*Desalination and Water Treatment*
www.deswater.com

doi: 10.5004/dwt.2022.28533

260 (2022) 279–290
June

# Water quality assessment of an artificial small-scale reservoir in the Moroccan Pre-Rif: a case study of Boudaroua Lake using multivariate statistical techniques and self-organizing maps

Hicham En-nkhili[a,*], Mohammed Igouzal[a], Issam Etebaai[b]

[a]*Laboratory of Electronic Systems, Information Processing, Mechanics and Energy, Department of Physics, Faculty of Science, Ibn Tofail University of Kenitra, Morocco, Tel. 0677957294; email: hicham.cpr@gmail.com (H. En-nkhili), Tel. 0673540318; email: igouzal@gmail.com (M. Igouzal)*
[b]*Laboratory of Research and Development in Engineering Sciences, Department of Earth and Environmental Sciences, Faculty of Science and Technique, Abdelmalek Essaadi University (Al Hoceima), Morocco, Tel. 0662600041; email: i.etebaai@uae.ac.ma (I. Etebaai)*

Received 29 December 2021; Accepted 27 April 2022

**ABSTRACT**

The present paper aims to evaluate the water quality of the Moroccan Boudaroua Lake and provide future management decisions. Accordingly, this study describes several multivariate statistical techniques like Pearson's correlation, principal component analysis (PCA), cluster analysis, and self-organizing maps (SOM) generated during June and October 2019, January 2020, and April 2021 through inspecting 15 different water quality surveillance parameters at five various sites. The PCA is imperative to distinguish the factors responsible for altering the surface water quality by reducing them to four principal components, resulting in 91.48% of the water quality's total variance. The obtained varifactors indicate that the parameters in charge of modifying the water quality are related to the wastewater discharge, temperature variation, soluble minerals (natural), and nutrients (agricultural activities). We have selected three different locations for our sampling sites: (Site 1), (Sites 2 and 3) and (Sites 4 and 5) based on their dissimilar water quality via cluster analysis. However, four distinct clusters have been identified using a self-organizing map model based on the temporal relationship between the water quality indicators. The results show that SOM 1 (January 2020) is characterized by lower ion concentrations due to weakly mineralized rain and higher dissolved oxygen (10.5 mg/L). Opposite are noticed in SOM 4 (June 2019). Nevertheless, SOM 2 (October 2019) seems to be highly loaded with nutrients, leading to organic pollution. SOM 3 (April 2021) shows an increase in the calcium ($Ca^{2+}$) and magnesium ($Mg^{2+}$) content which signifies the hardness of the water. The obtained results prove that the multivariate statistical techniques and self-organizing maps are very effective in monitoring surface water quality and, therefore, can be used to detect pollution sources.

*Keywords:* Water quality; Pearson's correlation; Principal component analysis (PCA); Cluster analysis (CA); Self-organizing map (SOM); Boudaroua Lake

* Corresponding author.

## 1. Introduction

Lakes and reservoirs are vital entities for both natural aquatic life and human needs; that's more, they are incredibly vulnerable to any changes in their ecological quality. For this reason, water pollution presents a significant preoccupation in developing countries such as Morocco due to unchecked economic growth and over-exploitation of natural resources. Moreover, climate change (precipitations rate and global overheating due to the increase of greenhouse gaze), urbanization, industrial development, and tourist activities [1] affect the surface water quality and lead to changing its characteristics between seasons [2,3]. On the other hand, the discharge of effluents that contains toxic components and originated from domestic sewage, and agricultural activities across rivers, streams and tributaries induce water quality problems and render aquatic life unsuitable for all living organisms [4]. Boudaroua Lake is one of the most important lakes in Morocco; it is very significant in terms of biodiversity and tourism. Nonetheless, there is a severe lack of knowledge regarding its water quality. Boudaroua Lake was built to increase the population's water supply and irrigation reserve. However, in previous dry seasons, the amount of its water decreased. As a result, the water quality of the reservoir, like other lakes around the world, varied greatly over time and space: salinity [5], acidification [6], temperature [7] and eutrophication level [8] have risen. Therefore, assessing the water quality is critical to truly comprehend the behavior and discrepancy of its components.

Water quality analysis will help restrict lake contamination and provide accurate information to management decision-makers in the aquatic environment. Hence, the data of spatial and temporal samples are more extensive and complex, containing a significant number of physicochemical properties that are typically difficult to evaluate.

Cluster analysis (CA) and principal component analysis (PCA) have been widely used to interpret complex data matrices and better understand water quality. Several studies have applied these techniques to analyze the effect of anthropogenic activities and natural processes on surface water quality. For instance, Barakat et al. [9] used multivariate statistical approaches to examine temporal-spatial fluctuations in the Oum rabia River (Morocco). They discovered that some water quality indices were above Morocco water standards, indicating domestic, industrial and agricultural activities contamination. Ti et al. [10] guesstimated the level of eutrophication in the lower lakes in Australia during drought (April 2008–September 2010) and post-drought (October 2010–October 2013) cycles using a variety of statistical methods (PCA, CA, FA, and discriminant analysis (DA)) and by doing so, they demonstrated that these lower lakes were eutrophic owing to high nutrient and algae concentrations in both drought and post-drought periods. Palma et al. [11] applied multivariate statistical techniques to evaluate spatial-temporal alterations and interpret the water quality dataset obtained at Alqueva Reservoir (south of Portugal). Zhao et al. [12] reviewed the water quality of Baiyangdian Lake (China) using multivariate statistical techniques such as principal component analysis (PCA) and cluster analysis (CA); the registered findings suggested that the main source of water quality degradation in this lake is the industrial, agricultural discharge and the domestic sewage coming from the upstream Fuhe River. Bouzekri et al. [13] quantified the metal pollution of surface water and its relationship to the mining environment (Moulouya River, Morocco) with the aid of the Pearson correlation and principal component analysis (PCA), which indicated the factors and sources of the surface water pollution.

Howbeit, the self-organizing map (SOM) approach is one of the most exalted topics within the neural network field, by virtue of it being a powerful tool capable of grouping similar inputs models from a multidimensional input space and turning them into two-dimensional [14]. It's a pattern analyzing method based on unsupervised learning that has been widely implemented in water quality data analysis. As a case in point, Berrada et al. [15] used a self-organizing map to distinguish seasonal fluctuations in Sidi Chahed dam sediments (Northern Morocco). They found four groups that correspond to water quality metrics. An et al. [16] determined the characteristics of surface water quality in Hong Kong using a self-organizing map. Olawoyin et al. [17] used an artificial neural network (SOM) for the categorization of water, soil and sediment quality in petrochemical regions.

Conversely, Kim et al. [18] characterized water quality and quantity profiles with poor quality data by a machine-learning algorithm. All these studies show that SOM is an effective technique in capturing and analyzing the behavior of the multivariable and nonlinear surface water quality data.

The goal of the current research is to analyze 15 water quality parameters from 5 sampling sites on a seasonal basis (a total of 20 observations) using multivariate statistical techniques and a self-organizing map model (SOM). Moreover, our specific objectives are: (1) to interpret the results obtained from the physico-chemical inquiry of Boudaroua Lake and compare them to the Moroccan Surface Water Guidelines (2002) [19] and the World Health Organization (WHO, 2004) [20], (2) to elaborate the relationship between physical and chemical parameters, (3) to extract the major factors responsible for water quality variability of the lake and finally (4) to determine the spatial and temporal similarities between the sampling stations.

## 2. Material and methods

### *2.1. Study area*

The Boudaroua Lake is located in the north of Morocco between latitudes 34°47′ N and longitudes 05°27′ W. It is about 5 km away from the city of Ouazzane (Fig. 1). The construction of the lake started in 1936, with the intent to provide drinking water to the city until 1970. Nowadays, the lake has become the ideal enjoyment and relaxation destination for most tourists. It is 210 m above sea level and has a surface area of approximately 13 ha. It also has a maximum depth of 8 m with a mean depth of 5 m.

