EMCEI 2017

CrossMark

Assessment of surface water quality using multivariate statistical analysis techniques: a case study from Ghrib dam, Algeria

Somia Hamil [1,2] · Siham Arab [1,3] · Amin Chaffai [1] · Mounia Baha [2] · Abdeslem Arab [1]

Abstract

This study explores the water quality status and pollution sources in Ghrib Dam, Algeria. It allows us to obtain more accurate information on water quality by applying a series of multivariate statistical techniques, including principal component analysis (PCA)/factor analysis (FA), hierarchical cluster analysis (CA), and multiple regression analysis (MRA). On 19 physicochemical parameters dataset over 5 years and from 6 different sites located in and around the lake. One-way analysis of variance (ANOVA) was used to investigate the statistically considerable spatial and seasonal differences. The results of ANOVA suggest that there exist a statistically significant temporal variation in the water quality of the dam for all parameters. On the other hand, only organic matter has a statistically significant spatial variation. In the multiple linear models, an association between organic and inorganic parameters was found; their origin comes from the mechanical erosion process of agricultural lands in the watershed. The PCA/FA identifies five dominant factors as responsible of the data structure, explaining more than 94.96% of the total variance in the water quality dataset. This suggests that the variations in water compounds' concentration are mainly related to the multiple anthropogenic activities, as well as natural processes. The results of cluster analysis demonstrate that the sampling stations were divided in two similar groups, which indicates spatial homogeneity. While seasonal grouping has showed that the source of pollution was related to the level of runoff in the seasons.

Introduction

Worldwide water quality (lakes and rivers) has deteriorated in the last years due to uncontrolled industrial discharges, the intensive use of chemical fertilizers in agriculture, and the disorganized exploitation of water resources. Surface water is highly prone to point and non-point source pollutions

This article is part of the Topical Collection on *Water resources and water management for environmental integration in the Euro-Mediterranean region*

✉ Somia Hamil
dj.soumia@hotmail.fr; hamil@ens-kouba.dz

1   Laboratory of Dynamic and Biodiversity (LaDyBio), FSB, USTHB, L.P. 32 El-Alia, Algiers, Bab Ezzouar, Algeria

2   Laboratory of Eco-Biology Animals (L.E.B.A.), École Normale Supérieure de Kouba Bachir El Ibrahimi, B.P. 92, 6050 Algiers, Algeria

3   FST/GAT, USTHB, L.P. 32 El-Alia, Bab-Ezzouar, Algiers, Algeria

(Singh et al. 2005), due to its easy accessibility for disposal of wastewaters and difficulty with its protection. The natural processes such as rainfall, erosion, and sediment loading, as well as anthropogenic processes contribute to the loss of water resources and define the water quality of the region (Khatri and Tyagi 2015; Singh et al. 2005; Kazi et al. 2009). This has generated great pressure on aquatic ecosystems, resulting in a decrease of biodiversity and loss of critical habitats. Therefore, it is essential to prevent and control water pollution and to implement regular survey programs. Also, the development of public awareness must be considered in order to get positive results in the long term for this kind of planning and nature conservation (Cetin 2016). Usually monitoring and evaluating the water quality are required for integrated management of these water resources. However, monitoring programs that provide representative and reliable information of the data are not easily implemented due to both spatial and temporal variations in water quality (Dixon and Chiswell 1996; Badillo-Camacho et al. 2015). In the scientific literature, different multivariate statistical techniques were applied: cluster analysis (CA), discriminant analysis (DA), principal

🪼 Springer

component analysis (PCA)/factor analysis (FA), for interpretation of the complex databases then to assess temporal and spatial variations in lake water quality and finally to identify the potential sources of water contamination.

In Algeria, as in most developing or industrialized countries, the deterioration of the surface water quality is increasing concern. Several studies have been carried out to determine the problems of deterioration in water quality of dams in Algeria, such as the research works of Bouguerne et al. (2016) and Mebarkia et al. (2017) on the water quality of the Ain Zada dam in the upper reaches of Wadi Boussellem, Bahroun and Chaib (2017) in the reservoir Mexa reservoir, and Guerraiche et al. (2016) in the Grouz dam (eastern Algeria). In western Algeria, the monitoring of Touhari et al. (2015) on four high-Cheliff dams and Djelita et al. (2016) on the spatiotemporal patterns of Boughrara dam water quality.

Ghrib Dam is subjected to an anthropic action; for this purpose, it is essential to highlight it as it supplies with water a large region of the north in Algeria. The water quality of the Ghrib Dam has deteriorated since 2005 due to the annual wastewater discharges that come from urban (2366 hm^3) and industrial areas (0.035 hm^3) (Harkat et al. 2011). These discharges contaminate the dam water mainly through the Cheliff River input.

The objective of the present study is to analyze and interpret database set obtained during 5 years (2012–2017) in Ghrib Dam. To achieve this objective, matrix correlation, PCA, and CA multivariate techniques were applied to (Badillo-Camacho et al. 2015) find similarities and dissimilarities among the different sampling sites, (Bahroun and Chaib 2017) evaluate the contribution of water quality parameters to temporal variations in surface water quality, and (Barakat et al. 2016) ascertain the influence of the pollution sources on the water quality variables.

Material and methods

Study area

Ghrib Dam is in the northwestern part of Algeria approximately 30 km from Medea City and 150 km from the Algerian capital Algiers (Fig. 1). The dam is at an altitude of approximately 435 m and is between 36°12′-36°16′N and 2°55′–2°60′ E. This dam was implanted in the wadi Cheliff, at about 20 km southwest of the Ain-Defla city, with an initial capacity of 280 hm^3. It has been put into service in 1939. Ghrib Dam has 14 km^2 of surface area, a drainage basin area of 1378.63 km^2, and a maximum depth of 40 m.

With regard to the climate, the region is characterized by sub-humid climate, relatively temperate winter with hot and dry summer. The annual average temperature is 18.06 °C, with an average annual precipitation around 619.27 mm.

The dammed water is used for the irrigation (the Mitidja Plain in the east and the upper and lower Cheliff Plains in the west) and drinking water purposes. Agricultural activities are present around surface waters in the western part of the dam.

Sampling

The water sampling has been carried out monthly for a period of 5 years (from 2012 to 2017) at the surface layers; only one sampling per month has been taken into account for the temporal monitoring of water quality. Six sampling points were selected for spatial quality monitoring for a short 2-year period (June 2013 to May 2015) (Fig. 1). The water samples were preserved in pre-rinse 1 L polypropylene bottle, at 4 °C in darkness and analyzed within 24 h. Temperature, pH, salinity, electrical conductivity, and dissolved oxygen were measured in situ using a multiparameter analyzer (the Multi 340i/SET WTW). The analysis methods are those advocated by Rodier (2004) and done in the dynamic and biodiversity laboratory; nitrate, nitrite, and phosphate were measured by molecular absorption spectrometry, while the ammonium is done by the spectrophotometric method, and the levels of calcium, magnesium, and chlorides were determined by the volumetric titration. Other parameters were carried out by the technical staff of the laboratory of the national agency of water resources (NAWR); BOD and COD were analyzed using an automated analysis SP2000, sodium and potassium were analyzed by flame spectrophotometry (Corning 410), and bicarbonates were analyzed by a pH meter Type E632 (METLOOHM). The obtained data are compared to the Algerian quality standards (Table 1) adopted by the national agency of water resources (NAWR 2009).

The monthly water inputs and the total monthly volume of water in the dam in the study period (2012–2017) was provided by the National Agency for Dams and Transfers (2017).

Statistical methods

Data for physicochemical parameters of water samples were presented as mean values and plotted using time series, to follow the temporal variation pattern of water quality variables in Ghrib Dam. With the objective of evaluating significant differences between spatial and temporal for all physicochemical variables, data was analyzed using one-way analysis of variance (ANOVA) at 0.05% level of significance. In all statistical analysis, the water quality parameters were grouped into four seasons: summer (from June to August), autumn (from September to November), winter (from December to February), and spring (from Marche to May).

