



# Statistical Modelling by Topological Maps of Kohonen for Classification of the Physicochemical Quality of Surface Waters of the Inaouen Watershed Under Matlab

R. El chaal\*, M. O. Aboutafail

*Engineering Sciences Laboratory, Data Analysis, Mathematical Modeling, and Optimization Team, Department of Computer Science, Logistics and Mathematics, Ibn Tofail University National School of Applied Sciences ENSA, Kenitra 14 000 Morocco*

## Abstract

Self-organizing maps (SOMs) and other artificial intelligence approaches developed by Kohonen can be used to model and solve environmental challenges. To emphasize the classification of Physico-chemical parameters of the Inaouen watershed, we presented a classification strategy based on a self-organizing map (SOM) artificial neural network in this study. The use of a self-organizing map to classify samples resulted in the following five categories: Low quantities of Sodium Na (mg/l), Potassium k(mg/l), Magnesium Mg(mg/l), Calcium Ca(mg/l), Sulfates SO<sub>4</sub>(mg/l), and Total Dissolved Solids TDS (mg/l) distinguish Classes 2 and 3. Bicarbonate HCO<sub>3</sub> (mg/l), Total Dissolved Solids TDS (mg/l), Total Alkalinity CaCO<sub>3</sub>(mg/l), Mg(mg/l), Calcium Ca (mg/l), and electrical conductivity Cond (s/cm) are slightly greater in Classes 1 and 4. Except for Dissolved Oxygen D.O. (mg/l) and Nitrate NO<sub>3</sub>(mg/l), Class 5 has exceptionally high values for all metrics. The results suggest that Kohonen's self-organizing maps (SOM) classification is an outstanding and fundamental tool for understanding and displaying the spatial distribution of water physicochemical quality.

DOI:10.46481/jnsps.2022.608

**Keywords:** Classification, self-organizing maps, SOM, physical-chemical parameters, cluster

## Article History :

Received: 24 January 2022

Received in revised form: 29 March 2022

Accepted for publication: 30 March 2022

Published: 29 May 2022

©2022 Journal of the Nigerian Society of Physical Sciences. All rights reserved.  
Communicated by: B. J. Falaye

## 1. Introduction

Self-organizing maps (SOM) are artificial neural network techniques based on unsupervised learning algorithms [1]. Because of their classification capabilities and visualization performance, they have been successfully used in environmental fields (soil, water, air, etc.). Shen *et al.* [2] studied the groundwater chemistry and quality of a multilayered groundwater system in the Northwest China Coal an approach based

on the SOM technique combined with multivariate statistical tools. Bigdeli *et al.* [3] applied self-organizing map (SOM) technique and K-means clustering algorithms to describe geochemical anomalies in Moalleman district, northeast Iran. Santos *et al.* [4] investigated the regional hydrogeochemical spatialization and controls of the Serra Geral aquifer system in the southern region of Brazil, using Kohonen's self-organizing maps (SOM) in combination with k-means clustering. Amiri *et al.* [5] conducted a Spatio-temporal assessment of groundwater quality in a coastal aquifer, based on Kohonen's linear discriminant analysis (LDA) and self-organizing maps (SOM).

\*Corresponding author tel. no: +212710115110

Email address: rachid.elchaal@uit.ac.ma (R. El chaal)

The goal of this study is to use a statistical technique (SOM) to display and analyze the spatial distribution of water samples and their Physico-chemical properties at the level of the Inaouen watershed. To emphasize the distinct classes and detect the spatial fluctuations of the Physico-chemical characteristics of the examined watershed, SOM's hierarchical classification (SOM-CHA) is utilized.

## 2. Materials and Methods

### 2.1. Sample Location and Description

The Oued Inaouene watershed is located in northeastern Morocco, between the Middle Atlas and the Pre-Rif, and covers approximately 5109 km<sup>2</sup> (Figure 1). The watershed has a Mediterranean climate with oceanic influences, with great seasonal variations and very evident abnormalities in rainfall due to its geographic location [6]. With an average annual rainfall of roughly 600 mm/year, this region is known for its very varied rainfall from month to month. The rainfall distribution reveals two bands: the first below 500 meters with an annual rainfall of roughly 800 mm, and the second between 500 and 1000 meters with an annual rainfall of 800 to 1500 mm. The rainy season in the study region lasts from November to April, with December and January being the wettest months. July and August, on the other hand, are the driest months of the year. The Oued Inaouène watershed is distinguished by a lithological contrast between the two banks. The Pre-Rif, whose outcrops are marly and support a clear sown flora, is on the right bank. The left bank corresponds to the Middle Atlas' northern boundary, with outcrops ranging from Tazzeka's Paleozoic formations to the triasi formations [6].

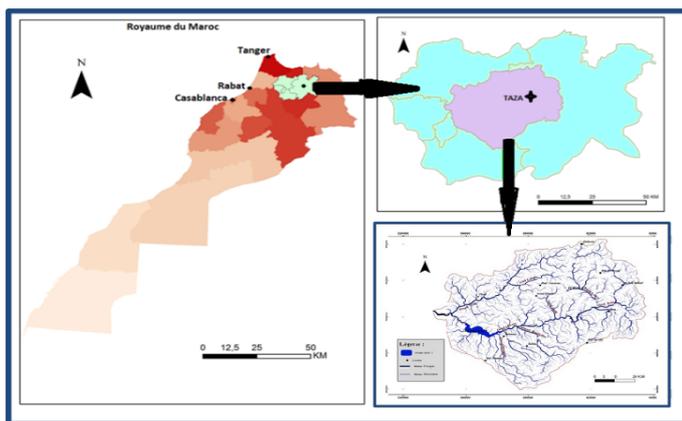


Figure 1: Location and DTM of the study area.