The lake's primary water sources suppliers are direct precipitation and intermittent streams [21,22]. During dry seasons, evaporation causes detrimental water loss, and the downstream dam allows excess water to be evacuated

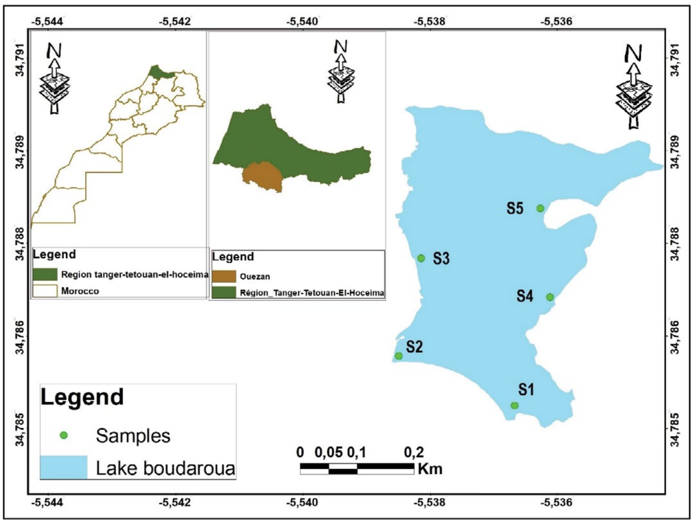


Fig. 1. Map of Boudaroua Lake and water quality stations.

during floods. The majority of the watershed area is covered by forests and supports large numbers of flora such as pine, eucalyptus, carob tree, arbutus, vitex thyme, and cistus; it also contains different fauna like fox, rabbits, hedgehog, hares, in addition to fish such as black bass and mirror carp.

The Ouazzane area is geologically related to the Mediterranean Alpine system; it is situated in the Rif Mountains, whose geological structure is made up of siliceous and limestone layers, resulting in a very rugged mountainous terrain with an altitude of fewer than 350 m, except in Mokrisset (rural commune) that's averaged at 600 m.

This area has a Mediterranean climate, with standard summer temperatures stretching from 20°C to 28°C and typical winter temperatures fluctuating from 10°C to 14°C. The precipitation of this area is characterized by being spread sporadically with an annual average of 800 mm.

The sampling site's placement has been carefully chosen to cover a large surface of the leading site that can accurately reflect the water quality of the lake system, which includes tributary and sewage inputs. Site 1 is a canal that allows the lake water to flow into agricultural lands. Sites 2 and 3 are situated on the downstream side of the lake, where local household waste is deposited and tourists practice their activities. Site 4, compared to other sites, has a small depth. The area's inhabitants use this site to irrigate their agricultural fields. Site 5 is upstream of the lake, where mainly freshwater reaches this latter during the rainy season.

### 2.2. Physical and chemical analysis

Water quality parameters such as temperature (T), pH and dissolved oxygen (DO) have been measured on-site with GLX apparatus multi parameters. The chemical properties of the samples have been gauged with the use of five polyethylene bottles. The date and station numbers are written on these sample bottles. On the day prior to usage, all glassware got washed with 1% $H_2SO_4$ solutions and were then rinsed with distilled water and dried. The sampling bottles were used after they were rinsed several times with surface water at the sampling site and immediately transported to the laboratory in coolers containing ice at a temperature of 4°C in order to prevent water quality degradation of the samples before analysis. Ammoniacal nitrogen ($NH_3$–N), ammonium nitrogen ($NH_4^+$–N), biological oxygen demand ($BOD_5$), chemical oxygen demand (COD), cations ($Na^+$, $K^+$, $Ca^{2+}$, $Mg^{2+}$) and anions ($Cl^-$, $SO_4^{2-}$, $NO_3^-$) are analyzed in accordance with standard procedures [23]. The data sets of 20 water quality surveillance stations comprise 15 water quality parameters inspected seasonally throughout four seasons: June 2019, October 2019, January 2020 and April 2021. Table 1 summarizes the water quality parameters, their units, and analysis methodologies.

### 2.3. Multivariate statistical techniques

#### 2.3.1. Principal component analysis

PCA is used to transform the original variables into new, uncorrelated variables (axes) called principal components, which are linear combinations of the original variables [24]. PCA offers information on the most critical parameters that define the entire data set interpretation, data reduction and summary of statistical correlation amid water constituents with the least amount of original data loss.

#### 2.3.2. Cluster analysis

Cluster analysis (CA) is a series of analytic processes to break down extensive multivariate data into smaller groups based on dissimilarity between the objects to be clustered [25]. The resulting object clusters ought to have high internal homogeneity (within clusters) along with high external heterogeneity (between clusters) [26].

Ward's procedure has been applied to appraise the distance between clusters [27], and their similarities are therefore measured via squared Euclidean distances [28]. The most popular method to do so is the hierarchical agglomerative clustering represented by a dendrogram

Table 1
Water quality parameters associated with their abbreviations. Units and analytical methods used

| Variables | Abbreviations | Analytical methods |
|---|---|---|
| Temperature, °C | T | Thermometer |
| pH | pH | pH meter |
| Electrical conductivity, µS/cm | EC | Electrometric |
| Dissolved oxygen, mg/L | DO | Oximeter |
| Biological oxygen demand, mg/L | BOD | DBO system |
| Chemical oxygen demand, mg/L | COD | DCO system |
| Ammoniacal nitrogen, mg/L | $NH_4^+$–N | Spectrophotometric UV-visible |
| Nitrate–nitrogen, mg/L | $NO_3^-$–N | Spectrophotometric UV-visible |
| Chloride, mg/L | $Cl^-$ | Spectrophotometric UV-visible |
| Bicarbonate, mg/L | $HCO_3^-$ | Volumetric dosing |
| Sulfate, mg/L | $SO_4^{2-}$ | ICP-optical emission spectrometry |
| Sodium, mg/L | $Na^+$ | ICP-optical emission spectrometry |
| Calcium, mg/L | $Ca^{2+}$ | ICP-optical emission spectrometry |
| Magnesium, mg/L | $Mg^{2+}$ | ICP-optical emission spectrometry |
| Potassium, mg/L | $K^+$ | ICP-optical emission spectrometry |

(tree diagram). The GraphPad Prism 8 and XLSTAT 2014 have been used to perform all mathematical and statistical calculations in this study.

### *2.4. Self-organizing map*

The SOM model contains two artificial neuron layers linked by computational weights: An input and a competitive layer (output layer). The samples are connected to the input layer, while the output layer is made up of an array of nodes (Fig. 2).

The SOM's procedure arranges neurons on a 2D map, with each neuron having a prototype weight vector linked with the input data variables, based on three processes. The first is the competition process which determines the best matching unit (BMU) as the known winner neuron (minimum distance between the weight and input vectors). The second is the cooperation process that identifies the adjacent neurons. As for the final process, its main feature is updating the weight vectors.

If two neighboring neurons on the output map space have similar weight vectors and carry out similar samples features, they can be grouped. In contrast, examples of neurons with diametrical opposite characteristics are predicted to be separated on the map [15].

Every mathematical and theoretical concern regarding this method has been described by [29–31]. The SOM algorithm can be summarized in the following steps [32,33].

*Step 1*: weights vectors initialized with random values, set learning rate α;
*Step 2*: For each $j$ neuron, compute the Euclidean distance between the input vector $x_i$ and weight vector $w_{ij}$;

$$D(j) = \sqrt{\sum_{i=1}^{j}\left(x_i - w_{ij}\right)^2} \tag{1}$$

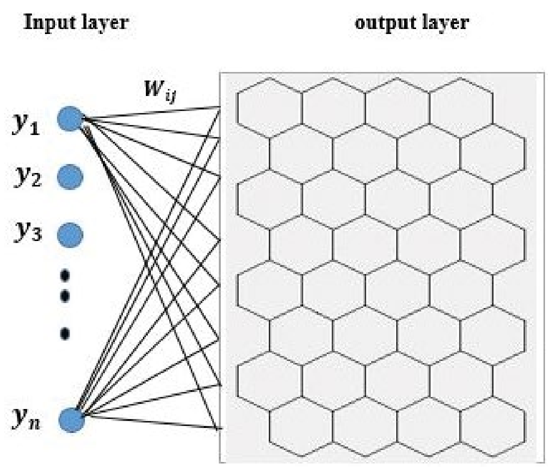

Fig. 2. Topological structure of SOM.