The Pearson parametric correlation coefficient was obtained to describe the degree of association between two variables under study (Peat et al. 2009) and gives preliminary
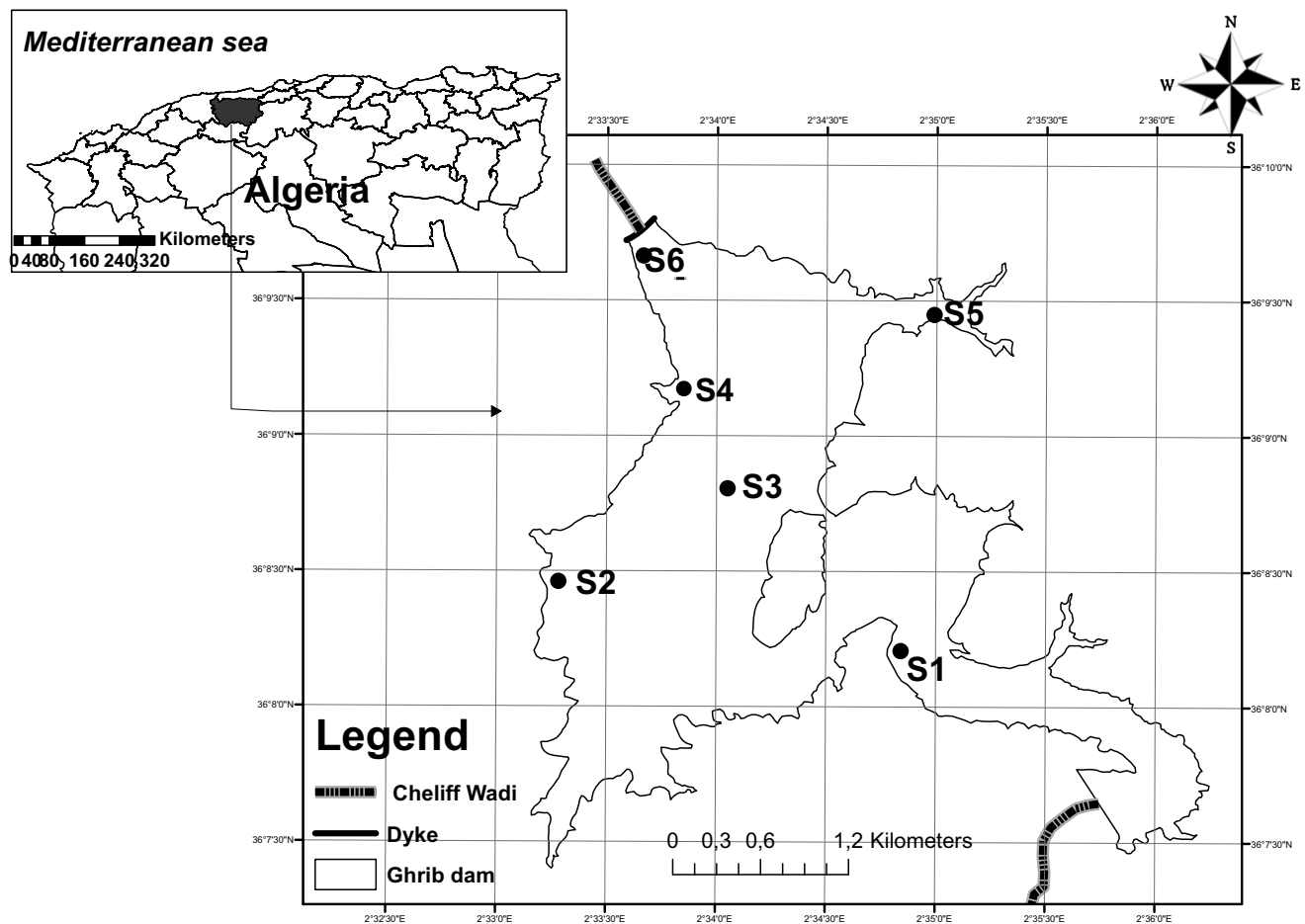
Fig. 1 Location of the study area and sampling stations within the Ghrib Dam reservoir

information on relationship models in the regression analysis. In this study, temporal variation of dam water quality parameters were evaluated through a season parameter correlation matrix, using Pearson parametric correlation coefficient (Pearson's R).

Multiple linear regressions is a multivariate statistical technique for examining the linear correlations between two or more independent variables and a single dependent variable (Charulatha et al. 2017). The variation of this dependent variable is accounted for by the variation in those independent variables of physical, chemical, or biological water characteristics. The general equation is: $Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m + \varepsilon$ where Y represents the dependent variable, $X_{1 \ldots} X_m$ represent the several independent variables, $\beta_o \ldots \beta_m$ represent the regression coefficient, and ε represents the random error.

The multivariate analysis of the data using PCA and CA enabled the identification of the sources of constituents and the distinguishing of the natural and anthropogenic contributions of pollutants into the lake system based on the level of association of the variables (Ndungu et al. 2015). All mathematical and statistical computations were made using the R program (R Development Core Team 2014).

Cluster analysis was used to explore the similarities between water samples (Otto 1998) and grouping each similar site or periods in the cluster with respect to a predetermined selection criterion. In the present analysis, hierarchical agglomerative CA was performed on the normalized dataset implementing the Ward's linkage model, employing Bray–Curtis distances regarded as a similarity measurement. Ward's method says that the distance between two clusters, A and B, is how much the sum of squares will increase when we merge them. In particular, CA allows the sorting of variables with the same descriptions into the same cluster; it is often illustrated in dendrogram form.

Factor analysis was employed to establish the possible relationships between the physicochemical variables. In this study, we used principal component analysis (PCA) of factor analysis. The main purpose of FA is to reduce the contribution of less significant variables to simplify even more of the data structure coming from PCA (Shrestha and Kazama 2007). The new set of the obtained variables, the principal components (PCs) are arranged in decreasing order of importance. The PCA is a data reduction technique and suggests how many varieties are important to explain the observed variance in the

Table 1   Algerian surface water guidelines, (NAWR 2009)

Quality classification	Good	Medium	Poor	Very poor
DO (mg/l)	< 7	7–5	5–3	> 3
NH_4^+ (mg/l)	≤ 0.01	0.01–0.1	0.1–3	> 3
NO_3^+ (mg/l)	≤ 10	10–20	20–40	> 40
PO_4^{-3} (mg/l)	≤ 0.01	0.01–0.1	0.1–3	> 3
BOD	< 5	5–10	10–15	> 15
COD	< 20	20–40	40–50	> 50
OM (mg/l)	< 5	5–10	10–15	> 15
Ca^{+2} (mg/l)	40–100	100–200	200–300	> 300
Mg^{+2} (mg/l)	< 30	30–100	100–150	> 150
Cl^- (mg/l)	10–150	150–300	300–500	> 500
Na^+ (mg/l)	10–100	100–200	200–500	> 500
TS (mg/l)	300–1000	1000–1200	1200–1600	> 1600

data (Bhat et al. 2014), with a minimum loss of original information.

Results and discussion

Water quality evaluation

The temporal variation of water inputs and the total monthly volume of water in the dam and 19 physicochemical parameters in the Ghrib Dam during the period of 5 years (June 2012 to May 2017) are presented as time series in Figs. 2 and 3 to assess the quality of this aquatic system. The mean values of physicochemical parameters at different sampling sites in Ghrib Dam during the period of 24 months (June 2013–May 2015) are presented in Table 2. To explain the nature and extent of the relationships between various physicochemical parameters, we used the Pearson correlation coefficient (*r*), which gives an idea of the possible relationships between

biophysical variables (Table 3, $p < 0.05$). All the studied parameters showed a significant temporal variation (ANOVA one-way, *p* value < 0.01), but their spatial distribution in the six study stations were homogeneous (ANOVA one-way, *p* value > 0.05) except organic matter had a spatial and temporal heterogeneity.