### 2.2. Data Source

The dataset for this study consists of 16 Physico-chemical characteristics measured on 100 surface water samples taken in the Inaouen watershed between 2014 and 2015. The processes for collecting, transporting, and stocking water samples is done by our team according to the protocol of the National Office of

### Drinking Water.

A portion of the analyses was completed on-site, while the remainder was completed at the CURI laboratory in Fez by our team [7].

The parameters used are pH, Bicarbonate (HCO<sub>3</sub>), Dissolved Oxygen (Oxy. Diss), Conductivity (Cond), Temperature (T°C), Total Dissolved Solids (TDS), Total Alkalinity (CaCO<sub>3</sub>), Potassium (K), Magnesium (Mg), Sodium (Na), Chlorides (Cl), Calcium (Ca), Sulfates (SO<sub>4</sub>), Nitrate (NO<sub>3</sub>), Phosphorus (P) and Ammoniac (NH<sub>3</sub>).

### 2.3. Self-Organizing Maps (SOM)

Kohonen, who was looking for a technique to represent huge multidimensional data, was the first to introduce topological maps. To do this, Kohonen uses machine learning [8] to divide data into "similar" groups whose neighborhood structure can be realized and represented using a discrete low-dimensional (1, 2, or 3D) space called a "topological map" [9]. Topological map methods allow data to be projected onto a low-dimensional environment while revealing the data's inherent structures. As a result, the SOM technique preserves both the topology of the data and the distance relationship between them (Figure 2). Unsupervised learning artificial neural network methods [10,11], often known as SOM maps, are a type of neural network [12]. The samples (input vectors) are supplied to a grid of  $d$  neurons (or nodes, or units) in these networks [13]. The choice of the parameter  $d$  (the map's dimension) is chosen ahead of time. The input vectors are connected to each grid neuron via  $d$  synapses and  $d$  weight vectors  $w$ . The map neuron in data space is represented by the vector  $w$ , which is also known as the prototype or referent vector [14]. The BMU is the closest data referent (the Best Matching Unit). The quantization and topological preservation capabilities of a self-organizing map are assessed. Topological error (Te) and quantization error (Qe) are commonly used to validate the SOM categorization [15,16].

**Quantization Error  $Q_e$ :** (or resolution measure) The average quantization error, which is defined by the average distance of the data from their referents (BMU), is used to determine the degree of deployment of the map on the data or the degree of quantization [1,14]. The better the quality of the SOM algorithm, the lower the value of  $Q_e$ . It is expressed as follows:

$$Q_e = \frac{1}{N} \sum_{k=1}^N \left\| x^{(k)} - w(x^{(k)}) \right\|^2 \quad (1)$$

where  $N$  is the number of data,  $x^{(k)}$  is the  $k$ -th individual, and  $w(x^{(k)})$  is the BMU of the individual  $x^{(k)}$

**Topographic error Te** [16]: (or a measure of topology preservation) This criterion assesses how well the SOM preserves the data set's topology [17]. Which is the percentage of data for which the two nearest referents do not correspond to adjacent

map units [18]. In contrast to the Quantification mistake, Te considers the SOM card's structure [19].

The criterion used to measure the number of observations where the first winning neuron ( $c_i$ ) and the second winning neuron ( $s_i$ ) are not neighbours on the map. The second winning neuron of observation has its closest referent vector to this observation after the first winning neuron [20]. The topographic error is a metric that is determined as follows:

$$Te = \frac{\sum_{i=1}^n E}{n} \quad (2)$$

$$E = \begin{cases} 1 & \text{si } r_{c_i} - r_{s_i}^2 \neq 1 \\ 0 & \text{si } r_{c_i} - r_{s_i}^2 = 1 \end{cases}$$

where  $r_c$  and  $r_s$  are respectively the locations of neuron  $c$  and neuron  $s$  on the map.

The topology is perfectly preserved when this criterion is 0.

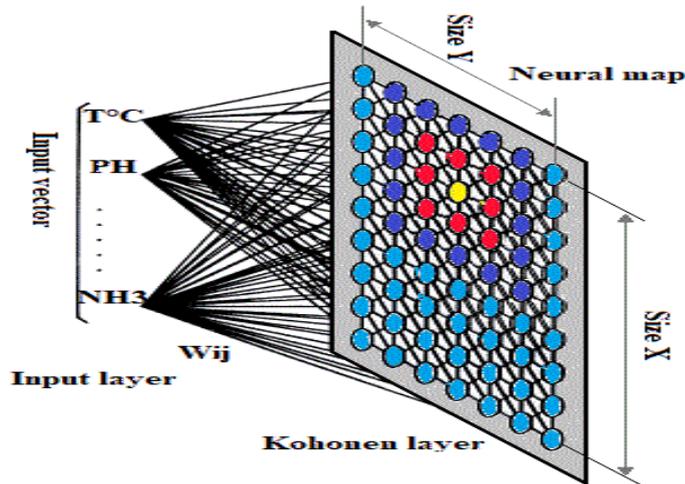


Figure 2: Structure of a topological map

The topological map has several advantages over the linear and classification methods [21] usually used to extract groups of collected samples, such as Principal Component Analysis (PCA) [22], Correspondence Analysis (C.A.) and Hierarchical Clustering (H.C.) [23]. Their limitations are well known. For example, for each of them, a strong distortion is observed when there are non-linear relationships between the variables [24].

