Discover Water

Research

Multivariate statistical methods for analysis of physicochemical and microbiological parameters of well water from the village M'Pody

Aubin Yao N'Dri[1] · Stanislas Egomli Assohoun[1] · Cyrille Gueï Okou[1] · Georges Aubin Tchapé Gbagbo[2,3] · Renaud Franck Djedjro Meless[2,3] · Christophe N'Cho Amin[2,3]

Received: 10 July 2023 / Accepted: 24 April 2024
Published online: 07 May 2024

Abstract
The pollution of surface water and groundwater is a real public health problem that is currently receiving particular attention throughout the world. The use of water for food or hygiene purposes requires excellent physicochemical and microbiological quality. Well water is used for many purposes by the inhabitants of M'pody, a village in the Anyama district of Côte d'Ivoire. In this village, an epidemic of diarrhoea was detected in January 2020. This epidemic claimed sixty-nine (69) victims. This study aims to evaluate well water quality controls using the methods of principal component analysis (PCA), correspondence factorial analysis (CFA), analysis of variance (ANOVA) and self-organizing map (SOM) algorithm. The parameters studied were, turbidity (Tur), conductivity (Cond), hydrogen potential (pH), temperature (T), nitrate (NO_{3-}), nitrite (NO_{2-}), ammonium (NH_{4+}), phosphates (PO_4^{3-}), chlorides (Cl^-), total hydrotimetric degree (DHT), sulfates (SO_4^{2-}), bicarbonate (HCO_{3-}), total alkalinity contents (TAC), escherichia coli (E.coli), thermotolerant coliforms (CTH) and enterococcus faecalis (E.faecalis). Data were collected from seventy-two (72) wells in the village during four campaigns in 2020. Physicochemical parameters were determined by electrochemical and spectrophotometric methods. Microbiological analysis was carried out using membrane filtration technique. Descriptive statistics revealed that Tur, Cond, pH and T parameters did not meet world health organization (WHO, 2017) standards. However, the parameters NO_{3-}, NO_{2-}, NH_{4+}, PO_4^{3-}, Cl^-, DHT, TAC, SO_4^{2-} and HCO_{3-} comply with WHO standards. The results of bacteriological analyzes confirm the presence in very significant numbers of indicators of faecal contamination (CTH, E.coli and E.faecalis) in all wells. The logical explanations for faecal pollution would come from the infiltration of septic tanks located near the wells and the run-off of waste water carrying human and animal faecal matter. The diarrhea epidemic would therefore be caused by the consumption of this polluted water. PCA, FCA and hierarchical cluster analysis (HCA) were jointly employed to identify the structure of wells and deduce the principal factors controlling the parameters of these well waters. ANOVA revealed the effect of human-induced activities as the main factors influencing the physicochemical and microbiological parameters of the studied well waters. Further studies should focus on multivariate statistical techniques for effective forecasting and monitoring of emerging pollution for improved water quality.

Keywords Principal component analysis · Factorial correspondence analysis · Analysis of variance · Self-organizing map

Stanislas Egomli Assohoun, Cyrille Gueï Okou, Georges Aubin Tchapé Gbagbo, Renaud Franck Djedjro Meless and Christophe N'Cho Amin have contributed equally to this work.

✉ Aubin Yao N'Dri, aubin_ndri@yahoo.com; Stanislas Egomli Assohoun, stanlasso@gmail.com; Cyrille Gueï Okou, okou.guei.cyrille@gmail.com; Georges Aubin Tchapé Gbagbo, aubintg2007@gmail.com; Renaud Franck Djedjro Meless, melessrenaud@gmail.com; Christophe N'Cho Amin, amin.christophe@ufhb.edu.ci | [1]Environmental Sciences and Technologies Laboratory, University Jean Lorougnon GUEDE, Daloa, Côte d'Ivoire. [2]Laboratory of water and Food Analysis, National Institute of Public Health, Abidjan, Côte d'Ivoire. [3]Department of Pharmaceutical and Biological Sciences, Félix Houphouët-Boigny University, Abidjan, Côte d'Ivoire.

Mathematics Subject Classification 62A09 · 62H25 · 62H30 · 62J10 · 62P12

1 Introduction

Well water is generally used for agricultural activities or as drinking water. Its quality affects human activities and, consequently, the health of the population. Wells are the main source of water for the inhabitants of M'pody, a village in Anyama, a suburb north of Abidjan (Côte d'Ivoire). The water that these populations use for drinking is not always treated. This can lead to illness. This was the case in the village M'pody, where an epidemic of diarrhoea was detected in January 2020. This epidemic affected sixty-nine (69) people, the majority were children aged 0 to 5. It is therefore necessary to quantitatively assess the characteristics of well water in order to find links between the quality of the water in these wells and this epidemic. The traditional method of assessing water quality involves analyzing physicochemical and microbiological parameters and comparing them with existing standards, in order to inform the public about the environmental conditions of these waters and the measures to be taken. With this in mind, Agbasi et al [2] studied the contamination of sachet water using an analysis of physicochemical parameters, heavy metals and microbial loads tested in sachet water in the six geopolitical zones of Nigeria, during the period 2020–2023. The manufacture, delivery, storage and sale of sachet water, as well as poor environmental hygiene, were identified as potential sources of contamination. Abba et al [1] used spatial, chemometric and indexical approaches to assess trace element pollution in the multi-aquifer groundwater system of the Al-Hassa oasis in Saudi Arabia. The average values revealed that chromium and iron concentrations exceeded the recommended limits for drinking water quality. The heavy metal assessment index, the heavy metal pollution index and the modified heavy metal index indicated low levels of groundwater pollution. Chemometric analysis identified human activities and geogenic factors as contributing to groundwater pollution. In a similar vein, Gobinder et al [13], assessed the seasonal suitability of groundwater for irrigation using indexed approaches, statistical calculations, graphical plots and machine learning algorithms. They concluded that seasonal changes in groundwater quality for irrigation are influenced by monsoon dynamics, showing significant changes in cation and anion chemistry. The artificial neural network models were found to have superior predictive capabilities for irrigation suitability.

When a pollution event occurs, the water can be treated and reused for a variety of purposes. However, the specific purpose of the reuse will determine the levels of treatment recommended. This is a difficult task, especially as the number of water points to be treated is large. It is therefore necessary to find techniques for grouping wells that take into account the physicochemical and microbiological characteristics of each group in order to provide the optimum treatment required. To this end, multivariate statistical analysis, such as principal component analysis (PCA), were used to study the interactions between multiple factors ([5, 9, 10, 24, 25, 31, 32]). This method serves as a theoretical basis for other multidimensional statistical methods called factorial, which appear as special cases. The quality of the estimates it produces depends on the choice of the number of principal components used to reconstruct the initial data. When the number of components is greater than two, it is necessary to look at the individuals projected on all the planes for a good interpretation. This becomes tedious. Additionally, PCA is limited to linear correlations. Kernels PCA or hierarchical cluster analysis (HCA) are often used to overcome these problems. The HCA method was the most widely used for studying the physicochemical and microbiological characteristics of water ([22, 23, 36]).

Unlike all these studies, which have combined several multivariate statistical techniques, this study attempts to find links between the quality of well water and an epidemic that has claimed many lives. In other words, this study attempts to determine how poor water quality contributed to an epidemic. The aim of this paper is to perform a detailed and comprehensive study of well water quality using conventional multivariate analysis techniques. The aim is to find relationships and conclusions that can help determine the state of water quality using biological, physical and chemical indicators in order to prevent future epidemics in other regions. Multivariate statistical analysis, including PCA, correspondence factorial analysis (CFA) and self-organizing map (SOM), is applied to a data set comprising three microbiological parameters (*escherichia coli* (E.coli), *enterococcus faecalis* (E.faecalis) and *thermotolerant coliforms* (CTH)) and thirteen physicochemical parameters (*chlorides* (Cl^-), *conductivity* (Cond), *total hydrotimetric degree* (DHT), *bicarbonate* (HCO_{3^-}), *ammonium* (NH_{4^+}), *nitrate* (NO_{3^-}), *nitrite* (NO_{2^-}), *hydrogen potential* (pH), *phosphates* ($PO_4^{3^-}$), *sulphates* ($SO_4^{2^-}$), *temperature* (T), *total alkalinity contents* (TAC) and *turbidity* (Tur)) sampled in seventy-two wells in 2020 over four campaigns (long dry season, long rainy season, short dry season and short rainy season). This paper can be used as a guide for future studies of water quality using multivariate statistics.

2 Materials and methods for classical data analysis

2.1 Materials

2.1.1 Description of the study area

M'Pody is a village in the Anyama commune in the autonomous district of Abidjan in Côte d'Ivoire (Fig. 1). The geographical coordinates are 5°34′29″ North latitude and 4°14′8″ West longitude in DMS (Degrees, Minutes, Seconds) or 5.57472 and −4.23556 in decimal degrees. The universal transverse mercator (UTM) position is UM61 and the Joint Operation Graphics reference is NB30–10. Anyama covers an area of 114 km^2 and its population is estimated at 325,209 inhabitants [29]. Natural vegetation has given way to intense agriculture. The highly developed culture of oil palms and rubber trees leads to maximum degradation of the natural environment [12]. The climate is equatorial, with four seasons in the annual cycle. A long rainy season from April to July followed by a short dry season from August to September; a short rainy season from October to November and a long dry season from December to March. Average annual rainfall varies between 1600 and 2500 *mm*. Humidity is the order of 80 to 90 percent. The study area is located in the onshore sedimentary basin to the north of the lagoon fault. The geological formations in the area are those of the ivorian coastal sedimentary basin (coarse sands, variegated clays, iron-bearing sands and sandstones, etc.) [33]. The hydrography of the area is composed of small rivers, the Niéké and the gbangbo, as well as several small non-permanent streams. The Niékéis a left bank tributary of the Agnéby river, which flows from north-east to south-west. The gbangbo flows in a north–south direction and empties into the Ebrié lagoon. The geological context of the study area makes it possible to define a single hydrogeological unit that contains groundwater: continuous aquifers. These aquifers are characteristic of the sedimentary basin. These are, the Quaternary aquifer, the Mio-Pliocene aquifer (Continental Terminal) and the Upper Cretaceous (Maestrichtian) aquifer ([14, 19]).

2.1.2 Equipment and sampling

The main measuring equipment consists of a Palintest photometer (Great Britain), a pH meter, a conductivity meter and a turbidity meter for physicochemical parameters, and a membrane filtration device for bacteriological parameters. Water sampling was carried out from the seventy-two wells in the village during four campaigns (long dry season, long rainy season, short dry season, short rainy season) of the year 2020. Samples were taken in 1000 ml polyethylene containers

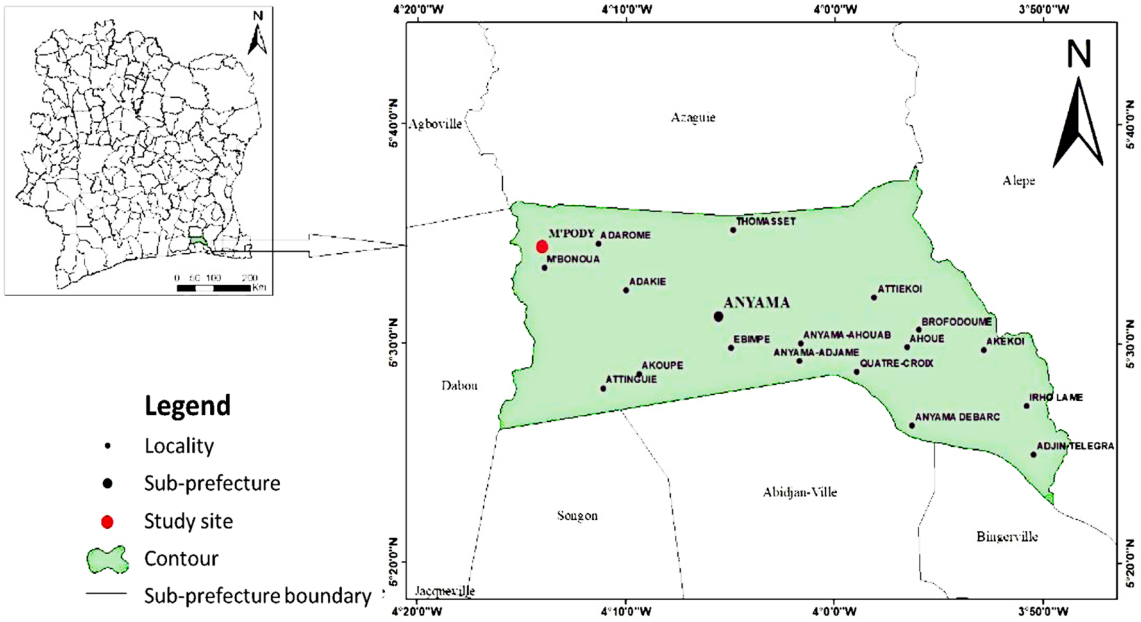

Fig. 1 Location map of the study area

for physicochemical parameters and 500 ml containers for microbiological parameters. The reagents used were of analytical quality. The reagents used to measure chemical parameters were PALINTEST brand (Great Britain). BIORAD Rapid E. coli 2 Agar, Bile Esculin Azide (BEA) agar and Tryptone Sulfite Neomycin (TSN) agar were used to enumerate markers of faecal contamination ([11, 12]).