The Ghrib Dam receives an average annual water of 7.40 Hm³/year. Usually, the dam receives a significant amount of water in winters and springs (Fig. 2), what makes the water level in the dam increases. Water temperature showed an annual characteristic cycle reflecting the atmospheric temperature and presents the most significant difference among seasons, with higher values during the summer (20–28 °C) and lower values in the winter season (9–13.5 °C). The water temperature causes significant environmental impacts, because most of the physical, chemical, and biological characteristics of aquatic systems are directly affected by temperature. The pH values of collected water samples ranged from 6.36 to 8.64, with the highest values in summer-autumn and the lowest in winter at most of the study sites, and they are generally within the recommended range 6–9 of Algerian standards. Dissolved oxygen is an essential factor for aquatic life. It is commonly used to assess the water resource quality (Sánchez et al. 2007; Barakat et al. 2014). Mean concentrations of DO in the study are generally level close to 7 mg L^{-1}, but the distribution of DO showed marked temporal variability and its lower value in summer and autumn (3.5–7.5 mg L^{-1}) and the higher (7–13.5 mg L^{-1}) in winter season and spring, qualifying the water quality as moderate to pollute in summer and autumn and good quality in winter and spring. On all studied stations, the concentrations of dissolved oxygen are homogeneous (Table 1). In correlation matrix, it was found that DO was negatively correlated to WT. This inverse relationship is a natural process because warmer water becomes more easily saturated with oxygen, and it can hold less dissolved oxygen (Shrestha and Kazama 2007). DO in our study present also negative correlation with BOD and positive

Fig. 2 Temporal variation of WI; water inputs (Hm³), VW; water volume (Hm³) in the Ghrib Dam from June 2012 to May 2017
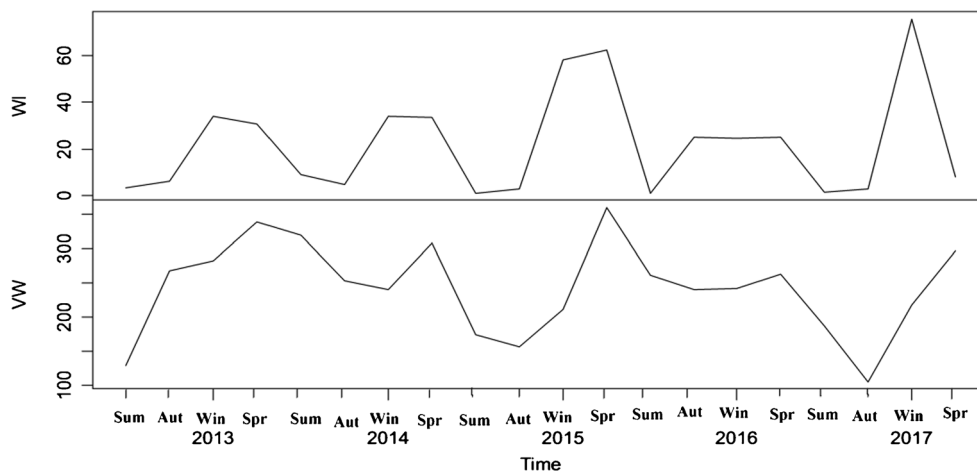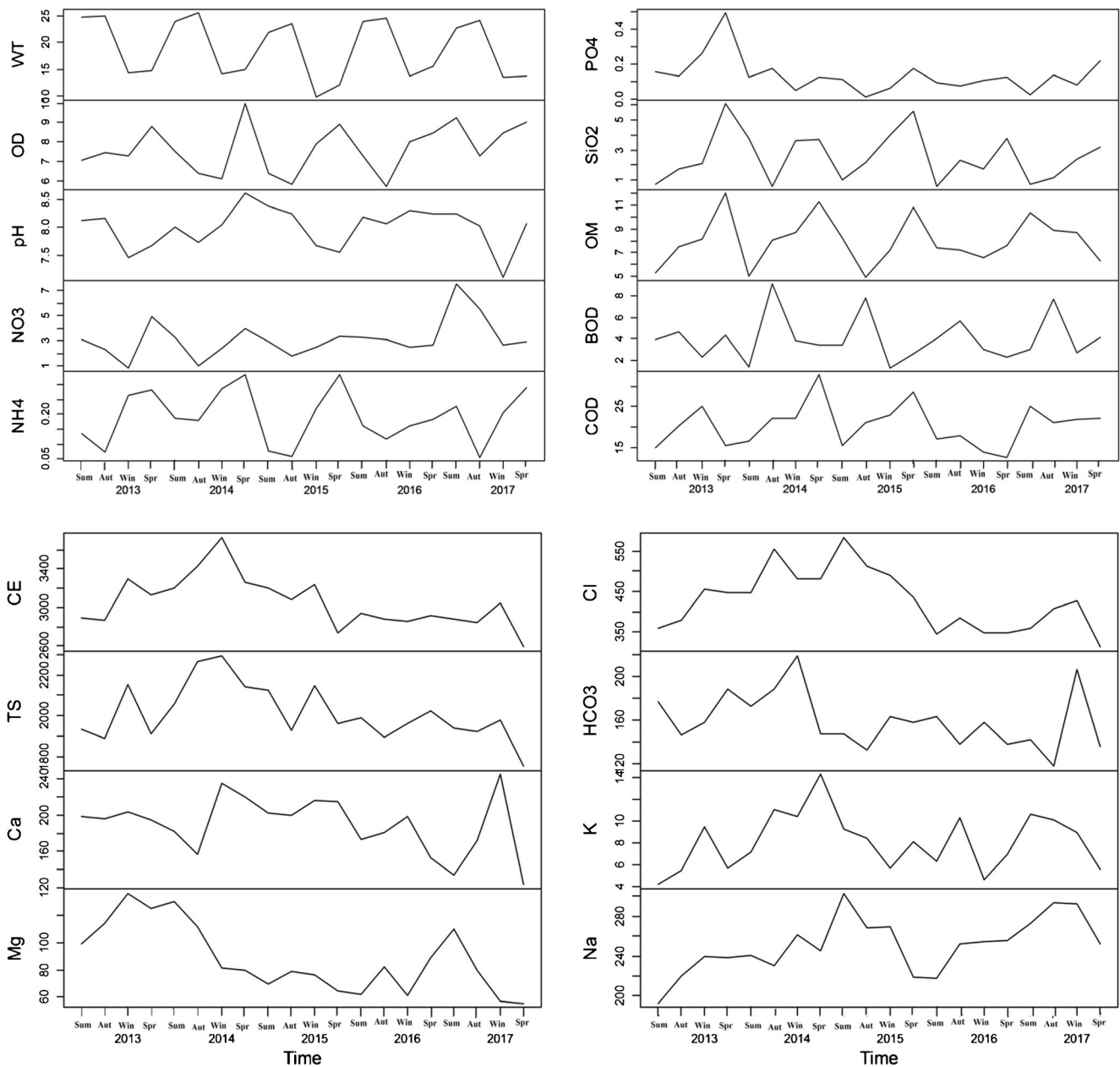
Fig. 3 Temporal variation of variables water quality in Ghrib Dam; WT water temperature (°C), pH, OD dissolved oxygen (mg l^{-1}), BOD (mg l^{-1}), COD (mg l^{-1}), OM (mg l^{-1}) organic matter (mg l^{-1}), NO$_3^-$ (mg l^{-1}), NH$_4^+$ (mg l^{-1}), PO$_4^-$ (mg l^{-1}), SiO$_2^-$ (mg l^{-1}), TS total solid (mg l^{-1}), CE electrical conductivity (μS cm^{-1}), Ca^{+2} (mg l^{-1}), Mg^{2+} (mg l^{-1}), Cl$^-$ (mg l^{-1}), K$^+$ (mg l^{-1}), Na$^+$ (mg l^{-1}), HCO$_3^-$ (mg l^{-1})

correlation with NH$_4^+$, COD, and water volume in Ghrib Dam (Table 3).