*Step 3*: Finding the index $j$ where $D(j)$ is a minimum;
*Step 4*: The weights vectors are updated for all neurons $j$ within a specified neighborhood of $j$ and for all $i$;

$$w_{ij}(\text{new}) = w_{ij}(\text{old}) + \alpha\left(x_i - w_{ij}(\text{old})\right) \tag{2}$$

*Step 5*: Update learning rate α. It is a decreasing function of the number of epochs;
*Step 6*: Reduce the radius of a topological neighborhood at specified times;
*Step 7*: Test of stopping typical conditions.

#### *2.4.1. Data pre-processing and map size*

The SOM algorithm has been employed to transform the input data using normalization techniques to reduce data redundancy and improve data integrity. Besides, normalization ensures that all data is present and read across all input vectors. Three distinct normalization procedures ('rang', 'Var', and 'log') developed by SOM Toolbox Team Helsinki University of Technology (http://www.cis.hut.fi/projects/somtoolbox/) have been explored in this study [34].

The chosen map was based on two criteria:

*The quantization error (QE)*: assesses the resolution of the Kohonen map by taking the average of the distances between each input vector and their BMU (Best Matching Unit) or winner neuron [29].

*The topographic error (TE)*: represents the percentage of all data vectors, in which the first and second BMU are not adjacent [35].

After multiple trials with different map sizes, the TE and QE (small values) have been used to measure the suitability and fitness of the optimized map size.

#### 2.4.2. Component planes

The component planes describe the values of the unique components of each neuron, in addition to the way each input vector differs in the space of the output map [15]. For each investigated sample, the color gradation on the component planes illustrates their levels (variables) according to the relative value of the respective neuron component [36]. The highest values are displayed in red, while the lowest are divulged in blue. It should be noted that; positively correlated variables have nodes of the same color in the same area of the map, whereas negatively correlated variables have nodes of the same color but with the opposite distribution. Furthermore, the SOM toolbox [34] has been used to implement all simulations in MATLAB R2018b.

#### 2.4.3. K-means clustering

K-means is a non-hierarchical, unsupervised clustering algorithm that enables the data set to be split into K separate clusters. This method minimizes the sum of squared distances between all points and their cluster centers. Accordingly, similar data will be grouped in the same cluster [37].

$$\sum_{j=1}^{c}\sum_{y_i \in Q_j}\left\| y_j - c_j^2 \right\| \tag{3}$$

where the index $j$ is the class label, $C$ is the number of clusters, $y_i$ is a data point, and $c_j$ is the centroid of the data point $j$. The k-means technique separates a data set into several clusters by iterative clustering. To determine the ideal number of clusters $C$ for the dataset, we have used the Davies–Bouldin validity index [38] which is defined by the subsequent equation:

$$\text{DB} = \frac{1}{C}\sum_{j=1}^{C}\text{Max}_{i \neq j}\left(M_{ij}\right) \tag{4}$$

with $M_{ij}$ is the cluster distance ratio between the $i$th and $j$th clusters and is calculated mathematically by the following relation:

$$M_{ij} = \frac{\left\{L_i + L_j\right\}}{L_{ij}} \tag{5}$$

where $L_i$, $L_j$ portrays the average distance between each point in the $i$th cluster and the centroid of the $i$th cluster, the $j$th cluster and centroid of the $j$th cluster respectively. $L_{ij}$ is the Euclidean distance between the centroids of the $i$th and $j$th clusters.

## 3. Results and discussion

### 3.1. Water quality evaluation

Table 2 summarizes the maximum, minimum, mean, and standard deviation of variables measured in the sample stations of the water lake. The results are compared to the values of Morocco's standards for surface water (M.S, 2002) [19] and (WHO, 2004) [20].

The Pearson correlation Index serves as an approach to quantify the association between the water quality parameters, as shown in Table 4. Meanwhile, Table 3 depicts the nature of the connection between them.

The electrical conductivity (EC) measures soluble salt concentration in water (cation and anion). The electrical conductivity measurements of the studied water observed at all stations of the lake range from 1,209.6 to 1,487 µS/cm, exceeding thus the (WHO, 2004) [20] guidelines (Table 2) for drinking water. This surge in the EC level is attributed to a tremendous amount of dissolved salt in the big springs that supply into Boudoroua Lake, although, these values do not exceed the limit value set by the Moroccan quality standards adopted by the Ministry of Energy, Mines, Water and Environment of Morocco (2002) [19] for surface water. Furthermore, the water contains a high electrolyte concentration because of evaporation and high conductivity in the dry season. However, the observed decrease in winter is ascribable to the strong dilution by the weakly mineralized rainfall (Fig. 3a). The compiled results prove the strong positive correlation between Water EC and $Cl^-$ ($r$ = 0.95), in addition to the high positive correlation with T ($r$ = 0.83), $Na^+$ ($r$ = 0.90), $SO_4^{2-}$ ($r$ = 0.84), not to mention the moderate correlation with pH ($r$ = 0.51), $Mg^{2+}$ ($r$ = 0.59) and $NH_4^+$ ($r$ = 0.55).

Water temperature is a key factor in regulating the ecological environment of lakes [39]. For instance: rising the water temperature leads to a decrease in gas solubility and increased mineral solubility. The taste and odor of water are affected by the temperature [40]. The measured temperature is that of the region's atmosphere, which is high in the summer and low in the winter and it changes between 9°C and 25°C (Fig. 3b). Similarly, this parameter demonstrates very high correlation with $Na^+$ ($r$ = 0.91), a high positive correlation with chloride $Cl^-$ ($r$ = 0.73), pH ($r$ = 0.80), $Mg^{2+}$ ($r$ = 0.88), a moderately positive correlation with $SO_4^{2-}$ ($r$ = 0.66), $BOD_5$ ($r$ = 0.57) and COD ($r$ = 0.55).

A pH scale is a unit of measurement that expresses the intensity of an acidic or alkaline state in a solution. In our case, the pH values indicate that the water in the Boudaroua Lake at all stations has an alkaline nature varying between 7.73 and 8.3 (Fig. 3c) but not exceeding the limit

Table 2
Statistical descriptive for the parameters analyzed at all stations

| Parameters | Mean | Minimum | Maximum | Standard deviation | Limit value (MS.2002)/(WHO.2011) |
|---|---|---|---|---|---|
| T | 17.32 | 17.32 | 17.32 | 17.32 | 17.32 |
| CE | 1,339.23 | 1,209.6 | 1,487 | 82.59 | 2,700 |
| pH | 7.97 | 7.73 | 8.3 | 0.15 | 6.5–8.5 |
| $Ca^{2+}$ | 75.15 | 65.02 | 96.82 | 7.3 | 200 |
| $Mg^{2+}$ | 13.50 | 9.7 | 15.94 | 2.18 | 50 |
| $Na^{+}$ | 184.71 | 132.25 | 231.47 | 40.19 | 200 |
| $K^{+}$ | 3.97 | 2.82 | 6.95 | 0.85 | 12 |
| $HCO_3^-$ | 195.71 | 136.74 | 277.5 | 37.68 | 200 |
| $Cl^-$ | 302.35 | 218.4 | 423.86 | 77.17 | 250 |
| $SO_4^{2-}$ | 44.67 | 39.18 | 54.86 | 5.82 | 250 |
| $NH_4^+$–N | 0.93 | 0.32 | 2.14 | 0.53 | 0.5 |
| DO | 7.27 | 5.02 | 10.5 | 1.88 | 7 |
| $NO_3^-$–N | 3.23 | 0.3 | 10.65 | 3.91 | 50 |
| $BOD_5$ | 51.10 | 10 | 96 | 32.09 | 10 |
| COD | 233.90 | 55.6 | 656.06 | 201.22 | 40 |