#### 2.4. Hierarchical Clustering by SOM (SOM-CHA)

Like other data analysis methods, of which it is a part, the SOM-CHA classification aims to obtain a simple schematic representation [25]. It consists in calculating a matrix expressing the mutual distances between the points to be classified, which are the nodes of the map, and then, based on this matrix, grouping together the closest points. This method allows the construction of a hierarchical tree [26], which reveals several possible partitions, where each point is assigned to one of the groups of a given partition. The choice of the best partition is made once the hierarchical classification is completed [27].

#### 2.5. Algorithm: Kohonen maps (SOM) [16]

Let  $(w_1^t, \dots, w_N^t) \in (\mathbb{R}^n)^N$  be neurons of the vector space  $\mathbb{R}^n$ . We designate by  $V(w_j)$  the set of neighbouring neurons of  $w_j$  for this Kohonen card. By definition, we have  $w_j \in V(w_j)$ . Let  $(X_1, \dots, X_K) \in (\mathbb{R}^n)^K$  a cloud of points. We use a sequence of positive real numbers  $(\alpha_t)$  checking  $\sum_{t \geq 0} \alpha_t^2 < \infty$  and  $\sum_{t \geq 0} \alpha_t < \infty$ .

##### initialization

The neurons  $(w_1^0, \dots, w_N^0)$  are distributed in the space  $\mathbb{R}^n$  in a regular way according to the shape of their neighbourhood.  $t \leftarrow 0$ .

##### closest neuron

We choose a point of the cloud  $X_i$  randomly; then, we define the neuron  $w_{k^*}^t$ , so that:  $w_{k^*}^t - X_i = \min_{1 \leq j \leq N} w_j^t - X_i$ .

##### update

for each  $w_j^t$  in  $V(w_{k^*}^t)$

$$w_j^{t+1} \leftarrow w_j^t + \alpha_t (X_i - w_j^{t+1}) \quad (3)$$

$$t \leftarrow t+1$$

As long as the algorithm has not converged, return to the nearest neuron step.

The update step can be modified to improve the convergence speed [28]:

$$w_j^{t+1} \leftarrow w_j^t + \alpha_t h(w_j^t, w_{k^*}^t) w_k (X_i - w_j^{t+1}) \quad (4)$$

where  $h$  is a function with value in the interval  $[0,1]$  which is 1 when  $w_j^t = w_{k^*}^t$ . And that decreases when the distance between these two neurons increases. A typical function is:

$$h_{ck}(\sigma(t)) = \exp\left(-\frac{d_{ck}^2(r_c, r_k)}{2\sigma^2(t)}\right) = \exp\left(-\frac{\|r_c - r_k\|^2}{2\sigma^2(t)}\right) \quad (5)$$

where  $r_c$  and  $r_k$  are respectively the locations of neuron  $c$  and neuron  $k$  on the map, and  $\sigma(t)$  is the radius of the neighbourhood at iteration  $t$  of the learning process.

Kohonen maps are used in data analysis to project a point cloud into a two-dimensional space in a non-linear manner using a rectangular neighbourhood. They are also used to perform unsupervised classification by clustering neurons where the points are concentrated. The edges connecting the neurons or vertices of the Kohonen map are either narrowed to indicate that two neurons are neighbours or distended to indicate a separation between classes.

### 3. Results and Discussion

#### 3.1. Classification of Surface Water Samples by SOM

The concept of the SOM algorithm is to conduct a non-linear classification of complicated datasets by recognizing similar patterns. In this work, the input layer consists of vectors representing individuals, each of which contains 16 components representing the 16 Physico-chemical parameters of the surface waters studied. The output layer is composed of

100 neurons (10 rows  $\times$  10 columns). This size was chosen for the output map because it minimizes the two error criteria ( $Q_e=0.268$  and  $T_e=0.03$ ).

The SOM component planes of the data set allow distinguishing two types of colors; dark red cells represent high values, while blue cells represent low values (Figure 3) [16]. The similar colors between the variables correspond to a positive correlation; this can be illustrated between the variables Bicarbonates ( $\text{HCO}_3$ ), Chlorides ( $\text{Cl}$ ), Magnesium ( $\text{Mg}$ ), Sodium ( $\text{Na}$ ), Sulfates ( $\text{SO}_4$ ), Conductivity ( $\text{Cond}$ ), Nitrate ( $\text{NO}_3$ ), Ammoniacal Nitrogen ( $\text{NH}_3$ ), Calcium ( $\text{Ca}$ ), Total Alkalinity (in  $\text{CaCO}_3$ ), Potassium ( $\text{K}$ ) and Total Dissolved Solids ( $\text{TDS}$ ). On the other hand, Dissolved Oxygen ( $\text{Oxy. Diss}$ ) and  $\text{P.H.}$ , Total Alkalinity ( $\text{CaCO}_3$ ), and Phosphorus ( $\text{P}$ ) show a negative correlation. The other variables, especially  $\text{T}^\circ\text{C}$ ,  $\text{P}$ , and  $\text{NO}_3$ , show neither positive nor negative correlations and vary autonomously from the others (Figure 3).

