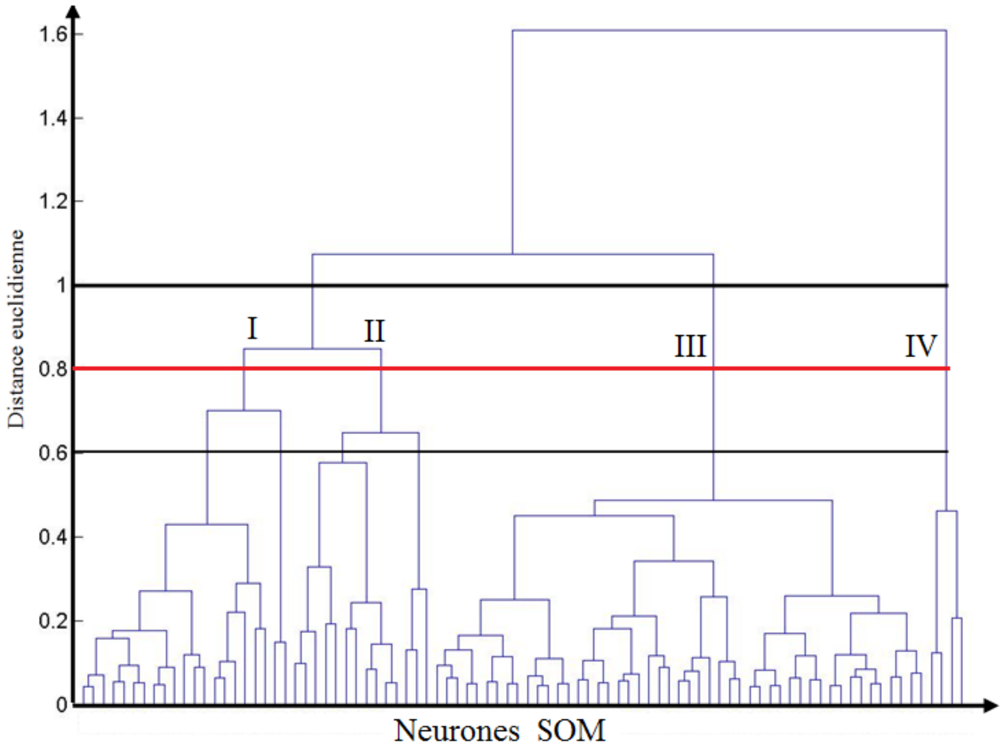

Figure 5: Tree of classification of the physicochemical parameters of subterranean waters of the F'kirina plain with the method of topological maps (SOM).

The choice of the number of classes resulting from this classification is left to the choice of the user: it must therefore be determined according to the objectives of the application and the assessment of the analysis. In the context of this work, we chose to distinguish four (4) classes with a break with the Euclidean distance of 0.8 because it is a good compromise between a restricted number of classes and a good representativeness of the differences of properties. The four classes thus obtained are represented in Fig. 6.