Table 3
Correlation matrix of water quality parameters (Pearson correlation coefficients ($r$))

| Variables | T | CE | pH | $Ca^{2+}$ | $Mg^{2+}$ | $Na^+$ | $K^+$ | $HCO_3^-$ | $Cl^-$ | $SO_4^{2-}$ | $NH_4^+$ | DO | $NO_3^-$ | $BOD_5$ | COD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T | **1.00** | | | | | | | | | | | | | | |
| CE | **0.83** | **1.00** | | | | | | | | | | | | | |
| pH | **0.80** | **0.51** | **1.00** | | | | | | | | | | | | |
| $Ca^{2+}$ | 0.04 | –0.07 | 0.16 | **1.00** | | | | | | | | | | | |
| $Mg^{2+}$ | **0.88** | **0.59** | **0.79** | –0.13 | **1.00** | | | | | | | | | | |
| $Na^+$ | **0.91** | **0.90** | **0.58** | –0.18 | **0.81** | **1.00** | | | | | | | | | |
| $K^+$ | –0.26 | –0.10 | –0.32 | 0.19 | –0.24 | –0.24 | **1.00** | | | | | | | | |
| $HCO_3^-$ | –0.15 | –0.30 | –0.21 | **–0.54** | 0.23 | 0.05 | –0.13 | **1.00** | | | | | | | |
| $Cl^-$ | **0.93** | **0.95** | **0.67** | 0.10 | **0.69** | **0.89** | –0.15 | –0.39 | **1.00** | | | | | | |
| $SO_4^{2-}$ | **0.66** | **0.84** | 0.43 | 0.37 | 0.30 | **0.63** | –0.04 | **–0.65** | **0.87** | **1.00** | | | | | |
| $NH_4^+$ | **0.46** | **0.55** | 0.08 | **–0.45** | 0.42 | **0.67** | –0.37 | 0.31 | **0.44** | 0.20 | **1.00** | | | | |
| DO | **–0.59** | –0.20 | **–0.81** | –0.19 | **–0.68** | –0.29 | 0.09 | 0.11 | –0.40 | –0.21 | 0.22 | **1.00** | | | |
| $NO_3^-$ | 0.09 | 0.13 | –0.21 | **–0.59** | 0.28 | 0.39 | –0.18 | **0.78** | –0.03 | –0.31 | **0.70** | 0.41 | **1.00** | | |
| $BOD_5$ | **0.57** | 0.34 | **0.77** | 0.38 | **0.45** | 0.24 | 0.02 | **–0.58** | **0.54** | **0.49** | –0.29 | **–0.84** | **–0.70** | **1.00** | |
| COD | **0.55** | **0.48** | **0.65** | 0.34 | 0.41 | 0.38 | 0.23 | **–0.61** | **0.63** | **0.60** | –0.16 | **–0.58** | **–0.50** | **0.82** | **1.00** |

Correlation is significant at $p < 0.05$ level (Number in bold).

value set by (WHO, 2004) [20] and (M.S, 2002) [19]. It exhibits a high positive correlation of $Mg^{2+}$ ($r$ = 0.79), a moderate correlation with $Cl^-$ ($r$ = 0.67) and $Na^+$ ($r$ = 0.58).

DO content is an essential parameter that maintains the equilibrium of aquatic ecosystems [41]; it comes primarily from the atmosphere and the photosynthetic activity of algae and aquatic plants. The water body of Boudaroua Lake is well oxygenated, with a slight increase of dissolved oxygen content in the winter due to the decrease in water temperature on the one hand and the stupendous mixing of the water by the wind on the other hand (Fig. 3d). The samples present values fluctuating from 5.02 to 10.5 mg/L, with an average of 7.27 mg/L, which reflects the excellent water quality of Boudoroua Lake. Concerning the correlation matrix, it has been found that DO reveal an extremely negative correlation to pH ($r$ = –0.81) and a moderately negative correlation to T ($r$ = –0.59) as well as $Mg^{2+}$ ($r$ = –0.68).

Ammonium ion ($NH_4^+$) is an emblematic sign of water contamination brought about by the biodegradation of waste and inputs from domestic, agricultural, and industrial sources. This compound is a water-soluble gas in low concentrations (0.1 mg/L) within natural water. Additionally, $NH_4^+$ is produced by organic nitrogen-containing matter and the gas exchange between water and the atmosphere. The values of $NH_4^+$ in our case are still varied between 0.32 and 2.14 mg/L. Besides, every value registered during the autumn season exceeds the limit value fixed by (WHO, 2004) [20]. This increase in levels is perhaps due to the seepage of the nitrogen matter that is present inside the agricultural soil in the lake during the first precipitation of the autumn rains, reflecting, therefore, agricultural pollution of the lake. Furthermore, $NH_4^+$ shows a significant positive correlation with $NO_3^-$ ($r = 0.70$).

The COD is extensively used in determining the concentration of wastes and is mainly applied to mixtures of harmful substances such as domestic wastewater, agricultural and industrial wastes. In our study, the COD presents high concentrations that varies from 55.6 to 656.06 mg/L, with notable fluctuations between seasons, especially for stations (S1, S2, S3) located downstream of the lake where tourists are known to be practising their activities. Moreover, high concentrations of the COD can be linked to the exudation and transport of domestic sewage, agricultural and industrial pollutants. All stations are moderate to highly polluted, particularly in the summertime when compared to (M.S. 2002) [18]. We have also noticed the presence of fairly positive correlation between COD, T ($r = 0.55$), $Cl^-$ ($r = 0.63$) and pH ($r = 0.65$) alongside a highly positive correlation to $BOD_5$ ($r = 0.82$) although COD is negatively correlated to DO ($r = -0.58$), and $NO_3^-$ ($r = -0.50$).

Table 4
Nature of relationship between water quality parameters

| Pearson correlation | Nature of the relationship |
|---|---|
| 0.9–1 | Very high |
| 0.7–0.89 | High |
| 0.5–0.69 | Moderate |
| 0.26–0.49 | Weak |
| 0–0.25 | Very weak |

$BOD_5$ is a term that refers to the amount of oxygen used in biochemical processes to disintegrate decomposable organic matter. These values range from 10 to 96 mg/L. The highest values are observed in autumn, indicating that the amount of non-degradable organic matter in the sample is high, possibly due to local anthropogenic pollution from occasional fishers and the addition of local household waste. $BOD_5$, displays a moderate correlation with T