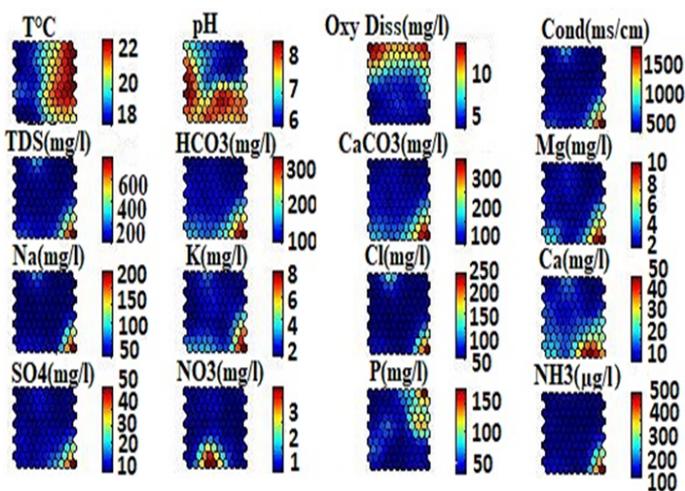


Figure 3: Gradient of values of Physico-chemical parameters on the Kohonen

### 3.2. Principal Component Analysis (PCA)

The PCA Result shows the score road composed of the two components, PC1 and PC2, which are regarded as the most informative ones since they contribute to most of the variance. In our case, PC1 and PC2 respectively contribute 51.1% and 11.2% of the total variance. Therefore, the first two components of the circle of correlations between variables in the subspace PC1 vs PC2 contain 62.3% of the data. Figure 4 illustrates the circle of correlations between the variables in the factor plane ( $\text{PC1} \times \text{PC2}$ ).

The correlation circle between variables in the factorial diagram ( $\text{PC1} \times \text{PC2}$ ) shows that the variables  $\text{Mg}$ ,  $\text{HCO}_3$ ,  $\text{CaCO}_3$ ,  $\text{K}$ ,  $\text{SO}_4$ ,  $\text{Na}$ ,  $\text{Cl}$ ,  $\text{NH}_3$ , and  $\text{TDS}$  are positively correlated with the PC1 axis with coefficients above 0.6. However, the element  $\text{Oxy Diss}$  is negatively correlated with the PC2 axis, with coefficients greater than  $-0.6$ . The elements  $\text{Mg}$ ,  $\text{HCO}_3$ ,  $\text{CaCO}_3$ ,  $\text{K}$ ,  $\text{SO}_4$ ,  $\text{Na}$ ,  $\text{Cl}$ ,  $\text{NH}_3$ ,  $\text{Cond}$  and  $\text{TDS}$  are

therefore positively correlated with each other.

The other variables do not have positive or negative correlations and are poorly represented in the circle, in particular the parameters  $\text{T}^\circ$ ,  $\text{P}$ ,  $\text{NO}_3$  and  $\text{P.H.}$  that vary autonomously.

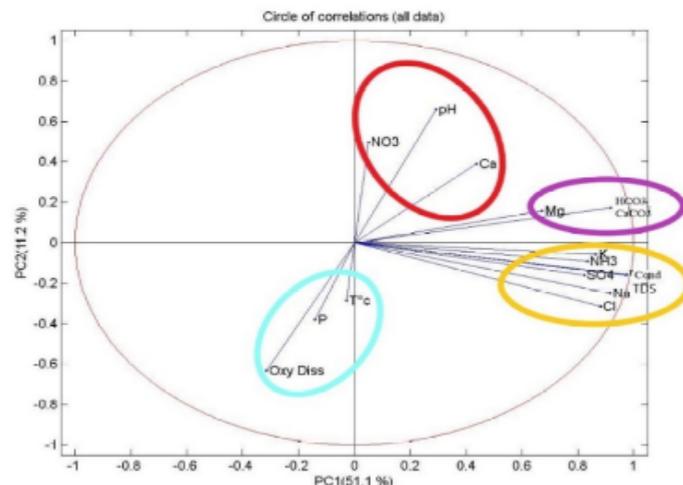


Figure 4: PCA circles of correlation ( $\text{PC1} \times \text{PC2}$ )

### 3.3. SOM-CHA Hierarchical Classification

Once the Kohonen map is obtained, we use a hierarchical classification based on the Ward method [29,30] and Euclidean distance. The hierarchical classification by SOM allows grouping the cells of the SOM map into groups of Physico-chemical parameters of the Inaouen watershed. The dendrogram obtained by SOM-CHA suggests that the 100 neurons should be grouped into five classes (Figure 5).

The first class contains 13 samples and represents 13% of the total data, it includes waters with average chemical element concentration respectively ( $\text{HCO}$  (105.39 mg/l),  $\text{TDS}$  (100.54 mg/l)  $\text{CaCO}_3$  (86.38 mg/l),  $\text{Mg}$  (4.41 mg/l),  $\text{Ca}$  (24.69 mg/l) and electrical conductivity (201.08  $\mu\text{s}/\text{cm}$ )) which are a little high and ( $\text{NH}_3$  (31.23  $\mu\text{g}/\text{l}$ ) and ( $\text{Na}$  (11.60 mg/l) which are low.

The second class includes the largest number of samples (48) and represents 48 % of the total database. It is characterized mainly by low concentrations of chemical elements such as:  $\text{Na}$  (9.00 mg/l),  $\text{K}$  (1.02 mg/l),  $\text{SO}_4$  (4.61 mg/l) and  $\text{TDS}$  (58.33 mg/l) and by high dissolved oxygen concentration (6.24 mg/l). The third class contains 13 samples and represents 13% of the total database. It is characterized mainly by low concentrations of chemical elements ( $\text{SO}_4$  (4.87 mg/l),  $\text{Cl}$  (3.19 mg/l) and  $\text{Na}$  (4.91 mg/l)),  $\text{Mg}$  (1.81 mg/l),  $\text{K}$  (0.84 mg/l),  $\text{Ca}$  (9.95 mg/l),  $\text{NH}_3$  (17.69 mg/l), and  $\text{TDS}$  (43.38 mg/l) and a very high concentration of  $\text{P}$  (225.39 mg/l) and dissolved oxygen (6.30 mg/l). The fourth class contains 23 samples that represent 23 % of the database. It is characterized by medium high concentrations of chemical elements  $\text{HCO}_3$  (117.76 mg/l),  $\text{Cl}$  (47.98 mg/l),  $\text{Mg}$  (4.85 mg/l),  $\text{CaCO}_3$  (96.52 mg/l),  $\text{Ca}$  (29.83 mg/l),  $\text{Na}$  (38.90