2.2 Methods for classical data analysis

Samples were taken in strict aseptic techniques to prevent any accidental contamination. Each sample was carried out in sterile flasks according to Jean Rodier's recommendations [30]. Collected samples were stored in a cooler (4 °C) and then transmitted to the laboratory on the same day for analysis. Physicochemical parameters were determined using electrochemical and spectrophotometric methods. Microbiological analysis was carried out using the membrane filtration method (100 *ml* on 0.45 *m* membrane). There are thirteen physicochemical parameters. They are, *Chlorides* (Cl^-), *Conductivity* (Cond), *Total Hydrotimetric Degree* (DHT), *Bicarbonate* (HCO_{3^-}), *Ammonium* (NH_{4^+}), *Nitrate* (NO_{3^-}), *Nitrite* (NO_{2^-}), *Hydrogen potential* (pH), *Phosphates* (PO_4^{3-}), *Sulfates* (SO_4^{2-}), *Temperature* (T), *Total Alkalinity Contents* (TAC), and *Turbidity* (Tur). There are three microbiological parameters. They are *Escherichia coli* (E. coli), *Enterococcus faecalis* (E. faecalis) and *thermotolerant coliforms* (CTH). For more details on the analysis of these parameters, see [12]. Descriptive analysis and multivariate analysis were performed using the two hundred and eighty-eight (288) samples. The analysis of the parameters is carried out on the average of the measurements of the physicochemical and microbiological parameters of the water samples from each well. Means determination was carried out using EXCEL 2010 software. PCA, CFA, analysis of variance (ANOVA), SOM and the location map of the study area were obtained using Python, R, GIMP and ArcGIS software.

2.2.1 Principal component analysis method

The context for PCA involves a data set with observations on p numerical variables, for each of n individuals. These data values define an $n \times p$ data matrix $Y = (Y_j)_{(1\leq j\leq p)}$. The observation of the vector Y_j on individual $1 \leq i \leq n$ is Y_{ij}. In most cases, the variables studied do not have the same unit of measurement. It is common practice to begin by standardizing the variables as in (1)

$$Z_k = (Z_{ik})_{(1\leq i\leq n)} = \left(\frac{Y_{ik} - \overline{Y}_k}{S_k}\right)_{(1\leq i\leq n)} \tag{1}$$

where $\overline{Y}_k$ and S_k are respectively the mean and the standard deviation of the variable Y_k. The principle of PCA is to reduce the dimension of the initial data, by replacing the initial p variables with $(q < p)$ new uncorrelated variables. These new uncorrelated variables are called the principal components of the data set, and denoted $(F_k)_{(1\leq k\leq q)}$ (2)

$$\begin{cases} F_1 &= a_{11}Z_1 + a_{21}Z_2 + \cdots + a_{p1}Z_p \\ &\cdots \\ F_k &= a_{1k}Z_1 + a_{2k}Z_2 + \cdots + a_{pk}Z_p \\ &\cdots \\ F_q &= a_{1q}Z_1 + a_{2q}Z_2 + \cdots + a_{pq}Z_p \end{cases} \tag{2}$$

Principal components are linear combinations of the initial p variables that successively maximize variance. The total variance captured by all the principal components is equal to the total variance in the original data set. The first principal component captures the most variation in the data, but the second principal component captures the maximum variance that is orthogonal to the first principal component, and so on. Before analyzing the results of a PCA, the correlation matrix between the initial variables must be studied. This gives an initial idea of the correlation structure between these variables. This correlation matrix is then used to create a table of the percentages of variance explained corresponding to the different eigenvalues. This table also contains the associated cumulative percentages. It is used to select the q dimensions used to interpret the PCA. This technique is used to calculate the linear correlation coefficients between each initial variable and each selected factor.

Let p_i be the weighting of individual i, c_i^k the coordinate of individual i on the k-th principal component F_k, the correlation of the variable Y_j with respect to F_k is given by formula (3)

$$cor(Y_j, F_k) = \frac{1}{n}\sum_{i=1}^{n} p_i Y_{ij} \frac{c_i^k}{\sigma(F_k)} \tag{3}$$

where, $\sigma^2(F_k) = \sum_{i=1}^{n} p_i (c_i^k)^2 = \lambda_k$. This correlation is used to construct graphs of the variables. The study of these graphics leads to the significance of the principal component. Another tool for interpreting principal components is the notion of contribution defined by formula (4). The contribution of the variable Y_j to the variance of the F_k axis is defined by

$$ctr(Y_j, F_k) = \frac{cor(Y_j, F_k)^2}{\sum_{l=1}^{p} cor(Y_l, F_k)^2}. \tag{4}$$

The contribution is also defined for individual X_i. The contribution of individual X_i to the dispersion of the F_k axis is defined by (5)

$$ctr(X_i, F_k) = \frac{p_i (c_i^k)^2}{\lambda_k}. \tag{5}$$

2.2.2 Correspondence factorial analysis method

Correspondence analysis is a factorial method of multidimensional descriptive statistics. Its aim is to analyze the relationship between two qualitative variables. The graphical results of these two analyzes are then superimposed to produce one or more scatter plots. This graph combines the modalities of the two variables under study. This makes it possible to study the relationship between the two variables. In this system, proximity between observations or between variables is interpreted as strong similarity. Proximity between observations and variables is interpreted as strong relationship. This proximity between two qualitative variables $X = (X_i)_{(i\in I)}$ and $Y = (Y_j)_{(j\in J)}$ is studied on N individuals. The cardinal of I is noted n and that of J is noted p. The number of individuals having the modality i of X and the modality j of Y is noted by x_{ij}. The contingency table is given by the matrix $(x_{ij})_{(1\le i\le n; 1\le j\le p)}$ or $(f_{ij})_{(1\le i\le n; 1\le j\le p)}$ with $f_{ij} = \frac{x_{ij}}{N}$. The column-profiles form a cloud of p points in space $\mathbb{R}^n$ and the array of column-profiles is $\frac{f_{ij}}{f_j} = P(Y = j|X = i)$ where $f_{.j} = \sum_{i=1}^{n} f_{ij} = P(Y = j)$, for $j = 1, \ldots, p$. The associated marginal column profile is $G_C = (f_{1.}, \ldots, f_{n.})$.

The χ^2 distance between two profiles columns j and j' is

$$d^2_{\chi^2}(j, j') = \sum_{i=1}^{n} \frac{1}{f_{i.}} \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{ij'}}{f_{.j'}} \right)^2. \tag{6}$$

The χ^2 distance between the profile column j and its marginal profile G_C is defined as follows

$$d^2_{\chi^2}(j, G_C) = \sum_{i=1}^{n} \frac{1}{f_{i.}} \left(\frac{f_{ij}}{f_{.j}} - f_{i.} \right)^2. \tag{7}$$

The total inertia of the cloud of profiles columns with respect to G_C is

$$\begin{aligned} I_{G_C} &= \sum_{j=1}^{p} f_{.j} d^2_{\chi^2}(j, G_C) = \sum_{i=1}^{n}\sum_{j=1}^{p} \frac{\left(f_{ij} - f_{i.}f_{.j}\right)^2}{f_{i.}f_{.j}} \\ &= \frac{\chi^2}{n} = \phi^2. \end{aligned} \tag{8}$$

This total inertia is decomposed into a sequence of axes of decreasing importance, each representing a synthetic aspect of the relationship between the two variables, and then a representation of the rows and columns is provided in which the position of a point reflects its participation in the independence gap. The definition of the χ^2 distance between

two profiles lines, between the profile line and its marginal profile and the total inertia of the cloud of profiles lines are respectively similar to those defined in (6), (7) and (8).

2.2.3 ANOVA method

One-way analysis of variance is used to study the effect of a qualitative variable X called a factor on a continuous quantitative variable Y. It shows whether the mean of the quantitative variable is the same in the different groups [4]. The different values taken by the factor X are called level (or population). For factor X, it is assumed that there are k levels, k samples of respective sizes $n_1, \ldots, n_k$. The total number of samples is $n = \sum_{i=1}^{k} n_i$. The value of the variable $Y = (Y_{ij})_{1 \le i \le k; 1 \le j \le n_i}$ is measured at each experiment. Then, the analysis of variance model is written as in (9)

$$\begin{cases} Y_{ij} &= m_i + \varepsilon_{ij},\ 1 \le i \le k\ ;\ 1 \le j \le n_i \\ &= \mu + \alpha_i + \varepsilon_{ij} \end{cases} \tag{9}$$

with

- $\varepsilon_{ij} \sim N(0, \sigma^2)$,
- μ average effect,
- α_i effect of level i of factor X,
- Y_{ij} observation of index j of level i of the factor X.

Constraints are, $\sum_{i=1}^{k} n_i \alpha_i = 0,\ \forall (i,j) \neq (k,l),\ \varepsilon_{ij}$ and ε_{kl} are independent. Then, the null and alternative hypotheses of the one-factor ANOVA are given by (10) or (11)

$$\begin{cases} H_0 &: m_1 = m_2 = \ldots = m_k \\ H_1 &: \exists\, i,j \in \{1, \ldots, k\} \text{ such as } m_i \neq m_j \end{cases} \tag{10}$$

or,

$$\begin{cases} H_0 &: \alpha_1 = \alpha_2 = \ldots = \alpha_k = 0 \\ H_1 &: \exists\, i \in \{1, \ldots, k\} \text{ such as } \alpha_i \neq 0. \end{cases} \tag{11}$$

The statistical test defined in (12) is used to determine the significance of the factorial variance in relation to the residual variance. This is the ratio test of these two variances, the formula for which is as follows:

$$F = F_{(k-1,n-k)} = \frac{SCF/(k-1)}{SCR/(n-k)}. \tag{12}$$

The quantities used in this report are defined by: $SCF = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\overline{Y_i} - \overline{Y})^2$, is the dispersion due to the factor and $SCR = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y_i})^2$, the residual dispersion; $\overline{Y} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij}$, is the overall average of the observations and $\overline{Y_i} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$ the mean of level i of factor X. Under the assumptions of normality and homogeneity of the residuals (differences between the observations and the group means), the F statistic follows a Fisher distribution with $k - 1$and $n - k$ degrees of freedom. If the value of F is greater than the theoretical threshold value according to the Fisher distribution, with a given alpha risk (usually 5 percent), then the test is significant. In this case, the factorial variability is significantly higher than the residual variability. We conclude that the means are globally different. If these hypotheses are not verified, it is always possible to apply a transformation at the level of the responses (log for example), or to use a non-parametric ANOVA (Kruskal-Wallis test), or to carry out an ANOVA based on permutation tests.

2.2.4 Self-organizing map method

SOM is a method of classification, representation and analysis of relationships. It was defined by Teuvo Kohonen, in the 80's, from neuromimetic motivations ([17, 18]). In practice, a Kohonen network is made up of N units arranged according to a certain topology. For each unit i in the network, a neighborhood of radius r denoted $V_r(i)$ is defined.