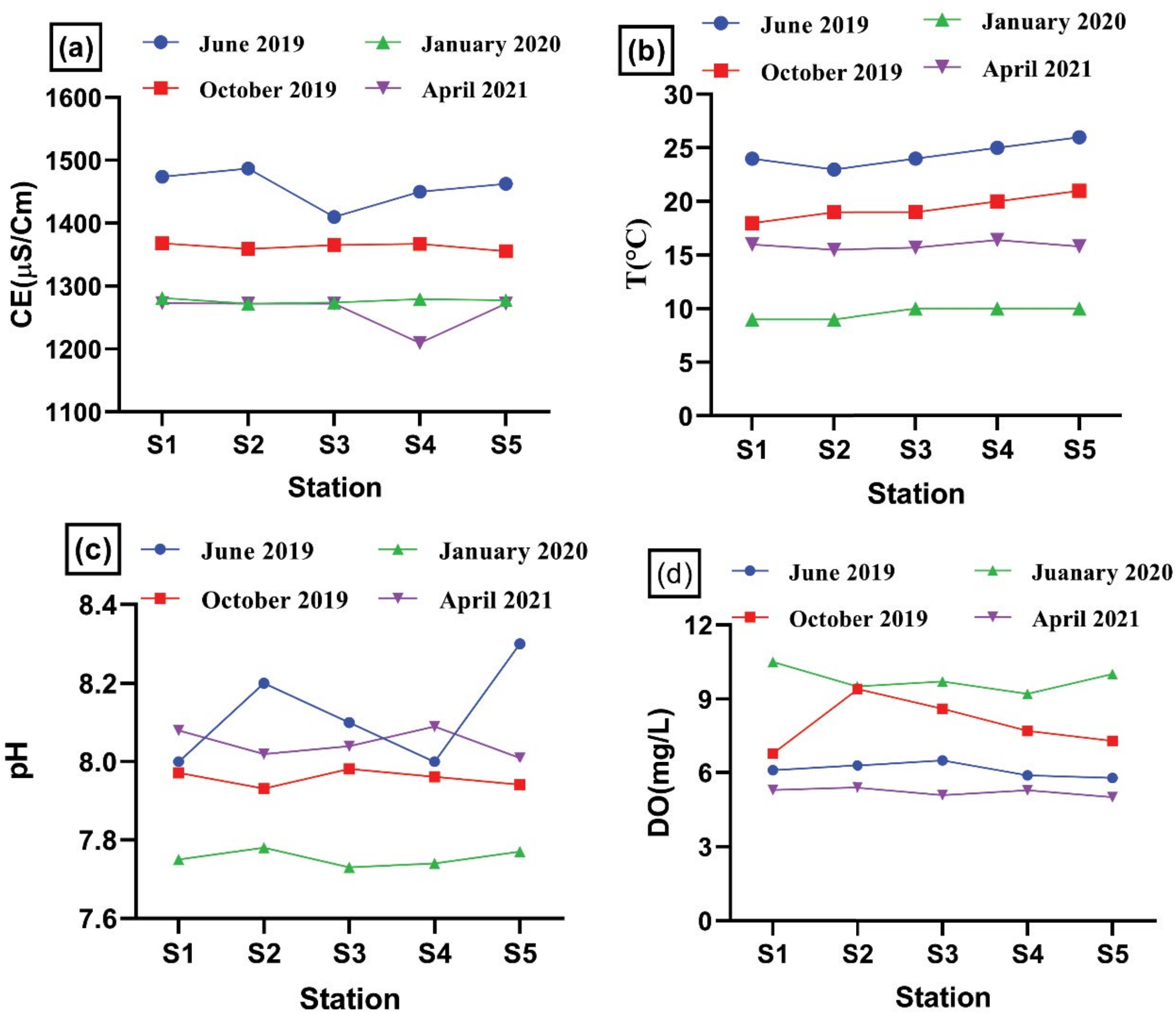


Fig. 3. Water quality parameters, electrical conductivity (a), water temperature (b), pH (c) and dissolved oxygen (d).

($r$ = 0.57), $Mg^{2+}$ ($r$ = 0.45), $Cl^-$ ($r$ = 0.54) and $SO_4^{2-}$ ($r$ = 0.49), on top of a moderate negative correlation with $HCO_3^-$ ($r$ = –0.58). However, it exhibits a solid positive correlation to COD ($r$ = 0.82) and a high negative correlation to DO ($r$ = –0.84).

When talking about the abundance of the cations, the order is the following: $Na^+ > Ca^{2+} > Mg^{2+} > K^+$, where sodium is the predominant mineral of Boudaroua Lake, with concentrations varying from 132.25 to 231.47 mg/L and an average value of 184.71 mg/L. Road salt, water treatment agents, domestic water softeners and waste effluents are all anthropogenic sources of sodium that can add large amounts of sodium to surface water [42]. Meanwhile, potassium indicates very low values that stretch between 2.82 and 6.95 mg/L, with a mean value of 3.97 mg/L. The main source of the latter in freshwater is rocks, and its concentration rises exponentially in polluted water. Potassium is less abundant than sodium because of its higher resistance to weathering and the formation of clay minerals. Sodium ($Na^+$) bears a high positive correlation with T ($r$ = 0.91), CE ($r$ = 0.90), $Mg^{2+}$ ($r$ = 0, 81) and $Cl^-$ ($r$ = 0.89) as well as a moderate correlation with pH (0.58) and $SO_4^{2-}$ ($r$ = 0.63).

The order of profusion of the anions is $Cl^- > HCO_3^- > SO_4^{2-} > NO_3^-$, where chloride is sovereign. The value of fluoride concentration in surface water samples varies between 218.4 and 423.86 mg/L with a mean value of 302.35 mg/L. The chloride ion content of surface water is mainly associated with sodium concentration, and its levels in water samples are higher than the WHO's proposed drinking water safety guidelines (WHO, 2004) [19]. If its concentration exceeds 200 mg/L, the quantity is considered hazardous to human health and can induce a bad taste in the water [43]. Chloride shows a very strong positive correlation with T ($r$ = 0.93), CE ($r$ = 0.95), a high correlation with $Na^+$ ($r$ = 0.89), $SO_4^{2-}$ ($r$ = 0.87), a moderate correlation to pH ($r$ = 0.67), $BOD_5$ ($r$ = 0.54) and COD ($r$ = 0.63).

The scrutinized concentrations of $HCO_3^-$ changes from 136.74 to 277.5 mg/L and the $SO_4^{2-}$ concentrations range from 39.18 to 54.86 mg/L with an average of 198.71 and 44.67 mg/L, respectively.

The $HCO_3^-$ content indicates a highly positive correlation with $NO_3^-$ ($r$ = 0.78) and a moderately negative correlation with $Ca^{2+}$ ($r$ = –0.54), $SO_4^{2-}$ ($r$ = –0.65), $BOD_5$ ($r$ = –0.58) and COD ($r$ = –0.61), whereas $SO_4^{2-}$ shows significant positive correlation with CE ($r$ = 0.84), T ($r$ = 0.66), $Na^+$ ($r$ = 0.63), and COD ($r$ = 0.60).

In the studied stations, $NO_3^-$ exhibits values between 0.3 and 10.65 mg/L with an average of 3.23 mg/L, additionally, there is a significant positive correlation between $NO_3^-$ and $HCO_3^-$ ($r$ = 0.78), $NH_4^+$ ($r$ = 0.70) and moderate negative correlation with $Ca^{2+}$ ($r$ = –0.59), $BOD_5$ ($r$ = –0.70) and COD ($r$ = –0.50).

### *3.2. Source identification of monitored variables*

The principal components (PCs) are usually determined using eigenvalues in the PCA process. The eigenvalue is an indicator of the factor's importance. The number of principal components (PCs) retained for further analysis is usually determined by eigenvalues more significant than one. As a result, the scree plot (Fig. 4) is used to determine the number of PCs to be held. The PC loadings are divided into 'strong', 'moderate', and 'weak', with absolute loading values of 0.75, 0.75–0.50, and 0.50–0.30, respectively [44].

Table 5
Loadings of experimental variables (15) on principal components for the whole datasets

| | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| T | **0.95** | 0.26 | –0.06 | 0.01 |
| CE | **0.83** | 0.31 | 0.41 | 0.07 |
| pH | **0.85** | –0.07 | –0.43 | –0.13 |
| $Ca^{2+}$ | 0.18 | –0.68 | 0.17 | –0.13 |
| $Mg^{2+}$ | **0.78** | 0.39 | –0.43 | 0.15 |
| $Na^+$ | **0.81** | **0.55** | 0.13 | 0.09 |
| $K^+$ | –0.17 | –0.37 | 0.23 | **0.87** |
| $HCO_3^-$ | –0.39 | **0.70** | –0.49 | 0.20 |
| $Cl^-$ | **0.95** | 0.16 | 0.26 | 0.01 |
| $SO_4^{2-}$ | **0.78** | –0.15 | 0.56 | –0.09 |
| $NH_4^+$–N | 0.28 | **0.84** | 0.25 | –0.10 |
| DO | **–0.65** | 0.29 | **0.66** | –0.06 |
| $NO_3^-$–N | –0.18 | **0.94** | –0.01 | 0.17 |
| $BOD_5$ | **0.74** | **–0.57** | –0.28 | –0.01 |
| COD | **0.75** | –0.46 | 0.02 | 0.25 |
| Eigenvalue | 6.94 | 3.97 | 1.83 | 0.99 |
| Variability (%) | 46.28 | 26.44 | 12.19 | 6.57 |
| Cumulative (%) | 46.28 | 72.72 | 84.91 | 91.48 |

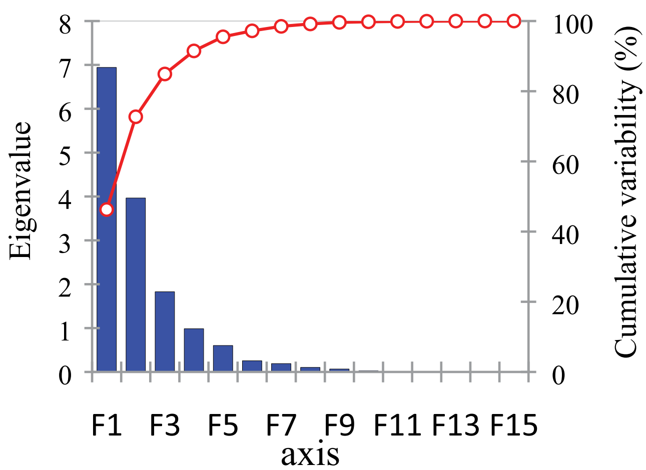


Fig. 4. Scree plot of the eigenvalues.

It has been found that four principal components explain nearly 91.48% of the total variance (Table 5) in the water dataset. Eigenvalues greater than one are derived from the PCs in this analysis. In contrast, variables with eigenvalues lower than one are eliminated due to their low significance.