BOD is a measure of the dissolved oxygen consumed by microorganisms during the oxidation of reduced substances in waters and wastes (Bhateria and Jain 2016). BOD in various samplings in Ghrib dam shows values between 0.89 and 12.7 mg L^{-1}. These values indicated that the water samples ranked as good to polluted quality. COD is defined as the amount of oxygen required for chemical oxidation with the sample under controlled conditions. In this study, COD of the Ghrib Dam water is ranged between 4.6 and 42 mg L^{-1}, which ranks Ghrib waters as polluted to good quality. Table 2, the

water of Station 1 (S1) presents high BOD and COD levels. High level of COD is due to increased human activities such as agriculture around the lake, transport of natural domestic sewage, and industrial pollutant; those factors have a negative impact on the water quality. COD showed positive correlation with NH$_4^+$, i.e., a decrease of one results in the decrease of the other one. In our result, there is no correlation between the BOD and the COD indicates that the major part of organic material is not biodegradable.

Organic matter in aquatic systems is a complex mixture of molecules (Hawkes et al. 2018) with different sources, waste-water and industrial particles, soil organic matter, living

Table 2 Spatial means and standard error of physicochemical parameters in Ghrib Dam (June 2013 to May 2015; $n = 24$) (mean ± standard error)

Stations	S1	S2	S3	S4	S5	S6
OD (mg/l)	7.73 ± 1.57	7.45 ± 1.55	8.51 ± 2.27	7.50 ± 1.87	7.36 ± 1.73	7.29 ± 1.71
NO_3^+ (mg/l)	3.54 ± 1.22	2.74 ± 0.76	2.35 ± 0.66	2.75 ± 0.45	2.47 ± 0.95	2.53 ± 0.58
PO_4^{-3} (mg/l)	0.097 ± 0.07	0.055 ± 0.02	0.050 ± 0.02	0.055 ± 0.07	0.042 ± 0.02	0.036 ± 0.02
NH_4^+ (mg/l)	0.67 ± 0.42	0.52 ± 0.4	0.35 ± 0.19	0.33 ± 0.26	0.21 ± 0.14	0.64 ± 0.56
OM (mg/l)	8.57 ± 3.34	4.12 ± 2.05	4.93 ± 1.11	5.04 ± 2.09	4.37 ± 0.14	7.28 ± 0.56
BOD	5.03 ± 3.25	3.45 ± 2.48	2.99 ± 1.7	3.42 ± 2.2	2.92 ± 2.04	3.38 ± 2.42
COD	27.73 ± 8.3	21.55 ± 7.7	24.17 ± 8.2	22.83 ± 7.1	24.82 ± 7.9	24.59 ± 7.8
EC ($\mu S\ cm^{-1}$)	3001.5 ± 308.8	3022.3 ± 268.8	3035.2 ± 288.4	3058 ± 289.9	3028.9 ± 279.2	3033.4 ± 283.6
TS (mg/l)	2057.1 ± 268.2	1959.5 ± 167.1	2005.6 ± 189.9	1978.5 ± 195.1	1965.8 ± 198.5	1993.1 ± 187.2

phytoplankton, and other plant matter. In Ghrib dam, mean concentrations of organic matter are generally level close to 8.01 mg L^{-1}, but they remain significant with values of 0.5–16.7 mg L^{-1} in temporal data. The water collected in the level of the stations (S1) and (S6) is the richest in organic matter (8.75, 728 mg/l respectively). The temporal variation of organic matter is positively correlated with water inputs (Table 3), so that the origin of the organic matter in the dam is mainly related to inputs entering into the lake. Assessing the level of nutrients is a main feature in determining lake productivity. It is a food source for aquatic organisms. Phosphorus is usually the most limiting nutrient for algae growth in lakes, since it is the nutrient in short supply in most fresh waters; even a modest increase in phosphorus can, under the right conditions, set off a whole chain of undesirable events in aquatic system (Bhateria and Jain 2016). The PO_4^- shows concentrations of from 0.001 to 0.68 mg L^{-1}, with a high concentration noted in spring season, including some peaks of concentration at the end of the summer. This concentration ranks Ghrib waters as excellent to polluted quality. Both nitrate (NO_3^-) and ammonia (NH_4^+) concentrations are highly variable during lake seasonal cycles. The measured NH_4^+ values vary between 0.006 and 0.51 mg L^{-1}, while the average concentrations of NO_3^- in this study ranged from 0.163 to 10.01 mg L^{-1}. According to Algerian standards, the water quality of Ghrib is good to polluted for both parameters. In the Ghrib Dam, nutrient concentrations show considerable values in S1 (Table 2). All three nutrients were positively correlated with organic matter (Table 3), and both NH_4^+ and PO_4^- were positively correlated with the volume of water in Ghrib Dam, but only the NH_4^+ is related to water inputs. Generally, it comes from the biodegradation of waste and inputs from domestic, agricultural, and industrial origin (Barakat et al. 2014). Its further degradation to nitrites and nitrates consumes dissolved oxygen. Nitrate and phosphorus in aquatic systems have a natural source, mainly coming from the decomposition of organic matter, land drainage, plant and animal debris, and the weathering of rocks from the catchment area (phosphorus-bearing rocks and igneous rocks). Domestic

wastewaters, industrial effluents, and use of manufactured fertilizers containing ammonium nitrate and potassium nitrate contribute to elevate its levels in surface waters. Increasing concentration of PO_4^- and NO_3^- in lakes has resulted in enhanced productivity elevating oxygen demand (Pandit and Yousuf 2002). Comparing our results with that of Touhari et al. (2015) in the period 1999 to 2008, the Ghrib dam presented a poor quality, which is associated with high nutrient levels. This latter is directly related to the inflow of water during the rainy season, which may be due to leaching of fertilizers used in agriculture soils situated in watersheds.

During the study period. SiO_2^- ranged between 0.1 and 7.92 mg L^{-1} with a mean value of 2.55 mg L^{-1}. The maximum concentrations were observed in spring and winter, also, in the transition period between autumn-winter. This concentration was significantly related to the volume of water and water inputs (Table 3); chemically, the rock erosion in catchment area solubilizes silicon, which is then transported by river runoff into the lake dam. Total solids is a measure of all the suspended, colloidal, and dissolved solids in a sample of water. TS can also come from various types of runoff, including dissolution and suspension minerals and organic matter from soils and geological formations, living aquatic microorganisms, and their decaying remains (Boyd 2015). In this study, TS ranged from minimum of 1356 (polluted) and to a maximum of 2608 (highly polluted) in Ghrib Dam. Agricultural runoff entering into the lake contains fertilizers, and suspended soil particles may be the main cause of increasing in TS.

Electrical conductivity in water is due to ionization of dissolved inorganic solids. It is used as a basic index to select the suitability of water for agricultural purposes. In this study, EC varied between 2020 and 3900 µS cm^{-1}; according to Rodier (2004), the waters of the Ghrib dam are strongly mineralized. Spatially, the electrical conductivity is homogeneous in the water of Ghrib (Table 2).