This network is then formed by all the units located at a distance less than or equal to r. Each unit i is represented in $\mathbb{R}^p$ space by a vector C_i called weight vector or code vector. The state of the network at time t is given by $C(t) = (C_1(t), C_2(t), \ldots, C_N(t))$. For a given state C and a given observation x, the winning class $i_0(C, x)$ is the one whose code vector $C_{(i_0(C,x))}$ is closest to the observation x in the sense of a certain distance. The winning class $i_0(C, x)$ is defined in (13)

$$i_0(C, x) = arg \min_i \|x - C_i\|. \tag{13}$$

For a given state C, the network defines an application ψ_C which associates to each observation x the number of its class. After convergence of the Kohonen algorithm, ψ_C respects the topology of the input space, in the sense that neighboring observations in the space $\mathbb{R}^p$ are associated to neighboring units or to the same unit. The code vector construction algorithm is defined in (14) iteratively as follows:

- At time 0, the N code vectors are randomly initialized,
- At time t, the state of the network is $C(t)$ and an observation $x(t+1)$ is presented according to a probability distribution P,

$$\begin{cases} i_0(C(t), x(t+1)) &= arg \min \left\{ \|x(t+1) - C_i(t)\|,\ 1 \le i \le N \right\} \\ C_i(t+1) &= C_i(t) - \varepsilon(t)\big(C_i(t) - x(t+1)\big), \forall i \in V_{r(t)}(i_0) \\ C_i(t+1) &= C_i(t),\ \forall i \text{ not in } V_{r(t)}(i_0) \end{cases} \tag{14}$$

where $0 \le \varepsilon(t) \le 1$is the adaptation parameter and $r(t)$ the radius of the neighborhoods at time t. After convergence of the algorithm, the n observations are classified into K classes according to the nearest neighbor method, relative to the distance chosen in $\mathbb{R}^p$. Graphical representations can then be constructed based on the network topology. For further details, please refer to [6] and [7].

3 Results and discussion

3.1 Descriptive statistics

The mean, maximum (max), minimum (min), median (med) and standard deviation (sd) were used to describe all the data corresponding to the sixteen (16) parameters studied for two hundred and eighty-eight samples (288). Means of

Table 1 Average concentrations of physicochemical and microbiological parameters in well water

Notations	Mean	WHO (2017)	Min	Max	Med	sd
Tur	22.41 NTU	< 5 NTU	2.71	162.43	16.04	23.55
Cond	157.776 $\mu S/cm$	180–1000 $\mu S/cm$	24.73	594	133.29	121.12
pH	5.21	6.5–8.5	4.22	11.51	5.08	0.88
T	27.79°C	≤ 25°C	25.88	29.08	28.03	0.83
NO_{3^-}	14.09 mg/l	≤ 50 mg/l	1.11	44.16	12.38	7.61
NO_{2^-}	0.08 mg/l	≤ 3 mg/l	0.01	0.54	0.04	0.11
NH_{4^+}	0.356 mg/l	≤ 0.5 mg/l	0.01	3.82	0.25	0.48
PO_4^{3-}	0.111 mg/l	0.5 mg/l	0.03	0.75	0.06	0.13
Cl^-	18.17 mg/l	≤ 250 mg/l	1.75	44.7	17.16	9.8
TAC	74.08 mg/l	–	26.25	216.25	65	40.23
DHT	25.78 mg/l	–	5	106.25	19.25	20.80
SO_4^{2-}	10.68 mg/l	250 mg/l	2.25	40.5	7.88	7.77
HCO_{3^-}	90.71 mg/l	–	30	216.25	73.75	42.9
CTH	1219.77 UFC/250 ml	0 UFC/250 ml	4.25	7400	818.88	1269.81
E.coli	947.85 UFC/250 ml	0 UFC/250 ml	2	5650	658.75	1050.09
E.faecalis	936.07 UFC/250 ml	0 UFC/250 ml	3.5	3913	692.50	1043.25

all the parameters were compared to WHO [35] standards in Table 1. Temperature influences the rate of chemical and biological reactions. It affects the level of dissolved oxygen in the water. In the present study, water temperature varied from $25.88 \pm 0.83\ ^{\circ}C$ to $29.08 \pm 0.83\ ^{\circ}C$ with mean $27.79 \pm 0.83\ ^{\circ}C$. The pH is used to measure the acidity or basicity of a solution. It varied between 4.22 ± 0.88 to 11.515 ± 0.88 with mean 5.21 ± 0.88, which means that the water from the wells is acidic. In all this wells, the pH is outside the world health organization (WHO) permitted limit [6.5, 8.5]. The characteristics of the M'pody soil (coarse sands, ferruginous sands and sandstones, etc.) could explain the acidity of these waters. In line with WHO standards, these well waters should not be consumed without being treated. Electrical conductivity is the ability of an aqueous solution to conduct electric current. It determines all the minerals present in a solution. It varied between $24.73 \pm 121.12\ \mu S/cm$ to $594 \pm 121.12\ \mu S/cm$ with mean $157.776 \pm 121.12\ \mu S/cm$. This means that well water is generally poorly mineralized. Turbidity varied from 2.71 ± 23.55 to 162.43 ± 23.55 NTU with a mean 22.41 ± 23.55 NTU. Turbidity levels in well water are on average higher than the WHO standard. In well water, turbidity is caused by small particles in suspension of various natures, such as, clays and silts, microsands, bacteria, organic matter and mineral salts, etc. Most of the time, they are the result of leaching from the surrounding soil and therefore indicate a well that is poorly protected from run-off water. In addition, mean of NO_{3-}, NO_{2-}, NH_{4+}, PO_4^{3-}, Cl^{-}, TAC, DHT, SO_4^{2-} and HCO_{3-} check WHO standards. Microbiological analysis of the well water showed the presence of germs. These microorganisms reached maxima of 7400 CFU/250 ml for thermotolerant coliform, 5650 CFU/250 ml for E. coli and 3913 CFU/250 ml for E. faecalis. The logical explanations for this situation of faecal pollution of the water could come, on the one hand, from the infiltration of septic tanks located near the wells and, on the other hand, from the run-off of waste water carrying human and animal faecal matter. These results are consistent with those of [12] and [15].

3.2 Results of principal component analysis and correspondence factor analysis

PCA is used to extract information from a table of quantitative data of the type individuals×variables to study the proximity between individuals (wells) on the one hand and the links between variables (parameters) on the other. Measuring the proximity between wells means determining which wells are similar in terms of physicochemical and microbiological parameters, in order to form groups of wells based on their proximity. Intuitively, two wells are close if their coordinates in $\mathbb{R}^p$, the space of parameters, are close. In other words, if the observations made on the p parameters are close. To quantify this proximity, we need to associate a measure of this proximity with the space $\mathbb{R}^p$. In other words, a measure of distance between the wells. Furthermore, PCA can also be used to obtain graphical representations of distances between individuals and correlations between variables. PCA is also a method of dimension reduction (construction of a small number of synthetic variables (axis) summarizing the initial variables as best as possible). In this study, the eigenvalue extraction method was applied to the correlation matrix (Fig. 2) to determine the principal components. The results are presented in Fig. 3 and Table 2. Combining the criteria of Kaiser, the scree plots and the proportion of variance explained, the number of factors to retain is five. Thus, in the analysis, only these first five principal components were chosen and the other components were omitted. It is very important to study the correlations between the new synthetic dimensions

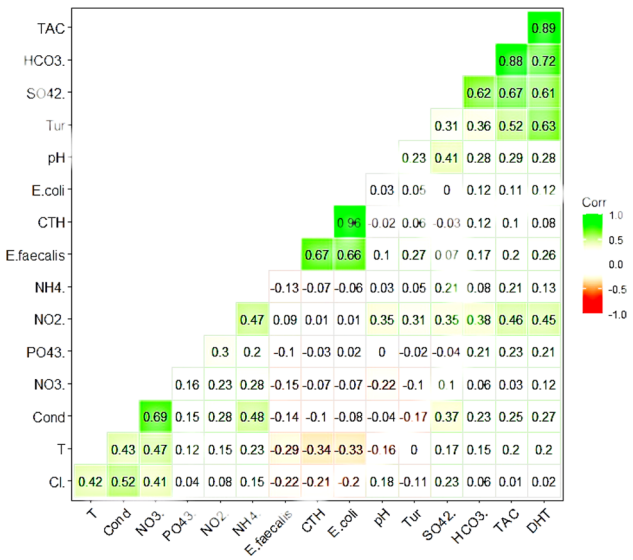

Fig. 2 Correlation matrix between the parameters

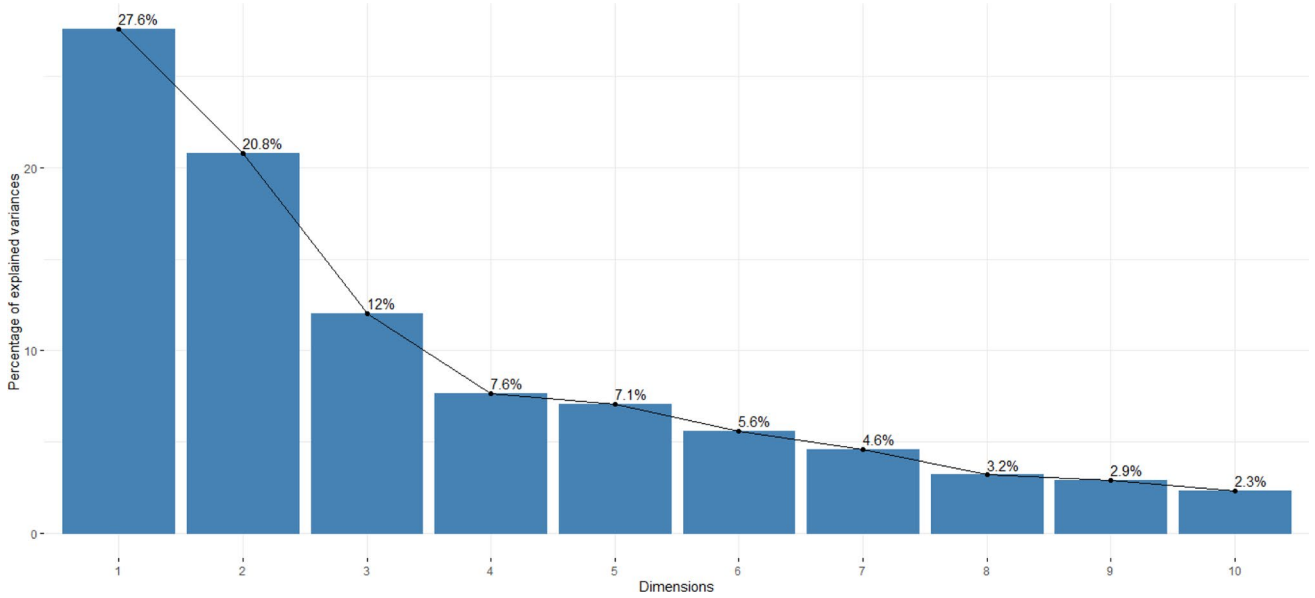

Fig. 3 Principal components explain of the variance

Table 2 Eigenvalues and percentage of variances on each principal component in PCA

Dimensions	Eigenvalue	Percentage variance	Cumulative percentage variance
Dim1	4.412	27.573	27.573
Dim2	3.328	20.797	48.370
Dim3	1.922	12.015	60.385
Dim4	1.220	7.626	68.011
Dim5	1.131	7.068	75.079
Dim6	0.893	5.583	80.662
Dim7	0.735	4.595	85.257
Dim8	0.509	3.184	88.441
Dim9	0.467	2.918	91.359
Dim10	0.370	2.314	93.673
Dim11	0.320	2.001	95.674
Dim12	0.277	1.730	97.404
Dim13	0.214	1.338	98.742
Dim14	0.128	0.797	99.539
Dim15	0.045	0.278	99.817
Dim16	0.029	0.183	100

and the original variables. These correlation coefficients will finally be used to estimate the relative contributions (ctr) of each original variable (Table 3) in the construction of the principal components. All the wells are then projected into the different planes defined by these principal components. An extract from these projections is shown in Fig. 4.