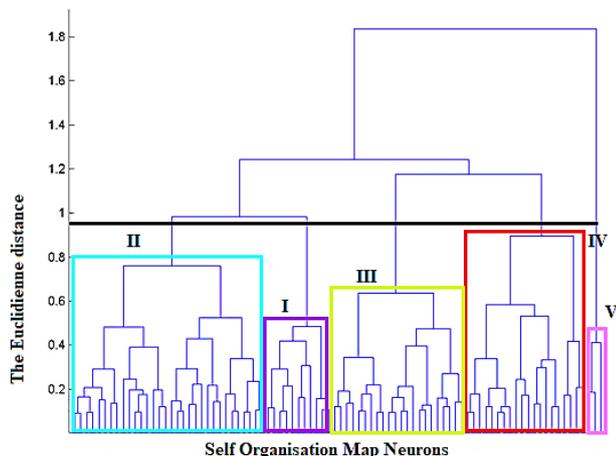


Figure 5: Classification tree (Dendrogram) of the Physico-chemical parameters of the surface waters of Inaouen watershed accessed with the method of topological maps (SOM)

mg/l), NH<sub>3</sub>(74.17 mg/l), electrical conductivity (375.87 μs/cm) and TDS (188.74 mg/l)

Finally, the Fifth class includes the smallest number with 3 samples that represent 3% of the database. It is characterized by very high concentrations of chemical elements NH<sub>3</sub>(770 mg/l), electrical conductivity (1801.33 μs/cm), TDS (901 mg/l), HCO<sub>3</sub> (435.13 mg/l), Na (303.33 mg/l), Cl (47.98 mg/l), Mg(4.85 mg/l), CaCO<sub>3</sub> (356.67 mg/l), Ca (31.40 mg/l), K (9.70 mg/l), SO<sub>4</sub> (59.67 mg/l) (Table 1 and Figures 6-8.

This enrichment would probably have a triple origin, on the one hand, the geological nature of the crossed grounds, the difference of altitude and on the other hand, the domestic discharges of the neighbouring communes.

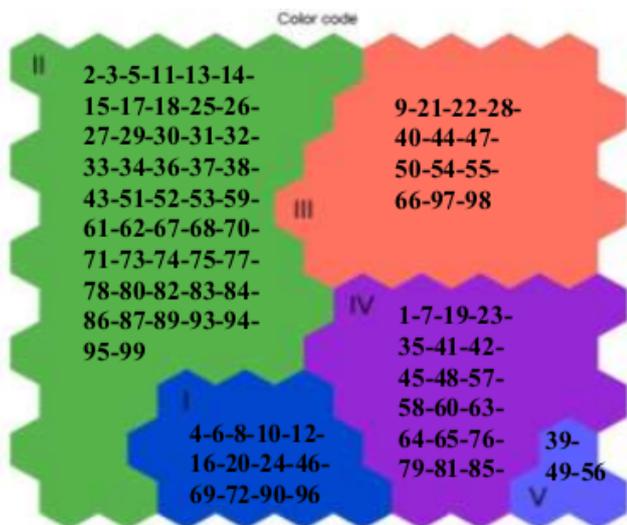


Figure 6: Picture of the five clusters with the samples accessed by the SOM hierarchical clustering.

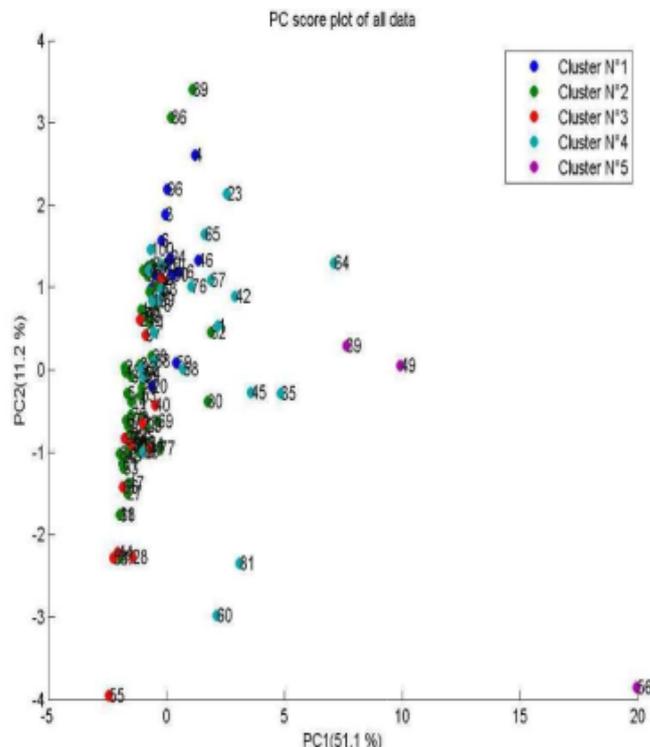


Figure 7: Projection of the samples relative to the five clusters on the plan factorial PC1 × PC2