Seasonally, EC was lower in spring and summer than in autumn and winter seasons. The higher EC is attributed to the high degree of anthropogenic activities such as waste disposal and agricultural runoff (Bhat et al. 2014). EC showed positive

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

Variables	WT	OD	pH	NO_3^-	NH_4^+	PO_4^-	OM	BOD	COD	Ca^{2+}
WT	1									
OD	−0.449***	1								
pH	0.252	0.021	1							
NO_3^-	0.190	0.162	0.123	1						
NH_4^+	−0.569***	0.468***	−0.117	0.030	1					
PO_4^-	−0.100	0.247	−0.173	0.269*	0.262*	1				
OM	−0.214	0.218	−0.123	0.470***	0.297*	0.334**	1			
BOD	0.447***	−0.361**	0.096	0.141	−0.148	0.105	0.064	1		
COD	−0.198	0.444***	0.096	−0.028	0.380**	−0.039	0.052	−0.100	1	
Ca^{2+}	−0.387**	−0.129	−0.189	−0.232	0.174	−0.074	0.207	−0.021	−0.024	1
Cl^-	−0.001	−0.323*	−0.145	−0.244	−0.058	−0.079	0.132	0.069	0.029	0.393**
CE	−0.103	−0.208	−0.094	−0.248	0.156	−0.027	0.104	0.039	−0.006	0.438***
HCO_3^-	−0.206	0.087	−0.193	−0.108	0.282*	0.193	0.033	−0.110	0.168	0.335**
K^+	0.055	0.075	0.011	0.116	0.270*	−0.019	0.170	0.294*	0.179	0.131
Mg^{2+}	0.219	−0.024	−0.098	0.003	0.008	0.324*	0.055	0.035	−0.037	−0.110
Na^+	0.045	−0.124	0.014	0.154	−0.170	−0.097	0.013	0.151	−0.060	0.069
SiO_2^-	−0.498***	0.240	−0.200	0.067	0.200	0.247	0.343**	−0.097	−0.105	0.289*
TS	−0.108	−0.187	−0.125	−0.157	0.211	−0.008	0.164	0.118	−0.165	0.407**
VW	−0.149	0.263*	0.019	0.084	0.341**	0.347**	0.243	−0.123	0.041	−0.171
AFF	−0.557***	0.377**	−0.251	0.046	0.337**	0.062	0.379**	−0.205	0.122	0.366**

Variables	Cl^-	CE	HCO_3^-	K^+	Mg^{2+}	Na^+	SiO_2^-	TS	VW	AFF
WT										
OD										
pH										
NO_3^-										
NH_4^+										
PO_4^-										
OM										
BOD										
COD										
Ca^{2+}										
Cl^-	1									
CE	0.642***	1								
HCO_3^-	0.015	0.274*	1							
K^+	0.196	0.281*	−0.035	1						
Mg^{2+}	0.109	0.359**	0.131	0.129	1					
Na^+	0.214	0.254	0.078	0.151	−0.067	1				
SiO_2^-	0.038	0.100	0.137	−0.081	0.089	0.006	1			
TS	0.465***	0.767***	0.111	0.286*	0.215	0.143	0.022	1		
VW	−0.109	−0.055	0.034	−0.059	0.112	−0.336**	0.398***	0.023	1	
AFF	−0.033	−0.005	0.076	−0.041	−0.186	−0.193	0.590***	0.053	0.388**	1

*Correlation is significant at p < 0.05 level (2-tailed),

**Correlation is significant at p < 0.01 level (2-tailed),

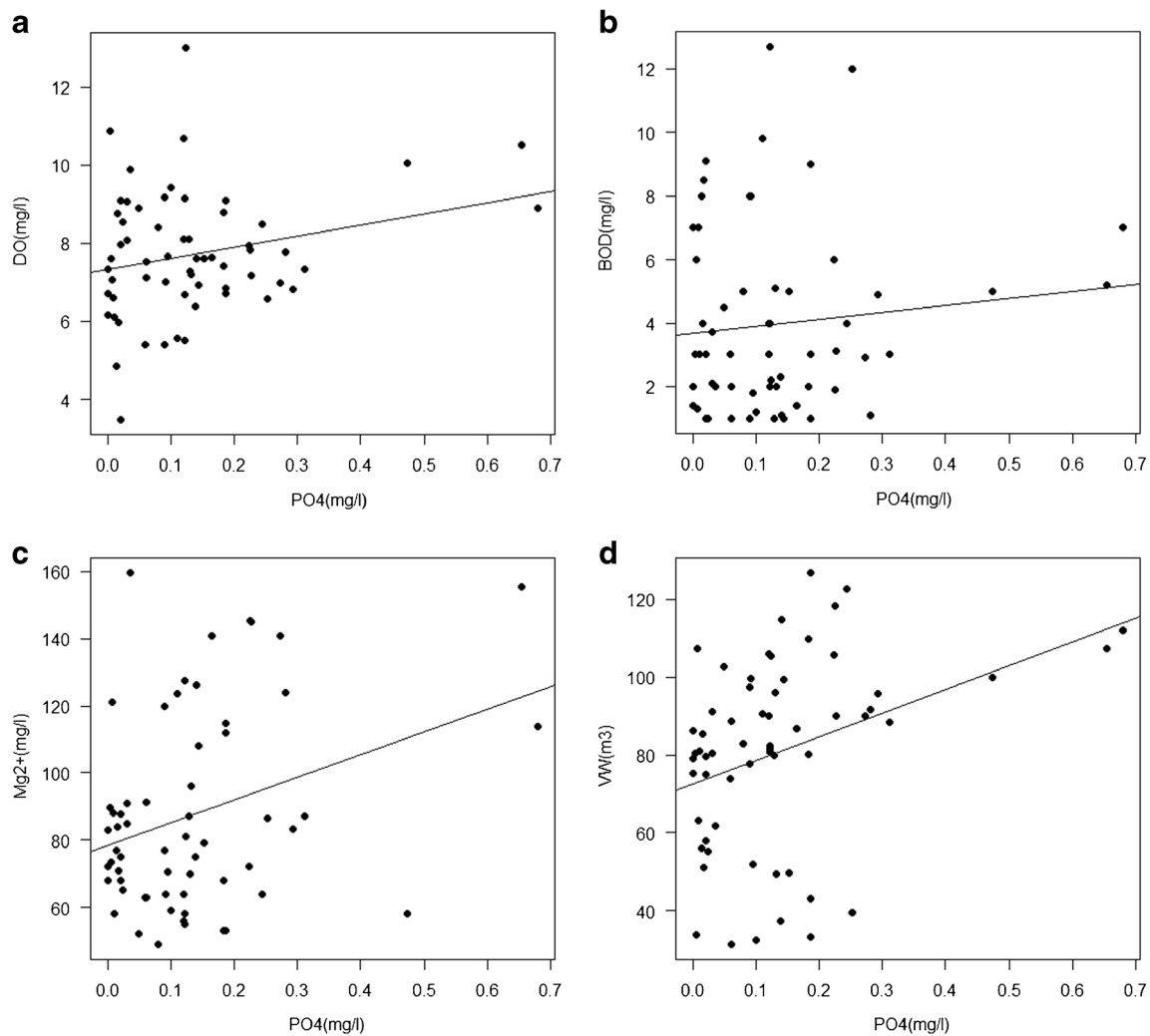***Correlation is significant at p < 0.001 level (2-tailed)

Fig. 4 Plots of water quality parameters as a multiple linear regression model (**a–d**); $PO_4^+ = 0.448 + 1.745\ DO + 2.269\ BOD + 2.119\ Mg^{2+} + 1.882\ VW$

and significant correlated with ions (Ca^{2+}, Mg^{2+}, Cl^-, K^+, HCO_3^-) and TS.

The main ions analyzed were Ca^{2+}, Mg^{2+}, Cl^-, K^+, Na^+, HCO_3^-, and their temporal variations are presented in Fig. 3. Naturally, these elements are very variable in surface waters due to local geology and geographical conditions. Generally, these ions maintained higher concentration in raining periods (winter and spring). Ion concentration varied between Ca^{2+} (90–259 mg/L), Mg^{2+} (49–159 mg/L), Cl^- (237–652 mg/L), K^+ (2.5–26 mg/L), Na^+ (172–370 mg/L), and HCO_3^- (97–335.5 mg/L). According to the Algerian guidelines (NAWR 2009), these measurement results also indicated that the majority of water samples ranked as moderate to highly polluted class in terms of water quality.