The factor loading classification method adopted by Liu et al. [21] is used to study the correlations between the variables and the returned principal components. In this classification, the load r is considered strong for $|r| \geq 0.75$, moderate if $0.5 \leq |r| < 0.75$ and weak if $0 < |r| < 0.5$. Fig. 2 shows that, in general, the values of the correlation coefficients show natural physical, chemical and microbiological behavior. Further evaluation of these coefficients shows that the strongest correlations are observed between TAC and DHT (0.891), TAC and HCO_{3^-} (0.875), E.coli and CTH (0.96); moderate correlations between E.coli and E.faecalis (0.655), CTH and E.faecalis (0.665), Tur and DHT (0.632), Cl^- and Cond (0.525), TAC

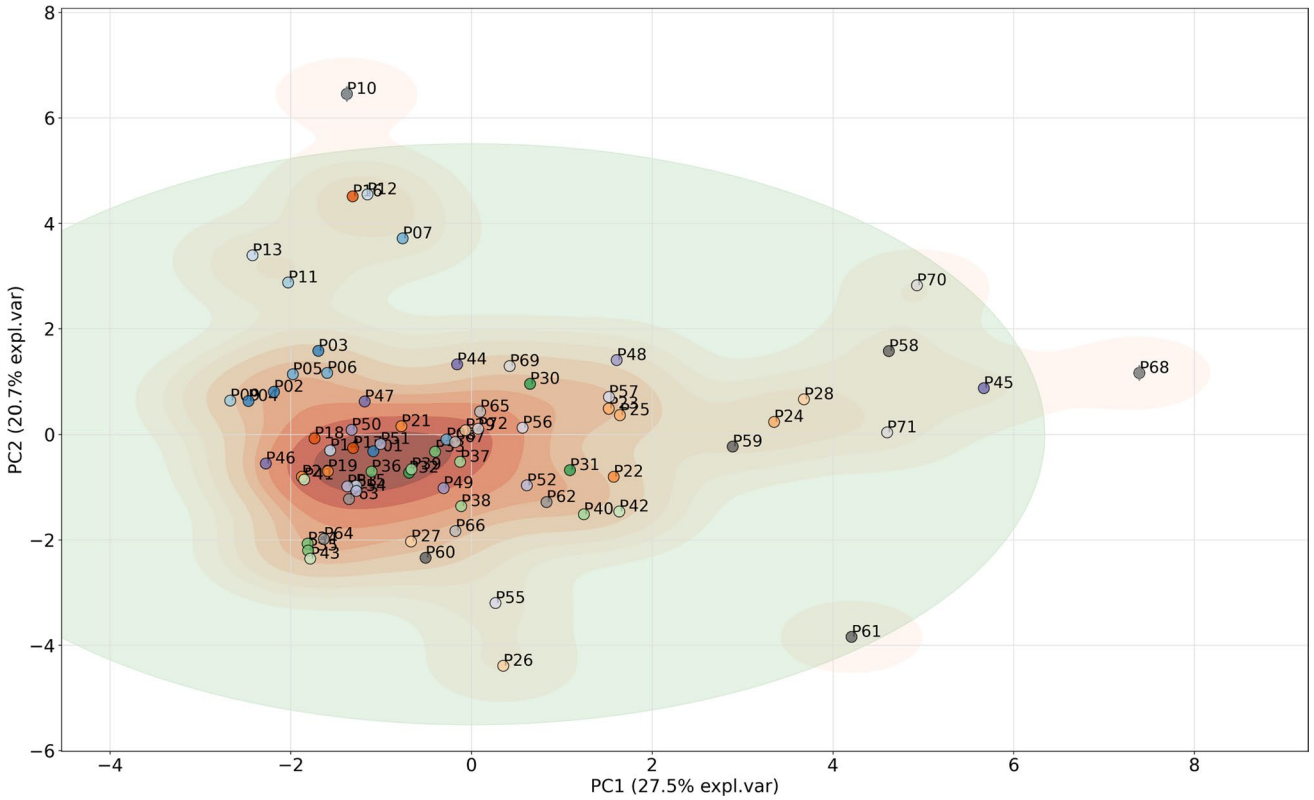

Fig. 4 Projection of wells on the factorial plane (1, 2) in PCA

Table 3 Correlations and contributions of variables on the different component principal

	Dim1		Dim2		Dim3		Dim4		Dim5	
	Cor	ctr1	Cor	ctr2	Cor	ctr3	Cor	ctr4	Cor	ctr5
Tur	**0.519**	6.102	0.338	3.427	− 0.352	**6.445**	− 0.091	0.681	**− 0.790**	2.829
Cond	0.489	5.411	**− 0.557**	**9.328**	0.494	**12.688**	0.170	3.362	0.044	0.170
pH	0.390	3.447	0.188	1.065	− 0.356	**6.580**	0.262	5.618	**0.658**	**38.234**
T	0.326	2.402	**− 0.627**	**11.828**	0.106	0.581	0.061	0.303	− 0.365	**11.800**
NO_{3-}	0.285	1.838	**− 0.549**	**9.051**	**0.565**	**16.618**	0.043	0.150	− 0.159	2.223
NO_{2-}	**0.652**	**9.642**	− 0.052	0.080	0.046	0.109	− 0.361	**10.684**	0.403	**14.352**
NH_{4+}	0.387	3.402	− 0.336	3.394	0.280	4.080	−0.355	**10.335**	0.395	**13.776**
PO_4^{3-}	0.275	1.711	− 0.141	0.597	0.150	1.169	**− 0.678**	**37.671**	0.020	0.035
CL^-	0.246	1.374	**− 0.557**	**9.321**	0.201	2.100	0.486	**19.318**	0.187	3.095
TAC	**0.902**	**18.422**	0.194	1.132	− 0.161	1.343	− 0.048	0.189	− 0.194	3.328
DHT	**0.875**	**17.363**	0.196	1.158	− 0.138	0.991	− 0.032	0.081	−0.251	5.583
SO_4^{2-}	**0.761**	**13.126**	− 0.002	0.000	− 0.162	1.364	0.359	**10.566**	0.010	0.729
HCO_{3-}	**0.809**	**14.826**	0.181	0.985	− 0.123	0.791	0.047	0.178	− 0.201	3.573
CTH	0.058	0.075	**0.737**	**16.307**	**0.617**	**19.824**	0.068	0.319	0.025	0.053
E.coli	0.082	0.151	**0.730**	**16.008**	**0.613**	**19.578**	0.059	0.288	0.042	0.156
E.faecalis	0.177	0.709	**0.737**	**16.318**	0.332	5.739	0.121	1.196	− 0.027	0.065

The moderate and strong correlations of the variables with the main axes are in bold. The same applies to some large-value contributions

and Tur (0.515), NO_{3-} and Cond (0.687), TAC and SO_4^{2-} (0.672), SO_4^{2-} and HCO_{3-} (0.612), HCO_{3-} and DHT (0.719). The other correlations have significantly low values.

The PCA extracted 5 principal components (Fig. 3) which accounted for 75.079 percent of the total variances with eigenvalues ranging from 1.131 to 4.412 (Table 2). PC1 accounted for 27.573 percent, PC2 accounted for 20.797

percent, PC3 accounted for 12.015 percent, PC4 accounted for 7.626 percent, while PC5 accounted for 7.068 percent. The parameters defining PC1 are TAC, DHT, SO_4^{2-}, HCO_{3-}, NO_{2-} and Tur; PC2 are CTH, E.coli, E.faecalis, Cond, T, Cl^- and NO_{3-}; PC3 are CTH, E.coli, Tur. Cond, pH and NO_{3-}; PC4 are NO_{2-}, NH_{4+}, PO_4^{3-}, Cl^- and SO_4^{2-} while PC5 is defined by pH, T, NO_{2-} and NH_{4+}. The PC1 can be interpreted as metal cations (calcium, magnesium), hydroxides, bicarbonates, carbonates, and turbid water component. The outcome for the PC1 is consistent with those of the study of [37] and [38]. The second and third components indicate microbial components. The fifth component indicates turbid and acidic water.

The parameters are then projected onto the different factorial planes. They are correctly projected onto the different planes when the end of the projected vector approaches the unit circle. An extract of the projections of the 16 variables onto the factorial planes (1,2) and (1,3) is given in Fig. 5. Three groups of variables can be distinguished. The first group is made up of the parameters DHT, TAC, HCO_{3-} and SO_4^{2-}. They are strongly correlated to the first axis. The second group is composed of the CTH, E.coli and E.faecalis parameters. They have a moderate correlation with the second axis. CTH and E.coli also have a moderate correlation with the third axis. While the third group is composed of the variables Cond, T, NO_{3-} and Cl^-. They have a moderate correlation with the second axis. This third group is opposed on the second axis to the second group. Taking into account Fig. 2, TAC, DHT, HCO_{3-} and are strongly correlated. Using this natural property of water, the TAC measurement is used directly to estimate the DHT and HCO_{3-} values of the water. This result is not consistent with those of [20, 28] and [34] who have shown that groundwater quality can be accurately predicted solely by measuring electrical conductivity. Moreover, the measurement of E.coli could be sufficient to predict water quality with regard to the parameters CTH and E.faecalis. The correlations obtained between the microbiological parameters studied are similar to those of [3] and [16]. Finally, the correlations obtained between the parameters (Cond, NO_{3-} and Cl^-) and also with (SO_4^{2-}, Tur, DHT and TAC) are similar to the results of [23, 26] and [36].

Then, the proximity of the wells is studied in order to determine the wells that are similar in terms of physicochemical and microbiological parameters. This will make it possible to form homogeneous groups of wells. This takes into account the respective coordinates of the principal components (Table 4) and the CFA method, which is used to study wells and parameters simultaneously in order to highlight correspondences. The eigenvalue extraction method was chosen for this purpose. Using the proportion of variance explained (Table 5), the number of factors to be retained is two. Consequently, the analysis will be limited to this design. Figure 6 shows the position of the wells and the parameters studied. Table 6 shows the partial correlations and partial contributions of the physicochemical and microbiological parameters in relation to the factors. The first factor, which accounts for 49.077 percent of the total variance, has a strong positive correlation with Cond (0.884), NO_{3-} (0.806) and CL^- (0.755); moderate correlation with pH (0.614), T (0.676), HCO_{3-} (0.613), TAC (0.599) and weak correlation with DHT (0.483), SO_4^{2-} (0.458), NH_{4+} (0.388), PO_4^{3-} (0.314). Parameters Cond, HCO_{3-}, TAC, CL^-, NO_{3-} and T contribute more to the inertia of this axis. The factor 1 represent physicochemical component presented in PCA study.