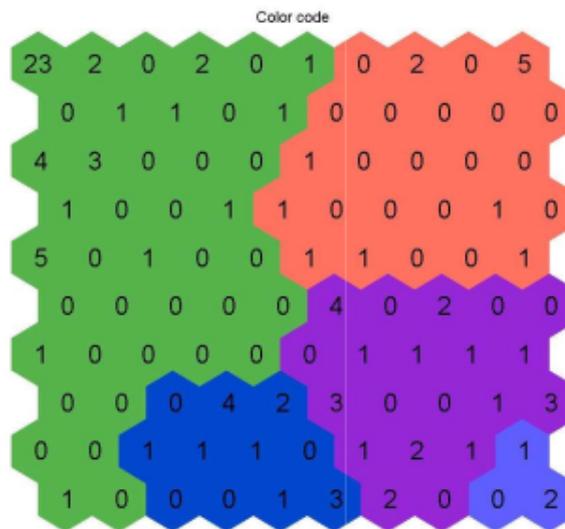


Figure 8: The five clusters generated by the SOM hierarchical clustering are illustrated. The number of patterns given to each neuron is indicated.

Table 1. Statistics on-base quantities (Min, mean, maximum) for Physico-chemical parameters, respectively, for the whole database and for classes 1, 2, 3, 4 and 5.

Table 1: Statistics on-base quantities (Min, mean, maximum) for Physico-chemical parameters, respectively, for the whole database and for classes 1, 2, 3, 4 and 5.

STAT OF ALL DATA				STAT OF DATA FROM CLUSTER 1			STAT OF DATA FROM CLUSTER 2		
Name	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
T °C	16,3	19,9	23	17,1	18,9846	21	16,3	19,7042	23
pH	5,21	7,1923	9,61	6,36	7,56462	9,42	5,21	7,04458	9,61
OxyDiss	0,09	5,0408	14,01	0,08	2,85	7,76	0,08	6,24229	13,26
Cond	20	233,69	2959	134	201077	315	20	116729	629
TDS	45	117,15	1480	67	100538	157	10	58,3333	314
HCO3	3,66	80,6298	634,4	68,32	105389	170,8	6,1	45,1654	170,8
CaCO3	3	66,09	520	56	86,3846	140	5	37,0208	140
Mg	0,2	3,1735	25	1	4,40769	11	0,2	1,82542	6,8
Na	1	24513	540	2,4	11,6	31	1	9,00208	76
K	0,1	1,7211	13	0,34	1,63615	4	0,1	1,01563	6,4
Cl	0,4	25,8394	550	0,6	5,98462	22	0,4	8,62792	99
Ca	1	17794	110	3	24,6923	54	1	11,4333	45
SO4	275	7,83175	150	1,65	7,08462	13	275	4,60781	15
NO3	0,02	0,4805	4,8	0,1	1,15846	4,8	0,02	0,464167	4,5
P	10	58,5	900	20	30,7692	80	10	33,3333	80
NH3	11	62,93	1000	11	31,2308	130	11	34,1875	130

STAT OF DATA FROM CLUSTER 3			STAT OF DATA FROM CLUSTER 4			STAT OF DATA FROM CLUSTER 5			
Name	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
T °C	18	20,8231	23	17,5	20,3174	23	19	19,8	21
pH	5,41	6,70692	8,86	5,63	7,44826	9	6,62	8,08333	9,04
OxyDiss	1,25	6,3	14,01	1,32	3,59913	10,24	0,7	0,906667	2,2
Cond	0.09	84,8462	198	100	375,87	1118	980	1801,33	2959
TDS	23	43,3846	99	50	188739	559	490	901	1480
HCO3	7,32	39,3215	82,96	3,66	117757	329,4	231,8	435133	634,4
CaCO3	6	32,2308	68	3	96,5217	270	190	356667	520
Mg	0,53	1,80923	4,5	0,61	4,84826	23	3,1	12,4667	25
Na	1,5	4,90769	13	1,4	38,8957	180	140	303333	540
K	0,16	0,84	2,4	0,43	2,6987	7,6	6,8	9,7	13
Cl	0,7	3,19231	11	0,4	47,9783	260	77	315667	550
Ca	1,9	9,95385	30	14	29,8261	110	2,4	31,4	86
SO4	1,65	4,86538	18	1,65	9,89783	37	11	59,6667	150
NO3	0,03	0,133077	0,4	0,02	0,364348	2,4	0,1	0,2	0,4
P	20	225385	900	20	34,7826	80	20	40	70
NH3	11	17,6923	50	11	74,1739	250	530	770	1000

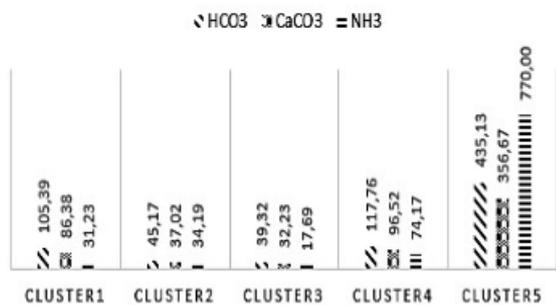


Figure 9: Variation of chemical elements in surface waters of the Inaouen watershed

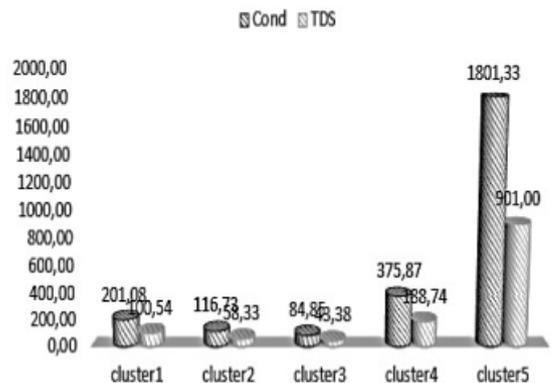


Figure 10: Variation of the total dissolved solids and conductivity in surface waters of Inaouen watershed