Regression analyses

To find out the best predictor of water quality variation in the Ghrib Dam, a stepwise multiple linear regression model was used. We checked the classical assumptions of the linear

regressions, and then we plotted the concentrations of all the dependent variables in relation to the independent variables. The results of the statistical analysis with the multiple linear regression models are shown in Figs. 4, 5, 6, and 7. The figured dependent variables have a significant linear model; the residual variances are equal or constant. The equation of each model was presented with the scatter plot (Figs. 4, 5, 6, and 7). The water quality variation in Ghrib Dam was explained by four predictor variables namely PO_4^-, NO_3^-, NH_4^+, and EC.

The PO_4^+ concentration showed strong significant and positive relationship ($p < 0.0001$) with dissolved oxygen, BOD, Mg^{2+}, and water volume in the dam, The R-square of 0.483 revealed that 48.3% of the variation of water quality is explained by this predictor. The estimated regression equation is $PO_4^+ = 0.448 + 1.745\ DO + 2.269\ BOD + 2.119\ Mg^{2+} + 1.882\ WV$.

The NO_3^- concentration showed strong significant positive relationship with water temperature, organic matter, and Na^+. The R-square of 0.533 revealed 53.3% of the variation of
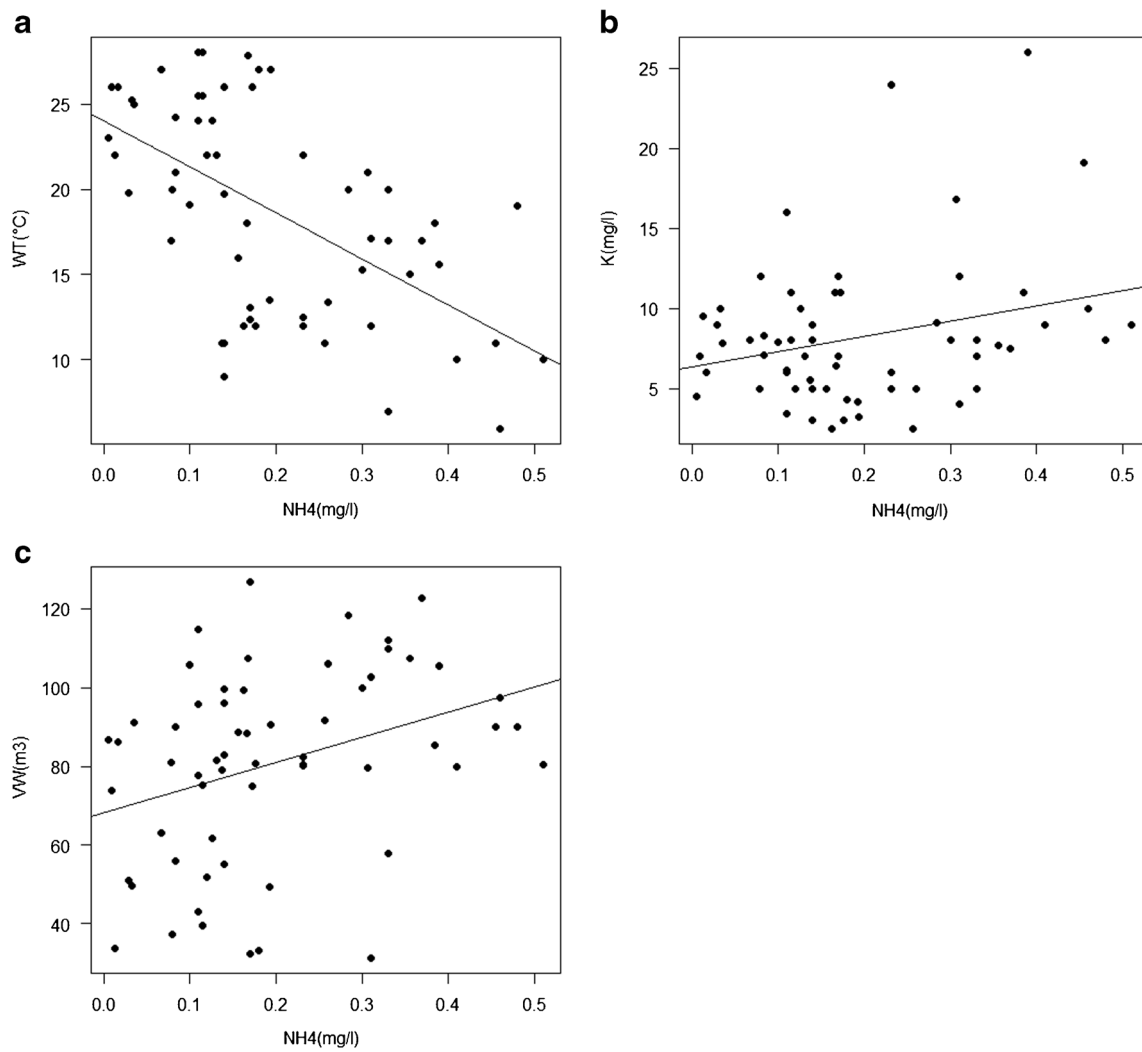
Fig. 5 Plots of water quality parameters as a multiple linear regression model (**a**–**c**); $NO_3^- = -0.542 + 2.173 \, WT + 3.685 \, OM + 1.707 \, Na^+$

water quality. The estimated regression equation is $NO_3^- = -0.542 + 2.173 \, WT + 3.685 \, OM + 1.707 \, Na^+$.

The NH_4^+ concentration showed strong significant and positive relationship with K^+, volume of water, and negative
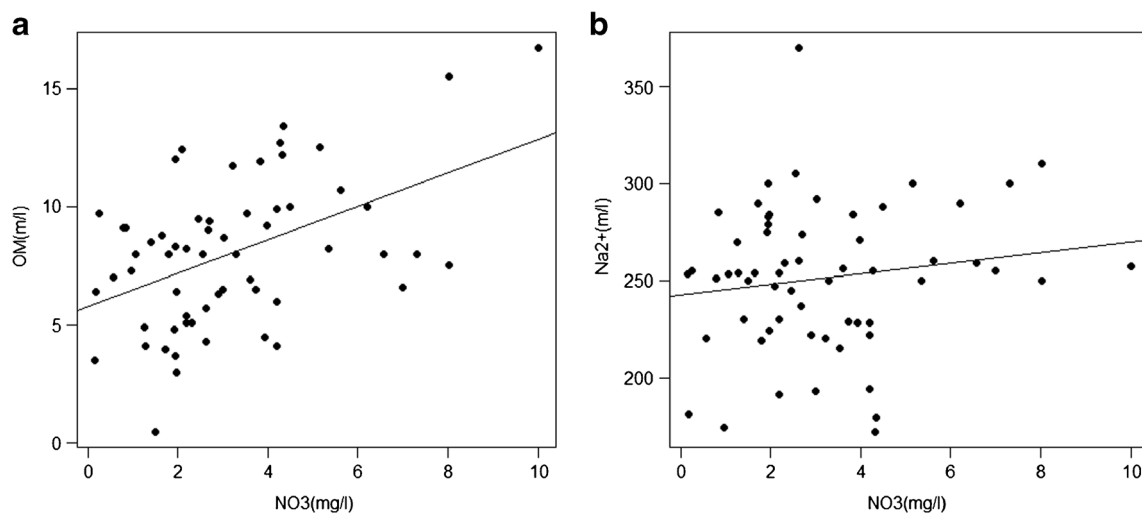