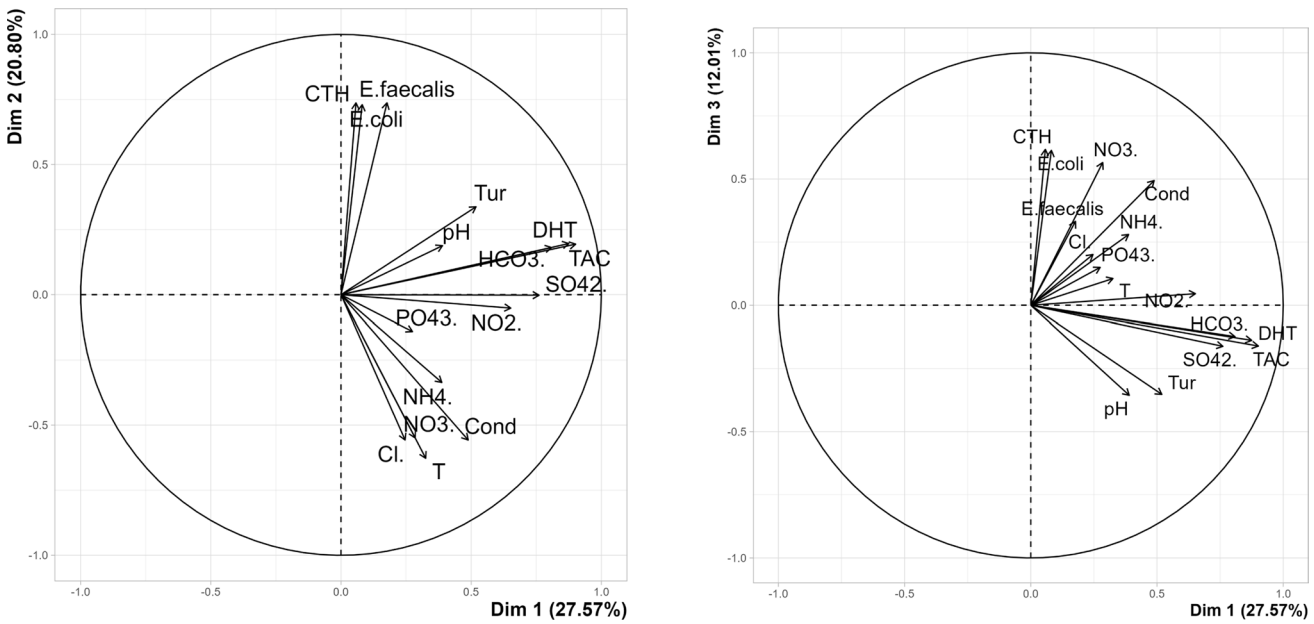

Fig. 5 Projection of variables on the factorial planes (1, 2) and (1, 3) in PCA

Table 4 Coordinates of the wells projected onto the principal components

Wells	PC1	PC2	PC3	PC4	PC5	Wells	PC1	PC2	PC3	PC4	PC5
P01	−1.09	−0.32	−1.56	0.22	0.28	P37	−0.13	−0.52	−0.27	−0.77	−0.63
P02	−2.19	0.8	−1.97	0.53	0.59	P38	−0.11	−1.36	1.24	2.21	1.62
P03	−1.69	1.58	0.04	0.06	0.31	P39	−0.66	−0.66	−1.51	−0.16	0.05
P04	−2.47	0.63	−1.54	1.04	0.37	P40	1.24	−1.51	0.08	−0.97	−0.19
P05	−1.98	1.14	−1.68	1.37	0.55	P41	−1.85	−0.86	0.42	0.12	−0.01
P06	−1.59	1.16	−1.63	0.27	−0.07	P42	1.63	−1.46	0.08	−0.88	−0.01
P07	−0.76	3.72	0.46	−0.18	−0.07	P43	−1.79	−2.36	0.79	−0.27	0.23
P08	−0.28	−0.09	−0.24	−1.61	1.14	P44	−0.16	1.32	−1.31	0.55	0.22
P09	−2.67	0.64	−1.32	0.35	0.47	P45	5.67	0.87	0.39	3.34	−0.13
P10	−1.38	6.46	3.73	−0.01	0.56	P46	−2.28	−0.55	−1.28	0.16	0.24
P11	−2.03	2.88	0.03	0.09	0.53	P47	−1.18	0.62	−2.25	0.63	0.37
P12	−1.15	4.55	2.65	−0.25	0.03	P48	1.61	1.41	−1.14	1.64	−1.09
P13	−2.42	3.39	1.21	0.30	0.64	P49	−0.31	−1.02	−0.08	−0.59	−0.69
P14	−1.56	−0.30	−0.49	0.24	−0.49	P50	−1.33	0.09	0.19	0.35	−0.24
P15	−1.28	−0.97	−0.96	−0.53	−0.33	P51	−1.01	−0.18	−0.11	−0.54	−0.07
P16	−1.32	4.52	3.17	0.02	0.19	P52	0.61	−0.97	0.12	−1.16	0.38
P17	−1.31	−0.26	−0.02	0.08	−0.32	P53	−1.38	−0.98	−0.04	0.56	0.28
P18	−1.74	−0.07	−1.63	0.83	−0.32	P54	−1.27	−1.07	−1.15	1.38	−0.19
P19	−1.59	−0.69	−0.25	1.28	−0.11	P55	0.27	−3.19	1.40	2.39	−0.39
P20	−1.88	−0.80	−0.95	0.29	−0.61	P56	0.57	0.13	1.58	−0.57	−0.47
P21	−0.78	0.16	−0.88	0.31	−0.36	P57	1.52	0.71	−0.73	−0.11	−0.46
P22	1.57	−0.8	−1.17	−0.89	−0.65	P58	4.62	1.58	−2.86	1.95	−2.03
P23	1.52	0.49	0.96	−0.51	−1.22	P59	2.89	−0.23	−1.11	−0.40	−0.92
P24	3.35	0.24	−1.03	−1.57	−0.99	P60	−0.51	−2.34	−0.05	−0.70	−0.19
P25	1.64	0.37	−2.79	−3.07	6.24	P61	4.21	−3.84	3.42	3.21	4.33
P26	0.35	−4.39	2.84	−1.26	−0.76	P62	0.83	−1.28	1.02	−0.74	−0.48
P27	−0.67	−2.03	0.99	−1.43	−0.36	P63	−1.36	−1.22	1.35	−0.8	0.29
P28	3.68	0.66	−0.02	−0.70	0.15	P64	−1.63	−1.98	0.21	0.24	−0.8
P29	−0.07	0.08	−0.25	−1.32	−0.03	P65	0.09	0.43	1.47	−0.84	−0.62
P30	0.67	0.96	−0.75	−0.69	−0.69	P66	−0.18	−1.83	2.52	0.61	−0.74
P31	1.09	−0.68	0.93	−0.55	−0.61	P67	−0.18	−0.15	1.46	1.14	−0.33
P32	−0.69	−0.72	−0.28	−0.51	−0.73	P68	7.39	1.16	0.77	−2.15	−0.71
P33	−0.4	−0.33	0.24	−0.92	−0.52	P69	0.42	1.29	0.76	−0.84	−0.04
P34	−1.81	−2.07	0.24	−0.41	−0.27	P70	4.93	2.83	−0.35	0.56	0.34
P35	−1.81	−2.2	0.22	0.14	−0.07	P71	4.59	0.04	−1.91	0.57	0.39
P36	−1.11	−0.70	1.13	−0.88	−0.09	P72	0.07	0.11	−0.58	0.74	0.28

Factor 2 explains 30.7 percent of the total variance and is strongly correlated with E.faecalis (0.896), moderately correlated with CTH (0.575) and E.coli (0.506). The parameters E.faecalis, CTH and E.coli contribute more to the inertia of this axis. The factor 2 represent microbiological component presented in PCA study.

Wells P22, P23, P24, P25, P28, P42, P45, P48, P57, P58, P59, P61, P68, P70 and P71 (Table 4) are well projected onto the first factorial plane because their coordinates on axis 1 are large. These wells share a high frequency on the axis for the variables DHT, TAC, HCO_{3^-} and $SO_4^{2^-}$. Wells P03, P07, P08, P10, P11, P12, P13, P16, P17, P23, P28, P29, P30, P31, P33, P36, P38, P44, P45, P48, P50, P51, P57, P57, P58, P61, P62, P63, P65, P66, P67, P68, P69, and P70 are also well projected on this plan but it is on axis 2 that their coordinates are large. These wells take on large values on axis 2 for the variables CTH, E.coli and E.faecalis. The wells P10, P12, P13, P16, P26, P55, P56, P61, P63, P65, P66, P67; P38, P45, P55, 58, P61; and P25, P38, P61 are well projected onto PC3, PC4 and PC5 respectively. The data makes it possible to characterize them. These wells share relatively high concentrations of certain parameters among all

Table 5 Eigenvalues and percentage of variances on each component principal in CFA

Dimensions	Eigenvalue	Percentage variance	Cumulative percentage variance
Dim1	0.139	49.077	49.077
Dim2	0.087	30.717	79.794
Dim3	0.030	10.526	90.320
Dim4	0.012	4.385	94.705
Dim5	0.006	2.039	96.744
Dim6	0.004	1.427	98.171
Dim7	0.001	0.526	98.697
Dim8	0.001	0.441	99.138
Dim9	0.001	0.326	99.464
Dim10	0.001	0.242	99.706
Dim11	0.001	0.194	99.900
Dim12	0.0001	0.051	99.951
Dim13	7.204E-05	0.025	99.977
Dim14	4.815E-05	0.017	99.994
Dim15	1.768E-05	0.006	100

Table 6 Correlations and contributions between variables and factors

Parameter	Factor 1	Factor 2	ctr1	ctr2
Tur	0.112	0.125	0.785	1.403
Cond	**0.884**	0.004	**53.478**	0.369
pH	**0.614**	0.006	0.662	0.01
T	**0.676**	0.002	4.138	0.023
NO_{3^-}	**0.806**	0.002	4.248	0.013
NO_{2^-}	0.296	0.006	0.012	0.00
NH_{4+}	0.388	0.008	0.085	0.003
PO_4^{3-}	0.314	0.002	0.032	0.00
CL^-	**0.755**	5.098×10^{-5}	4.443	0.00
TAC	**0.599**	0.021	7.551	0.415
DHT	0.483	0.067	2.422	0.538
SO_4^{2-}	0.458	0.011	1.179	0.045
HCO_{3^-}	**0.613**	0.006	**9.32**	0.154
CTH	0.216	**0.575**	4.076	**17.360**
E.coli	0.202	**0.506**	3.478	**13.893**
E.faecalis	0.089	**0.896**	4.089	**65.771**

The moderate and strong correlations of the variables with the main axes are in bold. The same applies to some large-value contributions

the parameters studied. For example, P25 has the highest pH (11.515); P26 (594 $\mu S/cm$) and P61 (545.25 $\mu S/cm$) have the greatest conductivities.

3.3 One-way ANOVA results

One-way analysis of variance is used to study the effect of wells on physicochemical and microbiological parameters. It shows whether the average for each parameter is the same in the different groups studied. Note that correlated parameters will have similar responses in the ANOVA. Based on PCA results, the result obtained with the *Cond* parameter will be similar to that obtained with NO_{3^-} and CL^-. The same is true of the result obtained with the TAC parameter. It will be the same as that for DHT, SO_4^{2-} and HCO_{3^-}. Finally, the result obtained with the parameter E.coli will also be

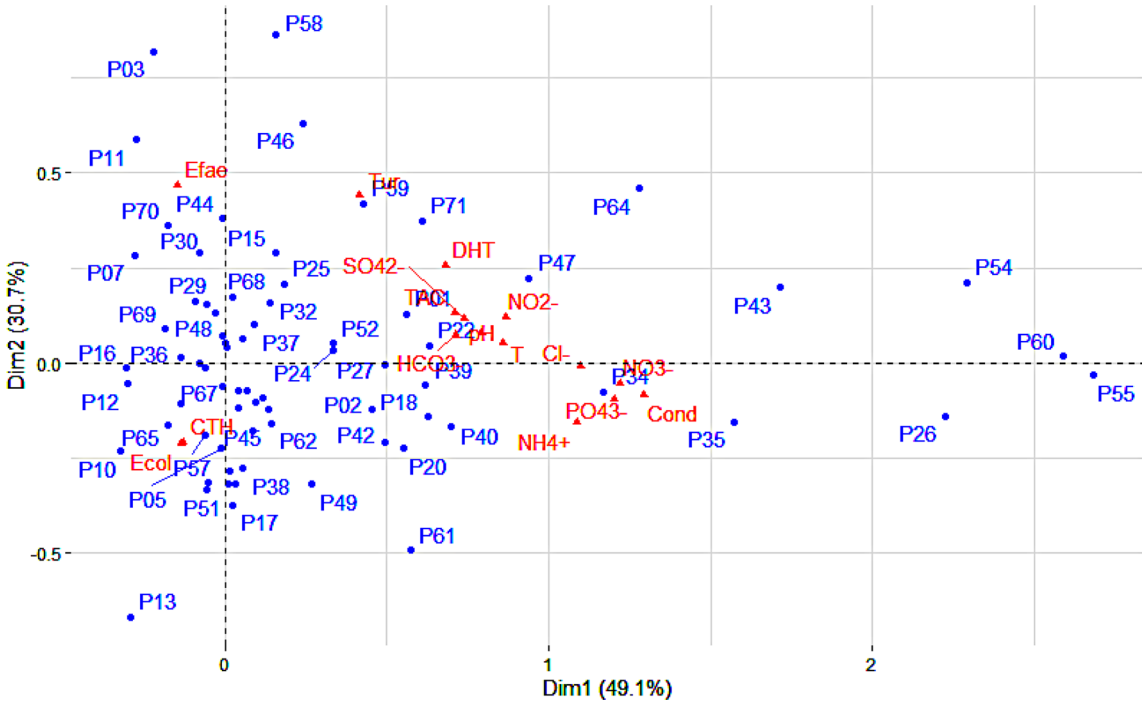

Fig. 6 Projection of the wells and parameters studied on the first factorial plane in CFA

Table 7 Wells classified by concentration levels of the conductivity parameter

Wells	Cond	Level	Wells	Cond	Level	Wells	Cond	Level
P26	594	a	P64	173.075		P51	84.275	ghij
P61	545.25	ab	P38	172.875		P54	67.475	
P27	440.95	abc	P22	170.7		P48	67.35	
P68	430.75		P37	166.1		P16	67.025	
P56	400.9	bcd	P67	159.45		P14	64.55	hij
P31	313.75	cde	P49	158.975		P58	62.6	
P52	311.75		P65	158.1		P11	62.425	
P24	295.75	cdef	P33	156.775		P01	62.35	
P55	295.25		P69	151.175		P46	58.7	
P66	273.75		P41	143.45	efghij	P19	57.55	
P45	272.9	cdefg	P32	138.575		P20	57.5	
P40	271.575		P50	138		P07	52.65	
P63	267.75		P25	128.575		P12	48.05	ij
P62	259.35	cdefgh	P36	127.075		P02	44.625	
P28	236.575	defghi	P53	120.8		P09	42.0225	
P71	224.825		P30	117.55		P18	40.7	
P43	222.7	defghij	P17	116.875		P06	40.15	
P60	221.2		P72	112.05		P05	37.925	
P35	193.65		P03	106.675	fghij	P47	35.65	
P59	180.025		P29	104.925		P15	34.55	
P08	179.275	efghij	P70	96.95		P10	32.5	j
P23	178.8		P21	94	ghij	P39	31.6	
P34	178.725		P44	90.4		P04	25.853	
P42	177.325		P57	87.175		P13	24.725	