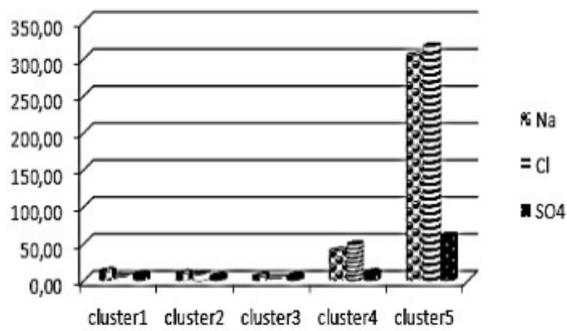


Figure 11: Variation of the Na, Cl and SO4 in surface waters of Inaouen watershed.

### 3.4. The U Matrix Classification:

For the U Matrix classification map, the hexagonal topology was chosen to achieve a better resolution and speedier results. The rectangular topology needed many fewer neurons to get a small quantization and topography error. The result of the U-matrix (Figure 12) from the selected SOM parameters. The precision of the classification via the BMUs in the U-matrix is considered almost exact and produces a high quality and very smooth mapping.

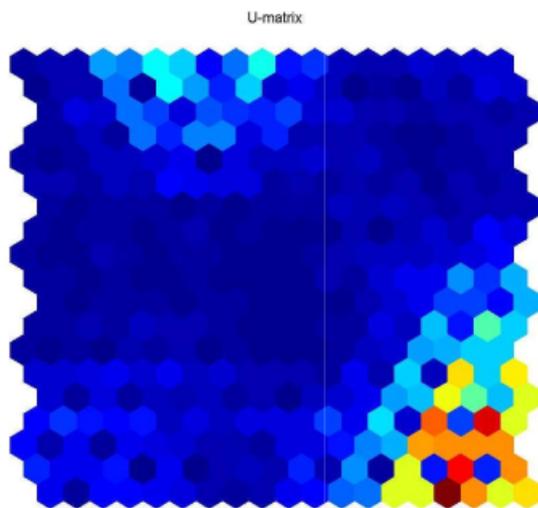


Figure 12: U-matrix representation of SOM. The training samples are the 16 Physico-chemical parameters taken from 100 Inaouen watershed sources.

## 4. Conclusion

Statistical analysis based on Kohonen's self-organizing map (SOM) approach was applied to a database consisting of 16 Physico-chemical parameters carried out on 100 samples of the surface waters of the Inaouen watershed between February 2014 and December 2015. It highlighted the different positive, and negative correlations between the different Physico-chemical parameters studied. The hierarchical classification of the SOM map (SOM-HC) detected spatial variations from one source to another, identifying the physicochemical behaviour of

the waters of the Inaouen watershed. This differentiation would probably be related to the geological nature of the land crossed, the difference in altitude and the domestic discharges of neighbouring municipalities.

## References

- [1] T. Kohonen, *Self-Organizing Maps*, 3rd ed., **30**, Berlin, Heidelberg: Springer Berlin Heidelberg, (2001).
- [2] S. Qu, Z. Shi, X. Liang, G. Wang, and J. Han, "Multiple factors control groundwater chemistry and quality of multi-layer groundwater system in Northwest China coalfield — Using self-organizing maps (SOM)," *J. Geochemical Explor.* **227** (2021) 106795. doi: 10.1016/j.gexplo.2021.106795.
- [3] A. Bigdeli, A. Maghsoudi, and R. Ghezelbash, "Application of self-organizing map (SOM) and K-means clustering algorithms for portraying geochemical anomaly patterns in Moalleman district, NE Iran," *J. Geochemical Explor.* **233** (2022) 106923. doi: 10.1016/j.gexplo.2021.106923.
- [4] M. R. Santos, A. Roisenberg, F. Iwashita, and M. Roisenberg, "Hydrogeochemical spatialization and controls of the Serra Geral Aquifer System in southern Brazil: A regional approach by self-organizing maps and k-means clustering," *J. Hydrol.* **591** (2020) 125602. doi: 10.1016/j.jhydrol.2020.125602.
- [5] V. Amiri and K. Nakagawa, "Using a linear discriminant analysis (LDA)-based nomenclature system and self-organizing maps (SOM) for spatiotemporal assessment of groundwater quality in a coastal aquifer," *J. Hydrol.* **603** (2021) 127082. doi: 10.1016/j.jhydrol.2021.127082.
- [6] B. Benzougagh, A. Dridri, L. Boudad, D. Sdkaoui, and B. Baamar, "Contribution of GIS and remote Sensing for the evaluation of the physical characteristics of Inaouene watershed (northeast Morocco) and their uses in the field of natural hazard management," *Am. J. Innov. Res. Appl. Sci.* (2019) 120 [Online]. Available @ www.american-jiras.com.
- [7] R. El Chaal and M. O. Aboutafail, "Development of Stochastic Mathematical Models for the Prediction of Heavy Metal Content in Surface Waters Using Artificial Neural Network and Multiple Linear Regression," *E3S Web Conf.* **314** (2021) 02001. doi: 10.1051/e3sconf/202131402001.
- [8] O. Olubi, E. Oniya, and T. Owolabi, "Development of Predictive Model for Radon-222 Estimation in the Atmosphere using Stepwise Regression and Grid Search Based-Random Forest Regression," *J. Niger. Soc. Phys. Sci.* **3** (2021) 132. doi: 10.46481/jnsp.2021.177.
- [9] T. Kohonen, "The self-organizing map," *Neurocomputing* **21** (1998) 1. doi: 10.1016/S0925-2312(98)00030-7.
- [10] David Opeoluwa Oyewola, E. G. Dada, J. N. Ndunagu, T. Abubakar Umar, and A. S.A, "COVID-19 Risk Factors, Economic Factors, and Epidemiological Factors nexus on Economic Impact: Machine Learning and Structural Equation Modelling Approaches," *J. Niger. Soc. Phys. Sci.* **3** (2021) 395. doi: 10.46481/jnsp.2021.173.
- [11] V. Umarani, A. Julian, and J. Deepa, "Sentiment Analysis using various Machine Learning and Deep Learning Techniques," *J. Niger. Soc. Phys. Sci.* **3** (2021) 385. doi: 10.46481/jnsp.2021.308.
- [12] N. P. Rougier and G. I. Dtorakis, "Randomized Self-Organizing Map," *Neural Comput.* **33** (2021) 2241. doi: 10.1162/neco\_a.01406.
- [13] D. Olszewski, "A data-scattering-preserving adaptive self-organizing map," *Eng. Appl. Artif. Intell.* **105** (2021) 104420. doi: 10.1016/j.engappai.2021.104420.
- [14] N. Vassilas, "Self-organization of the batch Kohonen network under quantization effects," *Int. J. Comput. Math.* **88** (2011) 3586. doi: 10.1080/00207160.2011.620094.
- [15] J. Oyelade et al., "Data Clustering: Algorithms and Its Applications," in 2019 19th International Conference on Computational Science and Its Applications (ICCSA), Jul. (2019) 71., doi: 10.1109/ICCSA.2019.000-1.
- [16] R. Ponmalai and C. Kamath, *Self-Organizing Maps and Their Applications to Data Analysis*. U.S, 2019.
- [17] G. Cabanes and Y. Bennani, "Learning Topological Constraints in Self-Organizing Map," *NEURAL INFORMATION PROCESSING: MODELS AND APPLICATIONS, PT II* **6444** 17th International Conference on Neural Information Processing (2010) 367.
- [18] H. Ritter, "Self-Organizing Maps on non-euclidean Spaces," in *Kohonen Maps*, Elsevier (1999) 97.