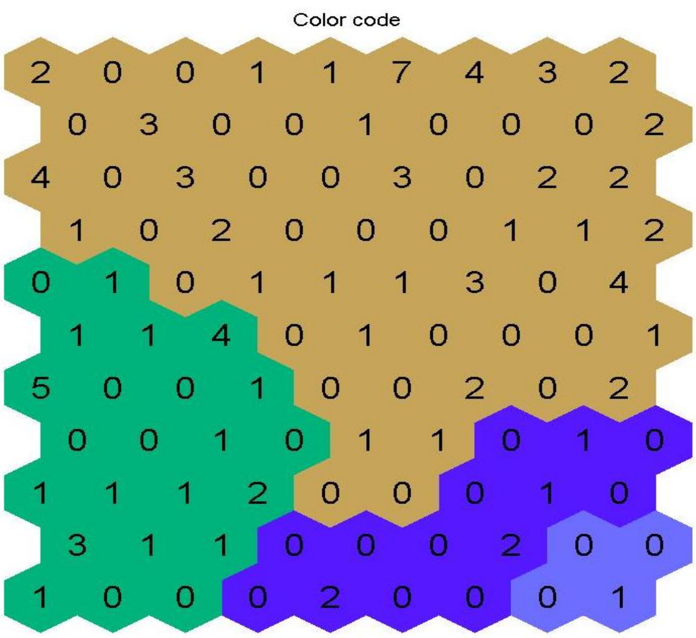

Figure 6: Distribution of the samples on the Kohonen map from the physic-chemical variables of groundwater studied.

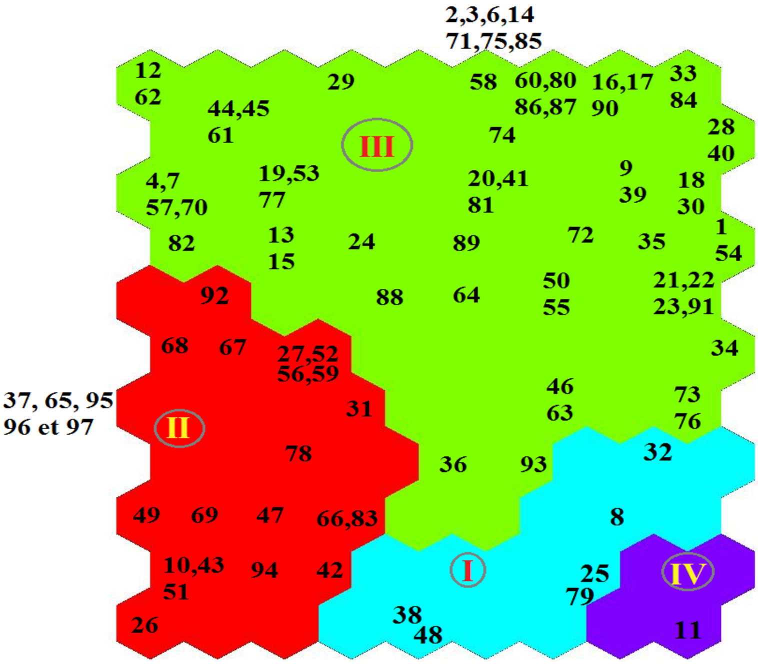

Figure 7: Final distribution of the stations according to the classes on the map of Kohonen from the physico-chemical variables of the F'kirina plain groundwater.

The first class contains 6 samples and represents 6.19% of the total data. It consists of waters with a very high concentration of chemical elements in relation to OMS standards: (Cl- (422.5mg/l), Na+ (215.33mg/l), Ca2+ (277.67mg/l), SO4--(492.83mg/l), Mg2+ (72.667mg/l)) and weak concentrations of ((NO3-) (35 mg /l) and (K+(1 mg/l)). The second class contains 24 samples and represents 24.74% of the total database. It is characterized essentially by concentrations of high chemical elements (HCO3-) (215.692 mg/l), Cl- (255.266 mg/l), and Ca2+ (166.577 mg /l)), NO3-(69.84mg/l),SO42-(344.04mg/l),Mg2+(57.1524mg/l) and the weak concentrations are (Na+) (131.753 Mg /l) and (K+ (2.65 mg /l)). The third class includes the largest number with 61 samples that represent 62.89% of the database. It is characterized by weak concentrations of chemical elements (Cl- (122.283mg /l), Na+ (65.7662mg /l), K+ (1.19589mg / l), SO4-- (167.138mg /l), NO3 (25.9785 mg /l), Mg2+ (39.77)) and relatively high concentrations in relation to standards of potability fixed by OMS (HCO-3 (249.686 mg / l and Ca2+ (105.959 mg /l)). Finally, the fourth class contains only one sample with very high and very alarming concentrations and represents 1.03% of the total database (Tab.2 and Fig. 6, 7).