Table 8 Wells classified by concentration levels of the total alkalinity contents parameter

Wells	TAC	Levels	Wells	TAC	Levels	Wells	TAC	Levels
P68	216.25	a	P29	73.75		P54	51.25	
P45	207.5	ab	P52	73.75		P60	51.25	
P71	178.75	abc	P33	72.5		P17	50	
P58	176.25		P37	72.5		P09	48.75	
P24	148.75	bcd	P51	72.5		P13	48.75	
P70	147.5	bcde	P07	70	fghijkl	P27	47.5	
P28	133.75	cdef	P72	70		P53	47.5	
P31	126.25	cdefg	P21	68.75		P03	46.25	hijkl
P59	109.25	defgh	P12	67.5		P26	46.25	
P22	107.5	defghi	P56	67.5		P04	45	
P48	105	defghij	P55	66.25		P11	45	
P42	101.25		P10	65		P19	45	
P40	100		P65	65		P20	43.75	
P23	98.75		P32	63.75		P05	42.5	
P30	98.75		P15	57.5		P36	42.5	
P39	98.75	defghijk	P50	57.5	ghijkl	P66	38.75	
P61	92.5		P06	56.25		P38	37.5	
P57	85		P67	55		P41	37.5	
P62	83.75		P16	53,75		P64	36.25	
P44	80		P18	53.75		P34	35	ijkl
P49	78.75	efghijkl	P02	51.5		P35	32.5	jkl
P47	75		P25	51.251		P63	31.25	kl
P69	75		P08	51.25		P43	28.75	
P01	73.75		P14	51.25		P46	26.25	l

similar to that obtained with the parameter CTH and E.faecalis. Consequently, the parameters (*Cond*, E.coli and TAC) were selected for testing. Shapiro-Wilk's normality test gave the following results. The $p - value < 2.2 \times 10^{-16}$ for the *Cond*, TAC and E.coli parameters. This confirm the non-normality nature of this parameters. The non-parametric Kruskal-Wallis test is therefore necessary for the study. Significant differences ($p - value = 0.0071 < 0.05$) were observed in E.coli between wells. Significant difference is observed in TAC ($p - value = 1.504 \times 10^{-7} < 0.05$). Electrical conductivity concentration shows significant differences between wells ($p < 2.2 \times 10^{-16}$). This indicates that the wells have an effect on these parameters. It also means that the factors influencing the well parameters are different. Duncan's multiple comparison test carried out on the *Cond*, TAC and E.coli parameters gave the results summarized in Table 7, Table 8 and Table 9 respectively. These results show significant differences between wells for the parameters studied. Table 7 shows sixteen significant levels for *Cond*. Nineteen significant levels for TAC are observed in Table 8. Eight significant levels for E.coli are observed in Table 9. With regard to the parameters studied, if three wells belong to different significant levels, namely *a*, *ab* and *c* for example, the well at level *a* is close to the well at level *ab* but different from the well at level *b*. Similarly, the well of level *b* is close to the well of level *ab* but different from the well of level *a* and so on. In other words, although they are in the same study area, the parameters of the well water evolve differently. The advantage of this classification is as follows. If we want to treat all the seventy-two wells for E.coli, these wells must be grouped into eight sub-groups. Each sub-group must be treated differently depending on the concentration level.

3.4 Classification of wells water samples by SOM

The concept of the SOM algorithm is to conduct a nonlinear classification of complicated data sets by recognizing similar patterns. In this work, the input layer consists of vectors representing seventy-two (72) wells, each of which contains sixteen (16) components representing the 16 physicochemical and microbiological parameters of the well water studied. The output layer is composed of 16 neurons (4 rows×4 columns). This size was chosen for the output map after convergence of the algorithm. Figure 7 shows the role of parameters in defining the different areas of the topological map and

Table 9 Wells classified by concentration levels of the escherichia coli parameter

Wells	E.coli	Levels	Wells	E.coli	Levels	Wells	E.coli	Levels
P10	5650	a	P33	957.25		P04	406.25	
P16	4242.75	ab	P48	950.75		P59	385	
P12	4125	abc	P17	895		P09	352.5	
P13	3867.5	abcd	P72	880		P22	329	
P65	2329.5		P24	867.5		P71	324.5	
P07	2205		P31	839		P51	293.25	
P69	2125		P03	785		P40	252.5	
P68	2065		P49	751.5		P39	210	
P45	1817.5		P37	717.5		P02	200.75	de
P67	1750.25		P19	692.5		P20	188	
P23	1710		P11	690.5		P58	180	
P70	1702.5		P41	660	bcde	P15	155	
P56	1666		P06	657.5		P35	146.25	
P28	1470	bcde	P14	627		P34	130	
P66	1286.5		P52	612.75		P46	126.75	
P57	1175		P27	572.5		P01	117.75	
P62	1148		P44	567.5		P26	112.5	
P36	1101.25		P05	565		P18	97.75	
P61	1096.5		P21	532.75		P43	63.75	
P29	1095		P38	532.5		P47	51.5	e
P50	1057.5		P32	532.25		P64	45.75	
P08	1057.5		P25	527.5		P55	12	
P30	1003.5		P53	488.5		P60	11	
P63	1000.25		P42	435	cde	P54	2	

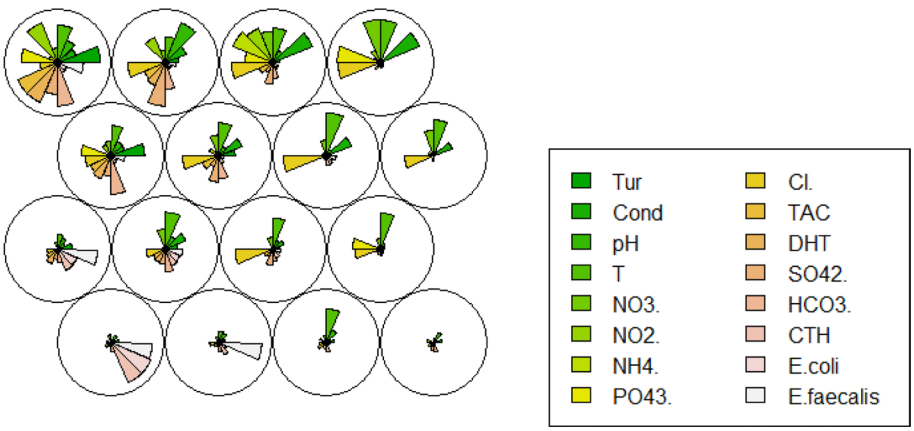

Fig. 7 Graph of the nature of the different zones on the map in relation to the parameters

Table 10 Distribution of wells in each node

P45 P58 P68 P70 P71	P24 P25P28	P38 P61	P26 P55 P66
P48 P59	P22 P31 P40 P42 P52 P57 P62	P27 P36 P49 P63	P34 P35 P41 P43 P60 P64
P07	P23 P30 P56 P65 P69	P08 P29 P32 P33 P37 P51	P14 P15 P17 P19 P20 P53 P54 P67
P10 P12 P13 P16	P03 P11	P01 P18 P21 P39 P46 P50 P72	P02 P04 P05 P06 P09 P44 P47

Table 10 shows the wells of each node. With the exception of the south-western part, almost the entire map is characterized by parameters in green and yellow (Tur, Cond, pH, T, NO_{3^-}, NO_{2^-}, NH_{4^+}, $PO_4^{3^-}$, Cl^-, TAC and DHT). The south-western part is characterized by the variables in pink (E.coli, CTH and E.faecalis). A graph (Fig. 8) of each variable is produced to

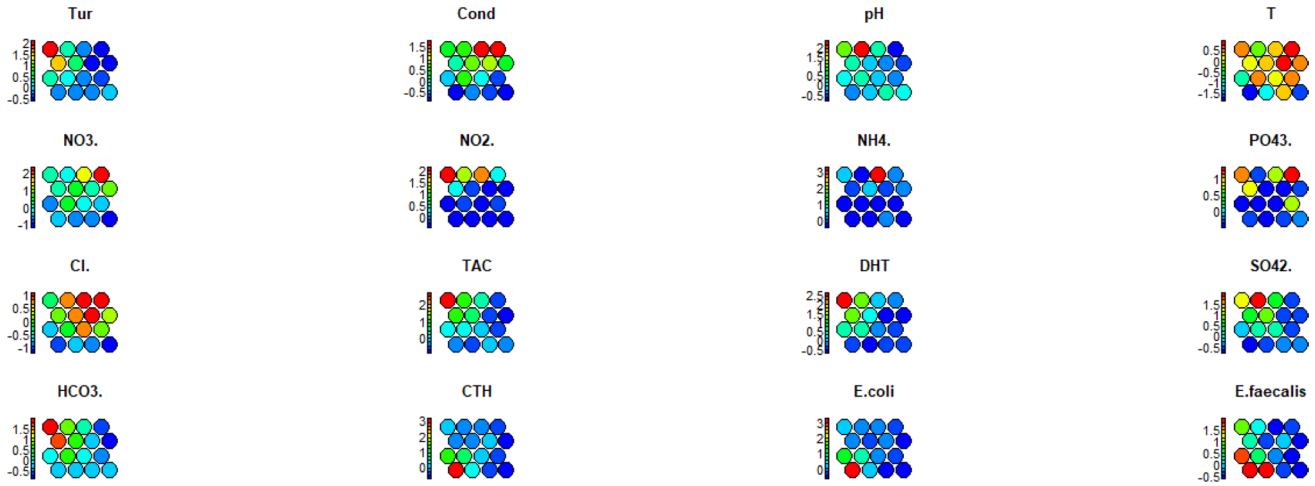

Fig. 8 Graph of the nature of the different zones on the map in relation to each parameter

Table 11 Relevance of parameters

E.coli	HCO_{3^-}	TAC	CTH	T	DHT	NO_{3^-}	$SO_4^{2^-}$
0.8988	0.8971	0.8886	0.8653	0.8536	0.8397	0.8281	0.8035
Cond	E.faecalis	NO_{2^-}	CL^-	Tur	pH	$PO_4^{3^-}$	NH_{4^+}
0.7807	0.7677	0.7325	0.7299	0.7242	0.6310	0.5123	0.4736

show the correlations between them and this graph can help to summarize the effective parameters of the wells in each node. 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 for each parameter [27]. The similar colors between the variables correspond to a positive correlation. This can be illustrated between the variables HCO_{3^-}, DHT, $SO_4^{2^-}$ and TAC. There is also a positive correlation between these parameters and Tur. Cond and NO_{3^-} are positively correlated. There is also a positive correlation between E.coli, E.faecalis and CTH variables. These results confirm those obtained previously. On the other hand, T, $PO_4^{3^-}$, CI^-, NO_{2^-}, NH_{4^+} and pH vary independently of each other. The same idea can be expressed by using a dispersion indicator such as variance. The variance weighted by the number of nodes is calculated. It then becomes possible to rank their role. The important variables (because they induce the strongest contrasts) appear in first position (Table 11). The parameters $PO_4^{3^-}$ and NH_{4^+} are the least influential. This means that the conditional averages tend to be homogeneous across the map. These results confirm what we have seen in the various graphs. A detailed summary of the parameters for each well is presented in (Fig. 8). The dark red nodes represent high values of each parameter.