Fig. 6 Plots of water quality parameters as a multiple linear regression model (**a**, **b**); $NH_4^+ = 0.308 - 3.633 \, WT + 1.702 \, K^+ + 1.934 \, VW$
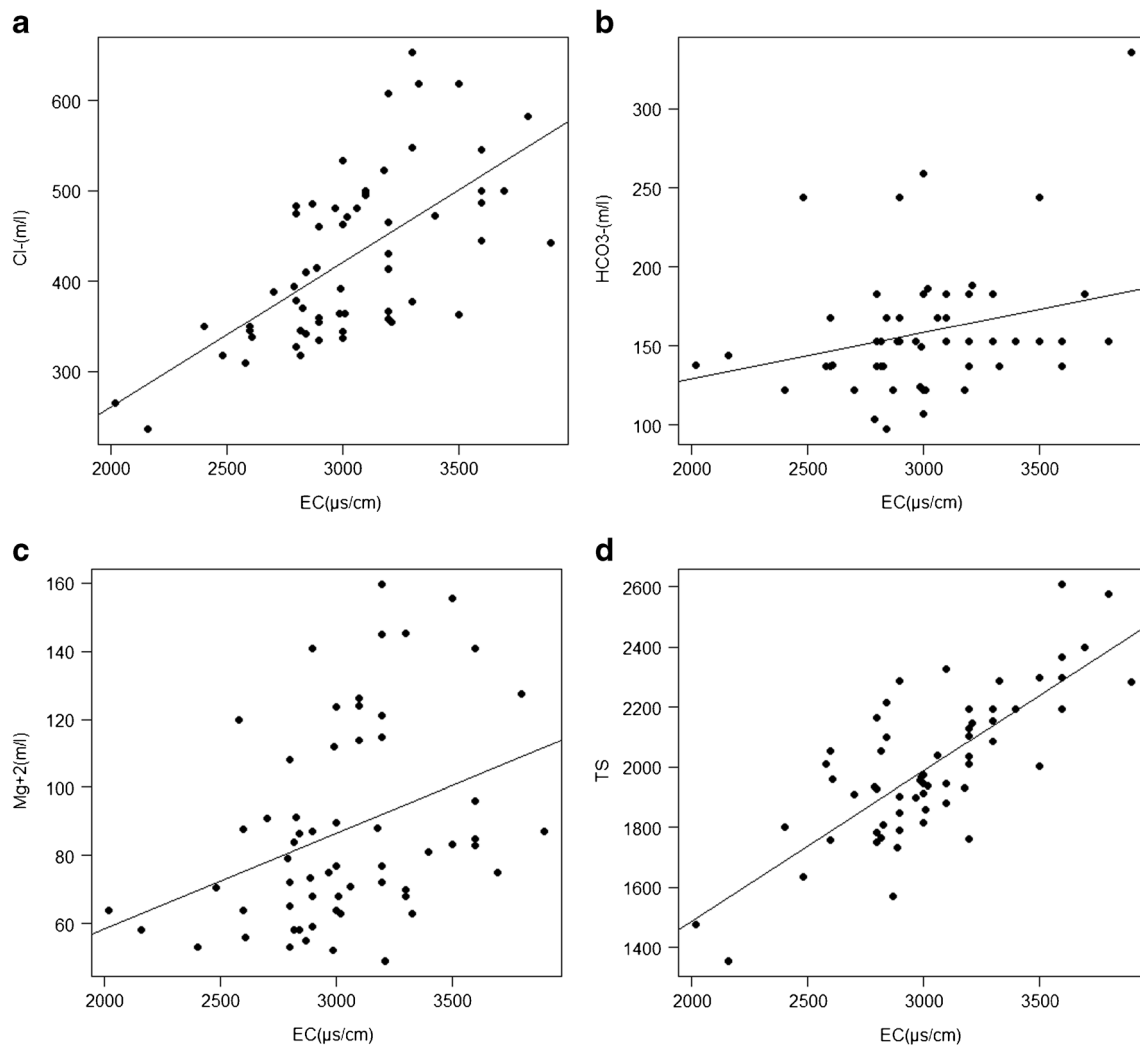
Fig. 7 Plots of water quality parameters as a multiple linear regression model (**a–d**); $EC = -0.582 + 3.097\,Cl^- + 1.824\,HCO_3^- + 2.419\,Mg^{2+} + 5.215\,TS$

with water temperature. The R-square was found equal to 0.65; the estimated regression equation is $NH_4^+ = 0.308 - 3.633\,WT + 1.702\,K^+ + 1.934\,VW$.

EC concentration showed strong significant and positive relationship with Cl^-, HCO_3^-, Mg^{2+}, and TS (R-square = 0.804); the estimated regression equation is $EC = -0.582 + 3.097\,Cl^- + 1.824\,HCO_3^- + 2.419\,Mg^{2+} + 5.215\,TS$. Significant variation of PO_4^+ and NH_4^+ with water volume could be due to the agricultural activity, which is associated with the use of organic and inorganic fertilizers. Generally, the volume of water in the dam increases with the runoff; the latter brings a large load of ions from the catchment area (Bhat et al. 2014). The associations between nutrients and minerals in the multiple linear model revealed that their origin comes from the process of mechanical erosion of agricultural lands in the watershed. Therefore, the presence of agricultural land near bodies of water might include effects on water chemistry, leaching of biocides (Moss 2008), and

increasing suspended loads from soil erosion (Brodie and Mitchell 2005). An excess of nutrients in a water source or eutrophication is often caused by runoff from the land and causes excessive plant growth, or pollution from inorganic or organic matter affecting water stress. This can cause neighboring areas to become severely damaged over time as water use is restricted (Sevik and Cetin 2015).

Multivariate analysis

Cluster analysis (CA)

Cluster analysis was used to detect the similarity groups between the sampling sites and seasons. It yielded a dendrogram, grouping the sampling sites and the seasons on the basis of a percentage of similarity and dissimilarity of water quality parameters. Figure 8 shows the results of the CA and indicates that the sampling sites were similar and clustered into two
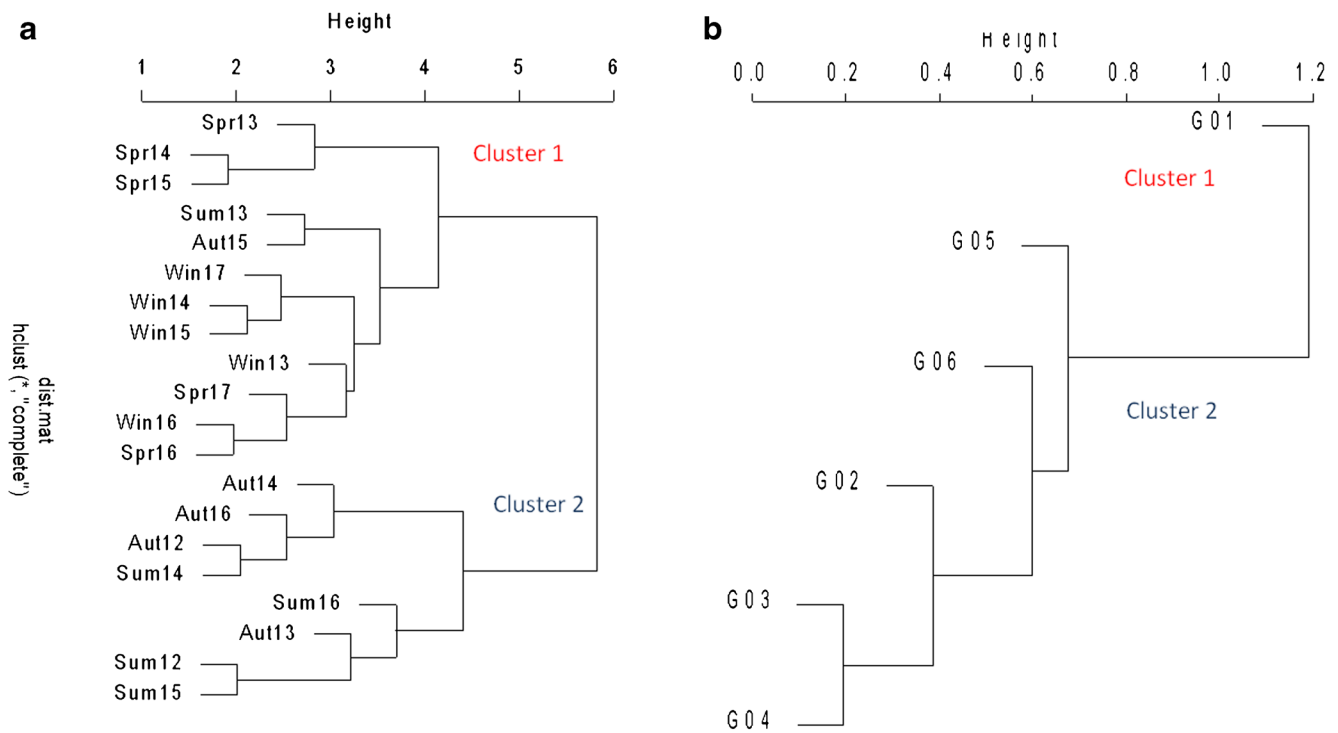
Fig. 8 Dendrogram showing clustering of sampling **a** stations and **b** seasons according to surface water quality characteristics of the Ghrib Dam