Table 2: Basic statistical quantities (minimum, average, maximum) for the physico-chemical parameters, respectively for the whole database and for the 3 classes obtained.

	Total data base			Classe I			Classe II			Classe III		
	Min	Average	Max	Min	Average	Max	Min	Average	Max	Min	Average	Max
HCO3--	116	253.65	1061	201	428	1061	116	215.692	329.4	122	249.686	561
$SO4^{-2}$	12	258.577	2660	12	492.83	919	165	344.04	880	30	167.138	460
Cl^-	25	185.207	1100	185	422.5	860	125	255.266	450	25	122.283	275
$NO3^-$	0	38.1299	134	1	35	134	2	69.84	124	0	25.9785	97
Ca^{2+}	52	138.534	720	166	277.67	490	100	166.577	330	52	105.959	196.4
Mg^{2+}	8	50.2769	427	30	72.667	100	23.76	57.1524	98	8	39.77	86
Na^+	14	98.3569	680	88	215.33	288	68	131.753	260	14	65.7662	185
K^+	0.212	2.12298	24	1	8.6667	24	0.898	2.64786	20	0.2	1.19589	3.34

5. Conclusion

Statistical analysis based on the approach of self organizing map (SOM) of Kohonen was applied to a database consisting of 8 physicochemical parameters from 97 groundwater samples from the F'kirina plain For the years 2003, 2008 and 2016. It highlighted the different positive and negative correlations between the different parameters studied. The hierarchical classification of the map SOM (SOM-HC) detected spatial variations from one sample to another by identifying the physicochemical behavior of subterranean waters in the F'kirina Plain. This differentiation would probably be related to the geological nature of the lands traverse [16], the difference of altitude and the domestic discharges of the neighboring townships.

References

[1] Kohonen T., Biol. Cybern., 43 (1982), 59-69.
[2] Khedairia S., Thèse de Doctorat, Faculté des Sciences de l'Ingénieur, Département d'Informatique, Université Badji Mokhtar Annaba, Algérie (2014).
[3] Berrada M., El Hmaidi A., Monyr N., Abrid D., Abdallaoui A., Essahlaoui A., and El Ouali A., Hydrological Sciences journal, 61, 3, (2016) 628-635. DOI:10.1080/02626667.2014.964717.
[4] Younssi H., Thèse de Magister, Faculté des Sciences de la Terre, Université Badji Mokhtar, Annaba, Algérie (2009) 97p.
[5] Kohonen T., Neurocomputing 21, 1-3, (1998)1-6.
[6] Vesanto J., Alhoneimi E., IEEE Trans, neural networks, 11 (3), (2000) 586-600.
[7] Cottrell M., Ibbou S., Letrémy P., Journal de la société française de statistique, 144(2003) 67-106.
[8] Kohonen T. Springer-Verlag, series in information sciences, Berlin, (2001) 501p.
[9] Kiviluoto K., International Conference on Neural Networks (1996) 294–299.
[10] Cabanes G., Bennani Y., In EGC, (2011) 137148.
[11] Uriarte E. A., Martin F.D., International journal of applied mathematics and computer sciences 1, 1(2005)19-22.
[12] Giraudel J.L., Lek S., Ecological Modelling 146, 1(2001) 329-339.
[13] Bamory K., Adama M., Luc S., Amidou D., Hamadoun B. and Droh Lancine G., In. Global Change: Facing Risks and Threats to Water Ressources (proc. of the Sixth World FRIEND Conference, Fez, Morocco), IAHS Pub. 340, (2010), 365-373.
[14] Olkowska E., Kudlak B., Tsakovski S., Ruman M., Simeonov V., Polkowska Z., Science of the Total Environment, 476 (2014) 477-484.
[15] Ward, J. H., Journal of the American Statistical Association 58 (1963) 236-24.
Berkani Ch., Houha B., J. Mater. Environ. Sci., 7, 6, (2016)1856-1865.