- [19] E. A. Uriarte and F. D. Martin, "Topology Preservation in SOM," PROCEEDINGS OF WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY, VOL 15, 15, no. Conference of the World-Academy-of-Science-Engineering-and-Technology. Univ Deusto, Fac Engn, Bilbao, Spain NR - 20 PU - WORLD ACAD SCI, ENG & TECH-WASET PI - CANAKKALE PA - PO BOX 125, CANAKKALE, 17100, TURKEY, pp. 187-190 WE-Conference Proceedings (2006).
- [20] K. Kiviluoto and IEEE, "Topology preservation in self-organizing maps," ICNN - 1996 IEEE INTERNATIONAL CONFERENCE ON NEURAL NETWORKS, IEEE International Conference on Neural Networks (ICNN 96) **1-4** (1996) 294.
- [21] A. B. Yusuf, R. M. Dima, and S. K. Aina, "Optimized Breast Cancer Classification using Feature Selection and Outliers Detection," J. Niger. Soc. Phys. Sci. **3** (2021) 298. doi: 10.46481/jnsps.2021.331.
- [22] D. Umar, O. V. Omonona, and C. Okogbue, "Groundwater Quality Assessment Using Multivariate Analysis and Water Quality Index in some Saline Fields of Central Nigeria," J. Niger. Soc. Phys. Sci. **3** (2021) 267. doi: 10.46481/jnsps.2021.183.
- [23] F. K. Nakano, R. Cerri, and C. Vens, "Active learning for hierarchical multi-label classification," Data Min. Knowl. Discov. **34** (2020) 1496. doi: 10.1007/s10618-020-00704-w.
- [24] J. L. Giraudel and S. Lek, "A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination," Ecol. Modell. **146** (2001) 329. doi: 10.1016/S0304-3800(01)00324-6.
- [25] J. Serrano-Pérez and L. E. Sucar, "Artificial datasets for hierarchical classification," Expert Syst. Appl. **182** (2021) 115218. doi: 10.1016/j.eswa.2021.115218.
- [26] C. Luo, T. Li, H. Chen, H. Fujita, and Z. Yi, "Incremental rough set approach for hierarchical multicriteria classification," Inf. Sci. (Ny). **429** (2018) 72. doi: 10.1016/j.ins.2017.11.004.
- [27] H. Zhao, S. Guo, and Y. Lin, "Hierarchical classification of data with long-tailed distributions via global and local granulation," Inf. Sci. (Ny). **581** (2021) 536. doi: 10.1016/j.ins.2021.09.059.
- [28] Z.-P. Lo and B. Bavarian, "On the rate of convergence in topology preserving neural networks," Biol. Cybern. **65** (1991) 55. doi: 10.1007/BF00197290.
- [29] M. Roux, "A Comparative Study of Divisive and Agglomerative Hierarchical Clustering Algorithms," J. Classif. **35** (2018) 345. doi: 10.1007/s00357-018-9259-9.
- [30] N. Randriamihamison, N. Vialaneix, and P. Neuvial, "Applicability and Interpretability of Ward's Hierarchical Agglomerative Clustering With or Without Contiguity Constraints," J. Classif. **38** (2021) 363. doi: 10.1007/s00357-020-09377-y.