The first factor PC1 accounts for 46.28% of the total information and shows high positive loading of T, CE, pH, $Mg^{2+}$, $Cl^-$, $SO_4^{2-}$, $BOD_5$ and COD, but a moderately negative loading of DO. This component can be interpreted as an ionic group of salts in the reservoir basin. It is disputed that all major minerals came from one common

Table 6
Performance comparison of three methods of normalization with different card sizes SOM

| Size of the SOM map | Types of normalization | | | | | |
|---|---|---|---|---|---|---|
| | Normalization var | | Normalization log | | Normalization range | |
| | QE | TE | QE | TE | QE | TE |
| 5 × 3 = 15 | 1.712 | 0.000 | 1.662 | 0.000 | 0.504 | 0.000 |
| 6 × 3 = 18 | 1.443 | 0.000 | 1.537 | 0.000 | 0.446 | 0.000 |
| 5 × 4 = 20 | 1.451 | 0.000 | 1.536 | 0.000 | 0.419 | 0.000 |
| 6 × 4 = 24 | 1.249 | 0.000 | 1.388 | 0.000 | 0.359 | 0.000 |
| 5 × 5 = 25 | 1.278 | 0.000 | 1.368 | 0.000 | 0.344 | 0.000 |
| 7 × 4 = 28 | 1.092 | 0.000 | 1.272 | 0.000 | 0.300 | 0.000 |
| 6 × 5 = 30 | 1.333 | 0.000 | 1.233 | 0.000 | 0.291 | 0.000 |
| 8 × 4 = 32 | 1.033 | 0.000 | 1.138 | 0.000 | **0.271** | **0.000** |

source, most likely the dissolution of limestone and gypsum soils. In addition, this particular component reflects the effect of discharge and temperature on the seasonal water quality. The inverse relationship between temperature and dissolved oxygen is natural due to cooler water getting filled with oxygen faster. PC2, which exemplifies 26.44% of the total variance, presents a highly positive correlation with $NH_4^+$, a moderately positive loading of $Na^+$ and $HCO_3^-$, while it exhibits a moderately negative correlation of $Ca^{2+}$ and $BOD_5$. This component represents the lake's level of eutrophication and organic pollution, implying that anthropogenic pollution is primarily caused by a discharge of domestic and agricultural sewage. $NO_3^-$ and $NH_4^+$ may also come from soil erosion during rainfall events, where inorganic nitrogen fertilizers are widely used.

PC3 stands for 12.19% of the total variance and showcases a moderately positive loading on dissolved oxygen and weak positive loadings on electrical conductivity. Yet, the loadings on $Ca^{2+}$ and $Mg^{2+}$ are feebly negative. This factor can be accredited to the cause of variability in physicochemical terms. Finally, PC4 portrays the lowest variance (6.56%) with a strong positive loading on $K^+$, which can be related to the wastewater inputs and nitrogen originating from organic fertilizers.

### *3.3. Spatial and temporal variation in reservoir water quality*

The CA technique has proven to be extremely beneficial in granting an accurate classification of surface waters throughout the area and providing optimal spatial assessment [45]. This technique is applied to lake water quality data to detect spatial dissimilarity and sort the sampling sites (spatial variability) spread over the lake. The resulted dendrogram (Fig. 5) groups all the five sampling sites into three statistically significant clusters.

Cluster 1 encloses one sampling site (Site 1) located downstream of the lake, which receives polluted effluents from non-point sources (NPS) originating from agricultural, industrial, and domestic activities, in addition to the pollution resulting from the remnants of tourism activities, especially in the summer and autumn seasons. Along with Cluster 1, we have category 2, which includes stations 2 and 3 situated downstream of the lake. The only probable cause of this pollution is the leaching of agricultural soils loaded with organic nutrients and fishing activities; in particular, these stations are renowned for having substantial fish. Cluster 3, which includes stations 4 and 5, is positioned upstream of the Boudaroua Lake, has relatively low levels of surface water pollution and corresponds to the middle water depth. Domestic wastewater, animal waste and manure pollute this cluster through soil lixiviation and overflow.

### *3.4. SOM results*

The SOM approach is used to group data on a map so that its size is more meaningful because if it is too small, it may not be able to illustrate some fundamental changes that should be noticed. However, the differences could become insignificant if the map size is too large [46]. The optimal number of map units was is determined by calculating the QEs and TEs of large and small map sizes. Table 6 summarizes the various map sizes that are targeted. It can be seen that the normalization range gives the best map quality with a 32-uint map size (8 × 4) with TE = 0.00 and QE = 0.275. The number of neurons ($n$ = 32) is close to the number of samples ($n$ = 20).

The component planes visualization is a valuable tool when determining the interrelationships between the different water quality measures. When comparing the component planes in Fig. 6, we see that some parameters show positive patterns. The red color denotes a high value of variables in their respective scale bar, while the blue indicates a low value. Indeed, three well-defined groups of correlated parameters are easily distinguished in the grouping of the parameter planes. The parameters of the first group T, CE, pH, $Mg^{2+}$, $Na^+$, $Cl^-$, $SO_4^{2-}$, $BOD_5$ and COD showcase a strong correlation between them, except for COD, which has nodes in the lower left. Additionally, this group's water quality parameters have a high value (red color) in the bottom portions. This outcome is consistent with the Pearson correlation, indicating that these characteristics forcefully impact the PC1 axe. The organic parameters indicate pollution caused by anthropogenic and agricultural activities, representing the second group (PC2 in the above section). Finally, the third group contains only potassium ions, corresponding to the PC4.

The k-means clustering algorithm is used to build a set of clusters from a dataset of only four clusters based on Davies–Bouldin clustering index (Fig. 7). For future analysis, the four-cluster structure of the map is illustrated in Fig. 8.

The following information on water quality parameters of clusters can be derived from Figs. 6 and 8:

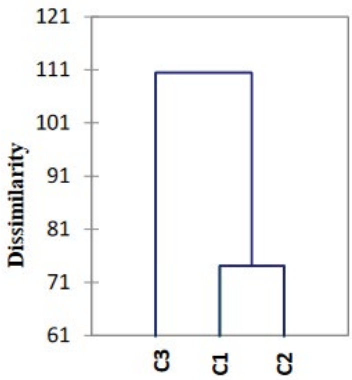


Fig. 5. Dendrogram for cluster analysis based on Ward's method.

- SOM 1: High DO, low T, low CE, low pH, low $Cl^-$, low $Na^+$, low $Mg^{2+}$, low $BOD_5$ and low COD.
- SOM 2: High $NH_4^+$, high $NO_3^-$, high $HCO_3^-$, high $Mg^{2+}$, low COD and low $BOD_5$.
- SOM 3: Low $NH_4^+$, low $NO_3^-$, low CE, low $SO_4^{2-}$ and low DO.
- SOM 4: High T, high CE, high pH, high $Ca^{2+}$, high $Mg^{2+}$, high $Na^+$, high $Cl^-$, high $SO_4^{2-}$, high $BOD_5$ and low DO.

The samples of January 2020 are featured in SOM 1. Seasonal fluctuation of the physical and chemical characteristics could be attributed to climatic change in extreme water dilution by weakly mineralized precipitation during the previous season. SOM 2 encompasses the samples of October 2019 that demonstrate organic contamination caused by the leaching of nitrogen-rich agricultural soil into the lake. SOM 3, on the other hand, consists of the samples of April 2021 that display a surge in calcium and magnesium because of rainwater seeping through the watershed soil. Nevertheless, the impact of evaporation on the proliferation of cations, anions and electrical conductivity is reflected in SOM 4, which embodies the samples of June 2019. The SOM results highlight the dependencies between the different parameters and their spatial-temporal distribution possible.