Wells P01, P03, P05, P06, P07, P08, P10, P11, P12, P13, P15, P16, P17, P18, P21, P22, P23, P24, P25, P28, P29, P30, P37, P42, P44, P45, P47, P48, P52, P56, P57, P58, P59, P62, P65, P67, P68, P69, P70, P71 and P72 are mainly characterized by high Tur concentrations (13.9 NTU, 162.43 NTU). Wells P08, P23, P24, P26, P27, P28, P31, PP34, P35, P38, P40, P42, P43, P45, P52, P55, P56, P59, P60, P61, P62, P63, P64, P66, P68 and P71 are mainly characterized by high Cond concentrations (172.875 μS/cm, 594 μS/cm). Wells P01, P03, P05, P08, P22, P23, P24, P25, P28, P29, P30, P39, P42, P44, P45, P46, P59, P61, P68, P69, P70, P71 and P72 are mainly characterized by high pH (5.24, 11.51). Wells P15, P22, P26, P27, P35, P36, P37, P39, P49, P58, P60, P62, P63, P64, P65, P66, P68, P69 and P72 are mainly characterized by high T (28.425 °C, 29.08 °C). Wells P23, P26, P42, P43, P55 and P66 are mainly characterized by high NO_{3^-} concentrations (27.018 mg/l, 44.16 mg/l). Wells P25, P26, P28, P38, P42, P45, P55, P58, P59, P61, P68, P70 and P71 are mainly characterized by high NO_{2^-} concentrations (0.11 mg/l, 0.54 mg/l). Wells P38, P61 and P72 are mainly characterized by high NH_{4^+} concentrations (1.1225 mg/l, 3.82 mg/l). Wells P01, P02, P03, P04, P05, P06, P07, P08, P10, P11, P12, P13, P14, P15, P16, P17, P18, P19, P20, P21, P22, P23, P24, P26, P27, P28, P29, P30, P31, P32, P35, P36, P37, P38, P39, P40, P41, P42, P43, P44, P45, P47, P48, P49, P50, P53, P54, P55, P56, P57, P58, P59, P60, P61, P62, P63, P64, P65, P66, P67, P68, P69, P70, P71 and P72 are mainly characterized by high $PO_4^{3^-}$ concentrations (0.0375 mg/l, 0.75 mg/l). Wells P27, P29, P34, P40, P42, P49, P61, P63, P66 and P68 are mainly characterized by high CI^- concentrations (27.975 mg/l, 44.7 mg/l). Wells P24, P45, P58, P68, P70 and P71 are mainly characterized by high TAC (147.5 mg/l, 216.25 mg/l). Wells P45, P58 and P68 are mainly characterized by high DHT concentrations (86.25 mg/l, 106.25 mg/l). Wells P24,

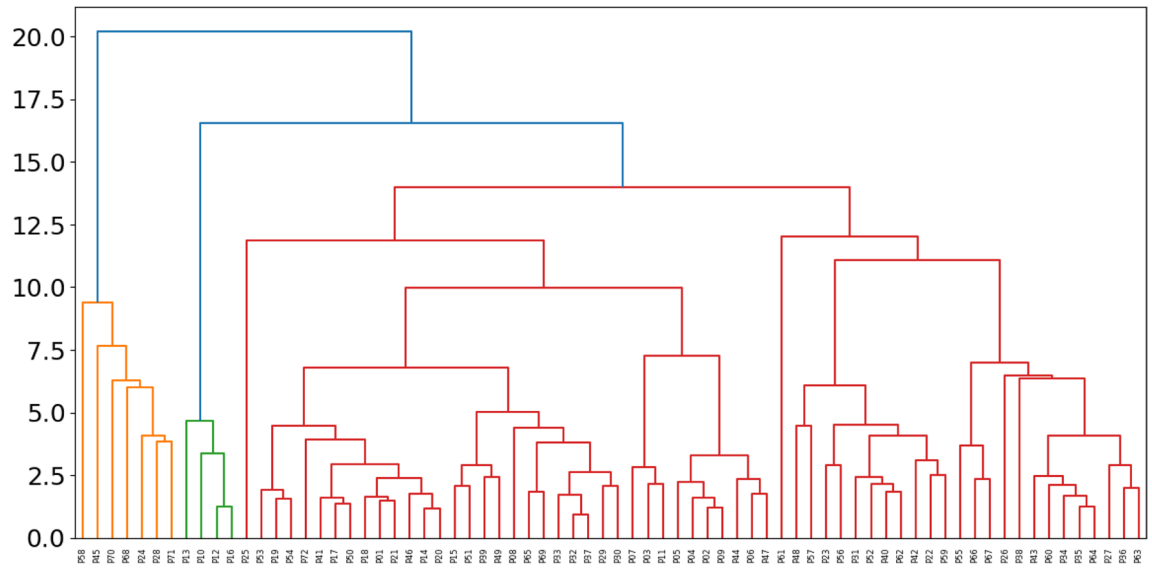

Fig. 9 Dendrogram of wells from M'pody village obtained using the Ward method

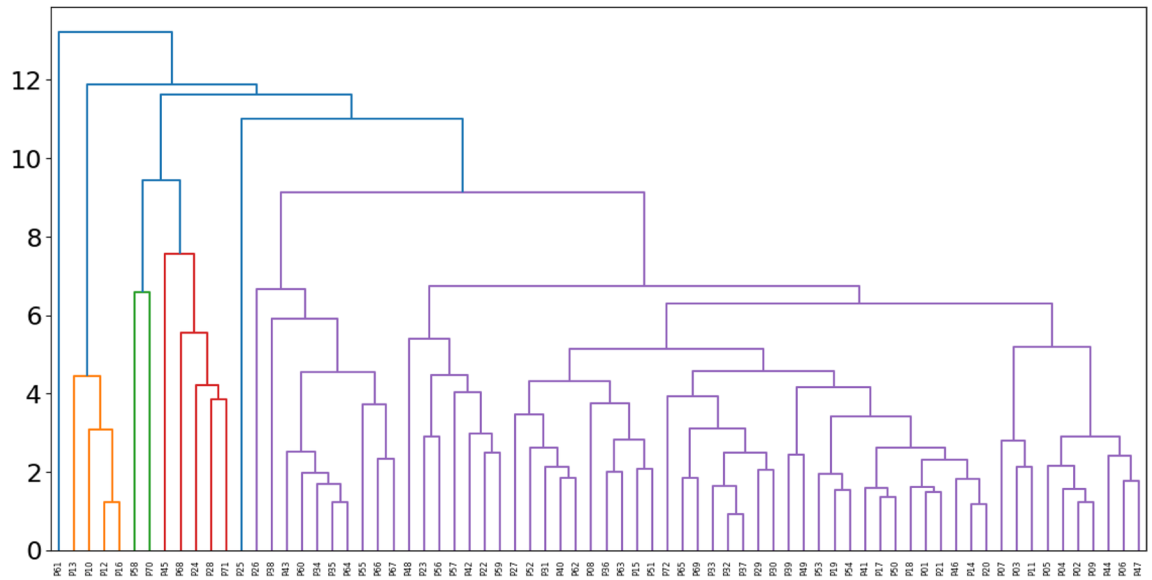

Fig. 10 Dendrogram of wells from M'pody village obtained using the Complete method

P25, P28, P51, P59, P68 and P71 are mainly characterized by high SO_4^{2-} concentrations (21.25 mg/l, 40.5 mg/l). Wells P12, P23, P24, P28, P30, P31, PP39, P40, P42, P44, P45, P48, P49, P56, P57, P58, P59, P62, P68, P70, P71 and P72 are mainly characterized by high HCO_{3-} concentrations (93.75 mg/l, 216.25 mg/l). Wells P10, P12, P13 and P16 are mainly characterized by high CTH concentrations (4367.5 UFC/250 ml, 7400 UFC/250 ml). Wells P10, P12, P13 and P16 are mainly characterized by high E.coli concentrations (3868 UFC/250 ml, 5650 UFC/250 ml). Wells P01, P02, P03, P04, P05, P06, P07, P08, P09, P10, P11, P12, P13, P14, P15, P16, P17, P19, P21, P22, P23, P24, P25, P27, P28, P29, P30, P31, P32, P33, P34, P36, P37, P38, P39, P40, P41, P42, P43, P44, P45, P46, P47, P48, P49, P50, P51, P52, P53, P56, P57, P58, P59, P62, P63, P64, P65, P66, P67, P68, P69, P70, P71 and P72 are mainly characterized by high E.faecalis concentrations (84.5 UFC/250 ml, 3913 UFC/250 ml).

Once the Kohonen map has been obtained, the HCA is used to group the seventy-two wells based on the similarity of the responses to physicochemical and microbiological parameters. Ward's method and the complete method give better results than other existing methods. Ward's method (Fig. 9) gives three clusters as in [8]. The Complete method

(Fig. 10) gives five clusters. Cluster1 of the Complete method is formed by well 61. It is the only well which has a good projection on all the factorial planes of the PCA (Table 4). Cluster 2 of the complete method is cluster 2 of Ward's method. Cluster 3 and cluster 4 of the complete method form Ward's cluster 1. Finally, cluster 5 corresponds to Ward's cluster 3. The results obtained using Ward's method will therefore be used for the analysis. The clusters obtained are very close to those obtained by PCA.

Moreover, cluster 1 contains seven wells (P24, P28, P45, P58, P68, P70 and P71) and represents 09.72 percent of the total number of the wells. These wells have the largest coordinates on PC1. It is mainly characterized by high concentrations of microbiological elements (CTH [362 UFC/250 ml, 2365 UFC/250 ml]; E.coli [180 UFC/250 ml, 2065 UFC/250 ml]; E.faecalis [716 UFC/250 ml, 3542.5 UFC/250 ml]). These waters are very turbid [13.54 NTU, 162.43 NTU]; acidic [5.09, 6.48], with conductivity [62.6 μS/cm, 430.75 μS/cm] and a temperature [27.875 °C, 28.525 °C] higher than the WHO standard. They verify the WHO standards with regard to the parameters: NO_{3^-}, NO_{2^-}, NH_{4^+} except P68 (1.02 mg/l), $PO_4^{3^-}$ except P45 (0.7525 mg/l), Cl^-, TAC, DHT, $SO_4^{2^-}$ and HCO_{3^-}.

Cluster 2 contains four wells (P10, P12, P13, P16) and represents 05.56 percent of the total number of the wells. These wells have the largest coordinates on PC2. It is mainly characterized by very high concentrations of microbiological elements (CTH [4367.5 UFC/250 ml, 7400 UFC/250 ml]; E.coli [3867.5 UFC/250 ml, 5650 UFC/250 ml]; E.faecalis [84.5 UFC/250 ml, 3913 UFC/250 ml]). These waters are turbid [15.7 NTU, 20.7475 NTU] and acidic [4.625, 5.19], with very low conductivity [24.725 μS/cm, 67.025 μS/cm] and a temperature [26.1 °C, 26.3 °C] higher than the WHO standard. They verify the WHO standards with regard to the parameters NO_{3^-}, NO_{2^-}, NH_{4^+}, $PO_4^{3^-}$, Cl^-, TAC, DHT, $SO_4^{2^-}$ and HCO_{3^-}. This cluster represent the microbiological component (F2) previously described in PCA/CFA study.

Cluster 3 includes the largest number of wells (sixty-one) and represents 84.72 percent of the total wells. It is characterized by concentrations of microbiological elements with high variability (CTH [4.25 UFC/250 ml, 2730 UFC/250 ml]; E.coli [2 UFC/250 ml, 2329.5 UFC/250 ml]; E.faecalis [3.5 UFC/250 ml, 3781.25 UFC/250 ml]). These waters are turbid [2.705 NTU, 83.775 NTU], more acidic [4.22, 6.065], with conductivity [25.8525 μS/cm, 440.95 μS/cm] and a temperature [25.875 °C, 28.675 °C] higher than WHO standard. They verify the WHO standards with regard to the parameters: NO_{3^-}, NO_{2^-}, $PO_4^{3^-}$ except P55 (0.5425 mg/l), Cl^-, TAC, DHT, $SO_4^{2^-}$ and HCO_{3^-}. With regard to ammonium, 18.03 percent of wells do not comply with WHO standards. These are the wells, P31 (0.76 mmg/l), P35 (0.565 mg/l), P38 (1.1225 mg/l), P40 (0.675 mg/l), P52 (0.542 mg/l), P55 (0.5525 mg/l), P57 (0.6375 mg/l), P61 (3.8175 mg/l), P62 (0.64 mg/l), P63 (0.52 mg/l) and P72 (1.215 mg/l). Cluster 1 and cluster 2 represent physicochemical component (F1) presented in PCA/CFA study.