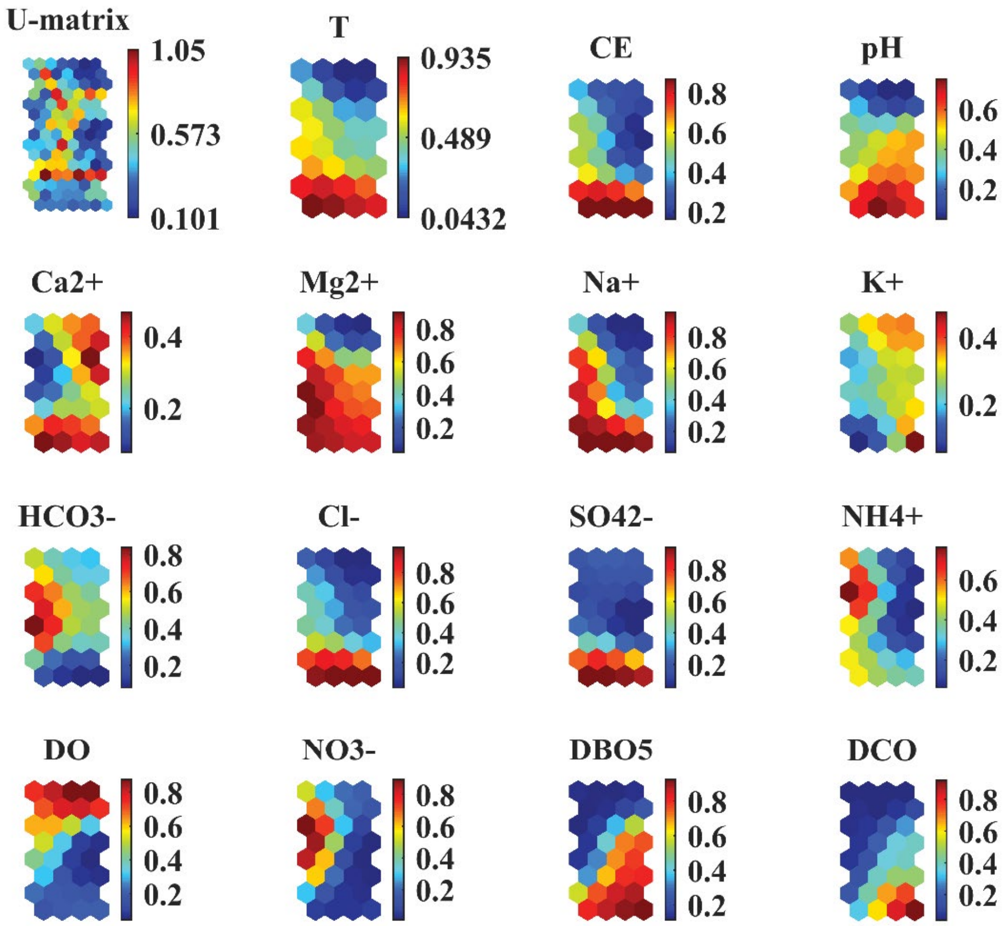


Fig. 6. Patterning analysis for the water quality parameters on the SOM plane.

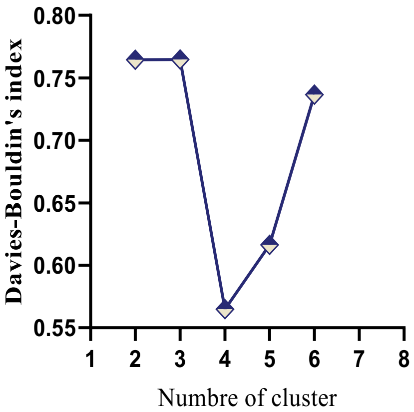


Fig. 7. Davies–Bouldin clustering index of the K-means clustering algorithm.

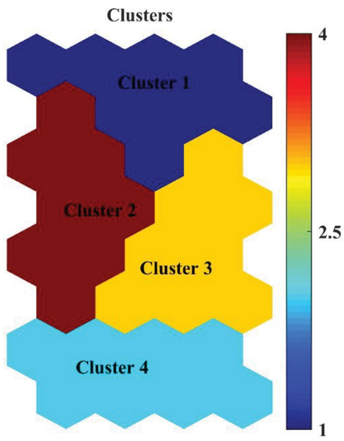


Fig. 8. Clusters of the SOM for the water quality dataset.

## 4. Conclusion

Boudaroua Lake is a vital part of the ecosystem that provides shelter for a diverse range of fauna and flora. In fact, according to the physicochemical analysis, this lake's water is subjected to robust climatic changes. The results show that the water quality is superlative upstream of the Boudaroua Lake, especially in the spring and summer. However, it is subpar, to say the least, in the autumn, when the ammonium nitrogen level reaches its maximum value (2.14 mg/L). On that account, serious measures must be taken to cease future contaminations.

The different multivariate statistical techniques are used to test the spatial and temporal variations in the water quality of Boudaroua Lake. The varifactors (VFs) reveal that the parameters responsible for altering the water quality are related to the mineral parameters and organic pollutants ($NH_4^+$ and $NO_3^-$), which is indicated in the Pearson correlation matrix and component planes in the SOM. Additionally, these last two techniques exhibit sturdy correlations between the temperature and the other physicochemical parameters reflecting the impact of climate change on the lake. The results point out that five sampling sites were grouped into three clusters based on water quality characteristics using cluster analysis. We obtained four clusters by self-organizing map model (SOM 1, January 2020; SOM 2, October 2019, SOM 3, April 2021; and SOM 4, June 2019) with the use of temporal variations of parameters. This research demonstrates that the multivariate statistical techniques and self-organizing map neural network are excellent tools for (1) analysis, interpretation, and classification of data sets, (2) identification of pollutant sources, (3) understanding water quality variability for successful lake water management and finally, it coerces us to recognize that more efforts should be made to protect the lake from the contamination by additional control of tourism and agriculture activities in order to avoid pollution problems in the lake. The results of this study can help also decision-makers to determine long-term management methods for the reservoir's water quality.

## Acknowledgments

The authors express their gratitude to all members of the food and environment laboratory (AGILABS)-Larache for their support with the physicochemical analyses and all and any others who contributed to the success of this work.

## References

[1] A.K. Singh, A. Mittal, S. Kumar, P. Raj, Water quality assessment of Nainital Lake by numerical indices, Pollut. Res., 40 (2020) 893–897.
[2] I. Etebaai, B. Damnati, H. Raddad, H. Benhardouz, O. Benhardouz, H. Miche, M. Taieb, Impacts climatiques et anthropiques sur le fonctionnement hydrogéochimique du Lac Ifrah (Moyen Atlas marocain), Hydrol. Sci. J., 57 (2012) 547–561.
[3] B. Damnati, I. Etebaai, H. Benjilani, K. El Khoudri, H. Reddad, M. Taieb, Sedimentology and geochemistry of lacustrine terraces of three Middle Atlas lakes: paleohydrological changes for the last 2300 cal BP in Morocco (western Mediterranean region), Quat. Int., 404 (2016) 163–173.
[4] M. Qadir, B.R. Sharma, A. Bruggeman, R. Choukr-Allah, F. Karajeh, Non-conventional water resources and opportunities for water augmentation to achieve food security in water scarce countries, Agric. Water Manage., 87 (2007) 2–22.
[5] Z. Liu, A.C.G. Henderson, Y. Huang, Alkenone-based reconstruction of late-Holocene surface temperature and salinity changes in Lake Qinghai, China, Geophys. Res. Lett., 33 (2006) L09707.
[6] L.M. Mosley, B. Zammit, A.M. Jolley, L. Barnett, Acidification of lake water due to drought, J. Hydrol., 511 (2014) 484–493.
[7] M.T.H. Van Vliet, J.J.G. Zwolsman, Impact of summer droughts on the water quality of the Meuse River, J. Hydrol., 353 (2008) 1–17.
[8] S. Li, R.T. Bush, R. Mao, L. Xiong, C. Ye, Extreme drought causes distinct water acidification and eutrophication in the Lower Lakes (Lakes Alexandrina and Albert), Australia, J. Hydrol., 544 (2017) 133–146.
[9] A. Barakat, M. El Baghdadi, J. Rais, B. Aghezzaf, M. Slassi, Assessment of spatial and seasonal water quality variation of Oum Er Rbia River (Morocco) using multivariate statistical techniques, Int. Soil Water Conserv. Res., 4 (2016) 284–292.
[10] T. Li, S. Li, C. Liang, R.T. Bush, L. Xiong, Y. Jiang, A comparative assessment of Australia's Lower Lakes water quality under extreme drought and post-drought conditions using multivariate statistical techniques, J. Cleaner Prod., 190 (2018) 1–11.
[11] P. Palma, P. Alvarenga, V.L. Palma, R.M. Fernandes, A.M.V.M. Soares, I.R. Barbosa, Assessment of anthropogenic sources of water pollution using multivariate statistical techniques: a case study of the Alqueva's reservoir, Portugal, Environ. Monit. Assess., 165 (2010) 539–552.

[12] Y. Zhao, X.H. Xia, Z.F. Yang, F. Wang, Assessment of water quality in Baiyangdian Lake using multivariate statistical techniques, Procedia Environ. Sci., 13 (2012) 1213–1226.

[13] S. Bouzekri, M.L. El Hachimi, K. Kara, M. El Mahi, E.M. Lotfi, Metal pollution assessment of surface water from the abandoned Pb mine Zaida, high Moulouya-Morocco, Geosyst. Eng., 23 (2020) 226–233.

[14] B.-H. Lee M. Scholz, A comparative study: prediction of constructed treatment wetland performance with k-nearest neighbors and neural networks, Water Air Soil Pollut., 174 (2006) 279–301.

[15] M. Berrada, A. El Hmaidi, N. Monyr, D. Abrid, A. Abdallaoui, A. Essahlaoui, A. El Ouali, Self-organizing map for the detection of seasonal variations in Sidi Chahed Dam sediments (Northern Morocco), Hydrol. Sci. J., 61 (2016) 628–635.

[16] Y. An, Z. Zou, R. Li, Descriptive characteristics of surface water quality in Hong Kong by a self-organizing map, Int. J. Environ. Res. Public Health., 13 (2016) 115, doi: 10.3390/ijerph13010115.

[17] R. Olawoyin, A. Nieto, R.L. Grayson, F. Hardisty, S. Oyewole, Application of artificial neural network (ANN)–self-organizing map (SOM) for the categorization of water, soil, and sediment quality in petrochemical regions, Expert Syst. Appl., 40 (2013) 3634–3648.

[18] Z. Kim, H. Jeong, S. Shin, J. Jung, J.H. Kim, S.J. Ki, Characterizing water quality and quantity profiles with poor quality data in a machine learning algorithm, Desal. Water Treat., 182 (2020) 127–134.

[19] Moroccan Standards for Surface Water Quality, Official Bulletin No 5062, 2002. Available at: http://www.sgg.gov.ma/BO/bulletin/Fr/2002/BO_5062_fr.pdf (Accessed Oct. 08, 2021).

[20] World Health Organization, Ed., Guidelines for Drinking-Water Quality, 3rd ed., World Health Organization, Geneva, 2004.

[21] H. En-nkhili, I. Etebaai, K. El kharrim, D. Belghyti, Water hydrochemistry of Lake Boudaroua in the Moroccan Pre-Rif west Mediterranean region, E3S Web Conf., 234 (2021) 00078.

[22] H. En-nkhili, M. Najy, I. Etebaai, F.Z. Talbi, K. El Kharrim, D. Belghyti, Application of Water Quality Index for the Assessment of Boudaroua Lake in the Moroccan Pre-Rif, Proceedings of the 4th Edition of International Conference on Geo-IT and Water Resources 2020, ACM, Al-Hoceima Morocco, 2020, pp. 1–5.

[23] J. Rodier, B. Legube, N. Merlet, L'Analyse de l'eau 9e édition, Entièrement mise à jour, Dunod, Paris, 2009.

[24] T.G. Kazi, M.B. Arain, M.K. Jamali, N. Jalbani, H.I. Afridi, R.A. Sarfraz, J.A. Baig, A.Q. Shah, Assessment of water quality of polluted lake using multivariate statistical techniques: a case study, Ecotoxicol. Environ. Saf., 72 (2009) 301–309.

[25] S.T. Leonard, M. Droege, The uses and benefits of cluster analysis in pharmacy research, Res. Social Adm. Pharm., 4 (2008) 1–11.

[26] S. Muangthong, S. Shrestha, Assessment of surface water quality using multivariate statistical techniques: case study of the Nampong River and Songkhram River, Thailand, Environ. Monit. Assess., 187 (2015) 1–12.

[27] S. Shrestha, F. Kazama, Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji River basin, Japan, Environ. Modell. Software, 22 (2007) 464–475.

[28] K.P. Singh, A. Malik, S. Sinha, Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques—a case study, Anal. Chim. Acta, 538 (2005) 355–374.

[29] T. Kohonen, Learning Vector Quantization, In: Self-organizing maps, Springer, 2001.

[30] Y.-S. Park, T.-S. Chon, I.-S. Kwak, S. Lek, Hierarchical community classification and assessment of aquatic ecosystems using artificial neural networks, Sci. Total Environ., 327 (2004) 105–122.

[31] A.M. Kalteh, P. Hjorth, R. Berndtsson, Review of the self-organizing map (SOM) approach in water resources: analysis, modelling and application, Environ. Modell. Software, 23 (2008) 835–845.

[32] B.-Y. Choi, S.-T. Yun, K.-H. Kim, J.-W. Kim, H.M. Kim, Y.-K. Koh, Hydrogeochemical interpretation of South Korean groundwater monitoring data using self-organizing maps, J. Geochem. Explor., 137 (2014) 73–84.

[33] Ö. Çinar, H. Merdun, Application of an unsupervised artificial neural network technique to multivariant surface water quality data, Ecol. Res., 24 (2009) 163–173.

[34] J. Vesanto, J. Himberg, E. Alhoniemi, J. Parhankangas, Self-Organizing Map in Matlab: The SOM Toolbox, Proceedings of the Matlab DSP Conference, Espoo, Finland, 1999, pp. 16–17.

[35] K. Kiviluoto, Topology Preservation in Self-Organizing Maps, Proceedings of International Conference on Neural Networks (ICNN'96), IEEE, Washington, DC, USA, 1996, pp. 294–299.

[36] T. Li, G. Sun, C. Yang, K. Liang, S. Ma, L. Huang, Using self-organizing map for coastal water quality classification: towards a better understanding of patterns and processes, Sci. Total Environ., 628–629 (2018) 1446–1459.

[37] A. Likas, N. Vlassis, J.J. Verbeek, The global k-means clustering algorithm, Pattern Recognit., 36 (2003) 451–461.

[38] D.L. Davies, D.W. Bouldin, A Cluster Separation Measure, IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: PAMI-1, Issue: 2, April 1979), IEEE, 1979, pp. 224–227.

[39] R. Adrian, C.M. O'Reilly, H. Zagarese, S.B. Baines, D.O. Hessen, W. Keller, D.M. Livingstone, R. Sommaruga, D. Straile, E. Van Donk, G.A. Weyhenmeyer, M. Winder, Lakes as sentinels of climate change, Limnol. Oceanogr., 54 (2009) 2283–2297.

[40] R.G. Abd Ellah, Physical properties of inland lakes and their interaction with global warming: a case study of Lake Nasser, Egypt, Egypt. J. Aquat. Res., 46 (2020) 103–115.

[41] B.L. O'Connor M. Hondzo, Dissolved oxygen transfer to sediments by sweep and eject motions in aquatic environments, Limnol. Oceanogr., 53 (2008) 566–578.

[42] C.A. Cooper, P.M. Mayer, B.R. Faulkner, Effects of road salts on groundwater and surface water dynamics of sodium and chloride in an urban restored stream, Biogeochemistry, 121 (2014) 149–166.

[43] A. Versari, G.P. Parpinello, S. Galassi, Chemometric survey of Italian bottled mineral waters by means of their labelled physico-chemical and chemical composition, J. Food Compos. Anal., 15 (2002) 251–264.

[44] C.-W. Liu, K.-H. Lin, Y.-M. Kuo, Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan, Sci. Total Environ., 313 (2003) 77–89.

[45] L.A. Karroum, M. El Baghdadi, A. Barakat, R. Meddah, M. Aadraoui, H. Oumenskou, W. Ennaji, Assessment of surface water quality using multivariate statistical techniques: El Abid River, Middle Atlas, Morocco as a case study, Desal. Water Treat., 143 (2019) 118–125.

[46] M.-L. Wu, Y.-S. Wang, J.-D. Gu, Assessment for water quality by artificial neural network in Daya Bay, South China Sea, Ecotoxicology, 24 (2015) 1632–1642.