4 Conclusion

A multivariate statistical approach was applied to a database comprising sixteen (16) physicochemical and microbiological parameters carried out on two hundred and eighty-eight (288) well water samples from the village M'pody. This technique is very promising, because it makes it possible to understand water quality while highlighting the different correlations that exist between the parameters studied. The study showed that turbidity, conductivity, hydrogen potential and temperature did not meet WHO standards. In addition, the water from all the wells is polluted with faecal bacteria (E.coli, E.faecalis and CTH). It is certainly this faecal pollution that is at the root of this diarrhoea epidemic. It indicates that poor well maintenance is the main factor controlling microbiological pollution of well water in the study area. The logical explanations for this situation could come, on the one hand, from the infiltration of septic tanks located near the wells and, on the other hand, from the run-off of waste water carrying human and animal faecal matter. To prevent epidemics, populations who use well water or surface water should use approved technicians for the construction of latrines. Erect perimeters to protect water points. Learn water treatment techniques. For example, filtering water through layers of granular materials or on granular activated carbon. However, all the physicochemical parameters NO_{3^-}, NO_{2^-}, NH_{4^+}, $PO_4^{3^-}$, Cl^-, TAC, DHT, $SO_4^{2^-}$ and HCO_{3^-} comply with WHO standards. The ANOVA method showed that there were significant differences between the wells for the parameters TAC, Cond and E.coli, due to the specificity and characteristics of each well. In addition, the ANOVA confirmed that human activities were the main factors influencing the physicochemical and microbiological parameters of the wells studied. PCA, CFA and SOM methods are the multivariate analysis techniques used to highlight certain specificities in the structure of the data. Five principal components identified by PCA accounted for 75.079 percent of the total variance. The PCA, CFA and HCA identified the structure of the wells and deduced the main factors controlling the physicochemical and microbiological parameters of the water in these wells. With regard to the CFA, two main factors were identified. The first factor was identified as the physicochemical component with 49.077 percent contribution and the second with 30.717 percent contribution was related to the microbial load. The

physicochemical component is mainly formed by the parameters Cond, HCO_{3^-}, TAC, CL^-, NO_{3^-}, pH, T and the microbial component is linked to E.coli, E.faecalis and CTH. The results of the PCA/FCA are broadly similar to those obtained by applying the Ward and complete methods. However, there are some additional differences, due to the specificity of each method. This study also showed that measuring DHT is therefore sufficient to predict water quality in terms of TAC, HCO_{3^-}, $SO_4^{2^-}$. Similarly, the measurement of E.coli could be sufficient to predict water quality with regard to the parameters CTH and E.faecalis. What's more, major difficulties are often encountered when using traditional PCA/CFA methods. An individual who is poorly represented but whose contribution is significant is eliminated from the analysis (extra individual). On the other hand, there are individuals whose contribution is too large and whose reliability is called into question. In this case, a new study is carried out. To overcome these difficulties, we plan to replace traditional PCA/CFA distances with robust distances such as the Hellinger distance. Subsequently, in order to implement effective planning and support methods for sustainable well water management, multiple linear regression and multi-layer perceptron models can be used to predict the dependent parameters. In practical terms, E.coli can be predicted from CTH and E.faecalis; TAC from DHT, $SO_4^{2^-}$, and HCO_{3^-}; Cond from T, NO_{3^-} and Cl^-.

Acknowledgements The anonymous reviewers and the editor are sincerely acknowledged for their useful and constructive comments.

Author contributions The measuring devices for the physico-chemical parameters, the membrane filtration device for the bacteriological parameters and the water sampling were carried out in the seventy-two wells in the village M'Pody by: Georges Aubin Tchapé Gbagbo, Renaud Franck Djedjro Meless and Christophe N'Cho Amin. Aubin Yao N'Dri defined the draft article. Aubin Yao N'Dri, Stanislas Egomli Assohoun and Cyrille Gueï Okou wrote the article.

Data availability The data sets analyzed during the current study is included in the submission of this paper as a supplementaryinformation file. The data sets are also available from the corresponding author on reasonable request.

Declarations

Competing interests The authors declare that they have no conflict of interest.

References

1. Abba SI, Mohamed AY, Syed MHS, Johnbosco CE, Hussam EE, Johnson CA, Gaurav S, Jamilu U, Nadeem AK, Isam HA. Trace element pollution tracking in the complex multi-aquifer groundwater system of Al-Hassa oasis (Saudi Arabia) using spatial, chemometric and index-based techniques. Environ Res. 2024;249: 118320.
2. Agbasi JC, Ezugwu AL, Omeka ME, Ucheana IA, Aralu CC, Abugu HO, Egbueri JC. More about making profits or providing safe drinking water ?A state-of-the-art review on sachet water contamination in Nigeria. J Environ Sci Health. 2024. https://doi.org/10.1080/26896583.2024.2319009.
3. Asefi M, Zamani-Ahmadmahmoodi R. Analysis of physiochemical and microbial quality of waters of the Karkheh River in southwestern Iran using multivariate statistical methods. Adv Environ Technol. 2018;4(2):75–81.
4. Azaïs JM, Bardet J. Le modèle linéaire par l'exemple : Régression, Analyse de la Variance et Plans d'Expériences. Illustrations numériques avec les logiciels R, SAS et Splus, Paris, Dunod; 2005.
5. Braga FHR, Dutra MLS, Lima NS, Silva GM, Miranda RCM, Firmo WCA, Moura ARL, Monteiro AS, Silva LCN, Silva DF. Study of the influence of physicochemical parameters on the water quality index (WQI) in the Maranhão Amazon, Brazil. Water. 2022;14:1546.
6. Cottrell M, Fort J, Pagès G. Theoretical aspects of the SOM algorithm. Neurocomputing. 1998;21:119–38.
7. Cottrell M, Letrémy P. How to use the Kohonen algorithm to simultaneously analyse individuals in a survey. Neurocomputing. 2005;63:193–207.
8. Dawood AS, Khudier AS, Ahmed Naeemah Bashara AN. Physicochemical Quality Assessment and Multivariate Statistical Analysis of Groundwater Quality in Basrah, Iraq. Int J Eng Technol. 2018;7(4.20):245–50.
9. Duan R, Li P, Wang L, He X, Zhang L. Hydrochemical characteristics, hydrochemical processes and recharge sources of the geothermal systems in Lanzhou City, northwestern China. Urban Clim. 2022;43(126): 101152.
10. El Mourabit Y, Assabbane A, Hamdani M. Study of correlations between microbiological and physicochemical parameters of drinking water quality in El kolea city (Agadir, Morocco): using multivariate statistical methods. J Mater Environ Sci. 2020;11(2):310–7.

11. Gbagbo TAC, Kpaibe SA, Gbokpeya KM, Able N, Seki TO, Bakayoko A, Meless DFR, Amin N. Caractérisation physicochimique et bactériologique des eaux de consommation de la nappe phréatique du village M'pody (Côte d'Ivoire). J Rech Sci Univ. 2020;22(3):771–88.
12. Gbagbo TA, Kpaibe SPA, Meless DFR, Seki TO, Bakayoko A, Agbessi KT, Amin NC. Spatio-temporal evolution of the quality of drinking water in m'pody, a village in the District of Anyama (Ivory Coast). Int J Environ Chem. 2022;6(2):42–50.
13. Gobinder S, Jagdeep S, Owais AW, Johnbosco CE, Johnson CA. Assessment of groundwater suitability for sustainable irrigation : a comprehensive study using indexical, statistical, and machine learning approaches. Groundwater Sustain Develop. 2024;24: 101059.
14. Jourda JP. Contribution à l'étude Géologique et Hydrogéologique du Grand Abidjan (Côte d'Ivoire). Thèse de Doctorat, 3ème cycle, Université Scientifique, Technique et médicale de Grenoble. 1987;319p.
15. Kandana MY, Aya NBK, Droh LG. Assessment of the borehole and well water quality intended for human consumption: multivariate analysis approaches. Int J Innov Appl Stud. 2023;40(4):1299–311.
16. Kumari P, Babu NJ, Singh S, Chandrasekar S, Raj A, Barik SK. Multivariate statistical analysis for water quality variation in Baraila Lake, Bihar, India. Aust Environ Sci. 2023;8(1):1090.
17. Kohonen T. Self-organization and Associative Memory. 3rd ed. Berlin: Springer; 1984.
18. Kohonen T. Self-Organizing Maps, Springer Series in Information Sciences, 30. Berlin: Springer; 1995.
19. Kouamé KJ. Contribution à la Gestion Intégrée des Ressources en Eaux (GIRE) du District d'Abidjan (Sud de la Côte d'Ivoire) : Outils d'aide à la décision pour la prévention et la protection des eaux souterraines contre la pollution. Thèse de Doctorat de l'Université de Cocody. 2007;227p.
20. Lagnika M, Ibikounle M, Montcho JC, Wotto VD, Sakiti NG. Caractéristiques physico-chimiques de l'eau des puits dans la commune de Pobè (Bénin, Afrique de l'ouest). J Appl Biosci. 2014;79:6887–97.
21. Liu CW, Lin KH, Kuo YM. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. Sci Total Environ. 2003;313(1–3):77–89.
22. Metaiche M, Djafer KH, Aichour A, Gaci N. Multivariate statistical analysis of groundwater quality of Hassi R'mel, Algeria. J Ecol Eng. 2023;24(5):22–31.
23. Mohammed MA, Szabó NP, Szűcs P. Multivariate statistical and hydrochemical approaches for evaluation of groundwater quality in north Bahri city-Sudan. Heliyon. 2022;8(11): e11308.
24. Muniz JN, Duarte KG, Braga FHR, Lima NS, Silva DF, Firmo WCA, Batista MRV, Silva FMAM, Miranda RCM, Silva MRC. Limnological Quality : seasonality assessment and potencial for contamination of the Pindaré river watershed, pre-amazon region, Brazil. Water. 2020;12:851.
25. Narmatha T, Jeyaseelan A, Mohan SP, Mohan VR. Integrating multivariate statistical analysis with GIS for groundwater in Pambar SubBasin, Tamil Nadu, India. Int J Geom Geosci. 2011;2(2):392–402.
26. Noori R, Sabahi MS, Karbassi AR, Baghvand A, Zadeh HT. Multivariate statistical analysis of surface water quality based on correlations and variations in the data set. Desalination. 2010;260:129–36.
27. Ponmalai R, Kamath C. Self-Organizing Maps and Their Applications to Data Analysis. U.S; 2019.
28. Raman SBK, Geetha G. Correlation analysis and prediction of characteristic parameters and water quality index of ground water. Poll Res. 2005;24:197–200.
29. RGPH (Recensement Général de la Population et de l'Habitat) Socio-demographic data. Permanent Technical Secretariat of the RGPH Technical Committee. 2021;p68.
30. Rodier J, Legube B, Merlet N. Water analysis: natural water, waste water, sea water : chemistry, physicochemistry, bacteriology, biology. Dunod Paris. 2016;9:78–1368.
31. Samson S, Elangovan K. Multivariate statistical analysis to assess groundwater quality in Namakkal district, Tamil Nadu, India. Indian J Geo Marine Sci. 2017;46(04):830–6.
32. Singh KP, Malik A, Mohan D, Sinha S. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of gomti River (India)-a case study. Water Res. 2004;38(18):3980–92.
33. Sombo BC. Etude de l'évolution structurale et sismo-stratigraphique du bassin sédimentaire off-shore de Côte d'Ivoire, marge passive entaillée d'un canyon. Thèse de Doctorat d'Etat, Université de Cocody, Abidjan, Côte d'Ivoire, no. 2002;355:304p.
34. Sunitha V, Sudarshan V, Reddy BR. Hydrogeochemistry of groundwater, Gooty area, Anantapur district Andhra Pradesh, India. Poll Res. 2005;24:217–24.
35. WHO. Guidelines for drinking-water quality : 4th edition incorporating the first additive. Geneva: WHO Document Production Services; 2017. p. 564.
36. Wu J, Li P, Wangan D, Rena X, Wei M. Statistical and multivariate statistical techniques to trace the sources and affecting factors of groundwater pollution in a rapidly growing city on the Chinese Loess Plateau. Human Ecol Risk Assess. 2020;26(6):1603–21.
37. Zango MS, Pelig-Ba KB, Anim-Gyampo M, Gibrilla A, Sunkari ED. Hydrogeochemical and isotopic controls on the source of fluoride in groundwater within the Vea catchment, northeastern Ghana. Groundwater Sustain Develop. 2021;12: 100526.
38. Zango MS, Pelig-Ba K-B-, Anim-Gyampo M, Gibralla A, Abu M. Assessment of the mineralogy of granitoids and associated granitic gneisses responsible for groundwater fluoride mobilization in the Vea catchment, Upper East Region Ghana. Sustain Water Sustain. 2022;8:4.