distinct regions (Fig.8a) as follows; the first cluster includes the wadi station (s1), and the second cluster includes all the stations of the body of water (s2:s6). The first station is characterized by a low level of water (at max. 3 m), which leads to the accumulation of some water pollutants and increase their concentration, e.g., the organic matter whose quantity is high in this station. This station receives pollution from non-point sources, i.e., mostly from agricultural activities and cereal plantations in the watershed. The grouping of the other stations in a single cluster proves their similarity and confirms the spatial homogeneity of the surface waters in the dam. There is a great similarity between the stations of the west (s2, s3, and s4); these are located near agricultural field. While station 5 is located far from human activities. In spite of that, these stations present a homogeneous distribution of the variables;

maybe this homogeneity is due to a relation with the daily and seasonal movements of water. Seasonal grouping (Fig. 8b) showed that there is a maximum similarity between summers and autumns and between winters and springs. This indicates that the source of pollution in winter and spring is non-point, coming from the catchment area. While the source of pollution in summer and autumn is a point from anthropogenic activities (i.e., agricultural, urban sewage) in the region. The results indicate that the CA technique is useful in offering reliable classification of surface waters in the region and will make it possible to design a future spatial sampling strategy in an optimal manner, which can reduce the number of sampling stations and associated costs.
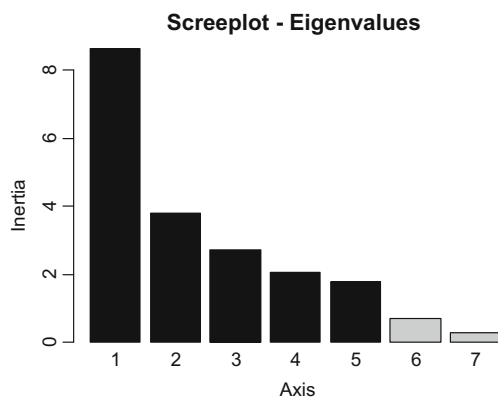
Principal component analysis/factor analysis

Principal component analysis/factor analysis were carried out on the dataset (21 variables) to extract the parameters that are most important in assessing variation in water quality. Eigenvalue selection criteria were selected to explain the sources of variance as one and greater than one. Five principal factors with Eigenvalues > 1 were extracted (Fig. 9) explaining more than 94.96% of the total variance in the water quality dataset (Table 4), where a correlation greater than 0.75 is considered "strong"; 0.75–0.50, "moderate"; and 0.50–0.30, as "weak" significant factor loading (Liu et al. 2003). The parameter loadings for the two identified factors from the factor analysis of the data are given in Table 4.

Fig. 9 Scree plot of the Eigenvalues

Table 4 Factor loadings with variable rotation for the PCs found

	PC1	PC2	PC3	PC4	PC5
WT	0.752	0.444	− 0.220	0.377	0.057
DO	−0.868	−0.024	0.348	−0.250	0.037
pH	0.781	0.285	0.440	0.064	0.161
NO_3^-	−0.544	0.367	−0.048	0.566	0.381
NH_4^+	−0.990	−0.000	−0.070	−0.088	−0.031
PO_4^-	−0.752	0.254	0.136	0.207	0.452
OM	−0.801	0.025	−0.168	0.351	0.442
BOD	0.711	0.310	−0.520	0.074	−0.319
COD	−0.529	−0.310	−0.678	0.159	−0.323
Ca^{2+}	−0.269	−0.698	−0.310	0.336	−0.373
Cl^-	0.303	−0.775	−0.348	0.317	0.169
CE	0.508	−0.790	0.140	0.064	0.102
HCO_3^-	−0.638	−0.233	0.464	0.420	−0.310
K^+	0.338	0.130	−0.756	0.258	0.473
Mg^{2+}	0.332	−0.022	−0.316	−0.728	0.449
Na^+	0.601	−0.636	0.0807	−0.045	0.368
SiO_2^-	−0.726	−0.205	−0.433	−0.467	0.114
TS	−0.107	−0.776	0.418	0.109	0.380
VW	−0.829	−0.309	−0.248	−0.297	−0.044
WI	−0.884	0.447	−0.103	0.028	0.063
Eigenvalue	8632	3806	2717	2051	1,78
Variability %	43,162	19,03	13,584	10,257	8924
Cumulative %	43,162	62,19	75,78	86,03	94,96

In the case of the Ghrib Dam, the F1, which explained 43.16% of the whole dataset, has strong negative loadings on DO, NH_4^+, PO_4^-, OM, water inputs, and volume waters of the Ghrib Dam (Table 4) and moderate loadings on HCO_3^-, NO_3^-; it has strong positive loadings on WT, pH and moderate loadings on BOD. This factor includes the organic variables that are related to the level of water supply in the dam, which may be associated to influences from municipal and industrial point-source discharges, agricultural non-point sources, livestock operations, and/or domestic sources (Simeonov et al. 2003; Barakat et al. 2016). The presence of PO_4^- and NO_3^- in Ghrib's water originate from fertilizer application in the farms; agricultural land use strongly influences stream phosphorus and nitrate nitrogen. This is evident as farmers practice cereal plantation around the area. These elements reflected the degree of eutrophication and organic pollution of the lake (Penn and Bowen 2018). A significant amount of nutrients from fertilizers and farm drainage enter the reservoir and cause excessive algal growth (Liu et al. 2009).

The F2 explained 19.03% of the total variance with strong negative loading on Cl^-, EC, TS and moderate loadings on Na^+ and Ca^{2+}. This factor accounts for the hardness and salinity of the dam's water. The presence of ions (cations and anions) may account for the electrical conductivity and total solids (Gumbo et al. 2016). These research findings suggest origins upstream may be associated mostly with the geological nature of the lands traversed by the Wadi Cheliff waters in the watershed.

The F3 and F4, which have moderate loadings on other mineral salts, and COD, explained respectively the 13.58 and 10.25% variance of the whole dataset (Table 4). These components can be a response to some changes introduced in the system by natural factors or a contamination sources.

All the results of the multivariate statistical studies clearly indicate that the waters of the Ghrib Dam are vulnerable to pollution from agricultural sources in the first degree, which is associated with the use of organic and inorganic fertilizers. The water quality of the Ghrib Dam is directly related to the wadi Cheliff that feeds it, which is primarily influenced by agricultural runoff and sewage discharge. This latter is located next to large lands of cereal cultivation. Principal components analysis associated the water inputs with nutrients (NH4-, NO3-, and PO4) and organic matter, which were the key parameters characterizing the mouth of Ghrib dam, which is under the influence of agricultural activities in the catchment. Cluster analysis resulted in two major regions: (i) southern region (mouth of Ghrib dam, station 1) and(ii) main lake of

Ghrib Dam (stations 2, 3, 4, 5, and 6). The southern region is the most influenced by external discharge, which consists in runoff from the agricultural activities in the catchment and urban waste water. The results of the multi-linear regression model also indicate a linear correlation model between water volume in the dam and nutrients with some ions, indicating that the dam water quality is fully dependent on those of the wadi Cheliff, which feeds it.

Conclusion

In this study, different multivariate statistical techniques were used to evaluate spatial and temporal variations in surface water quality of the Ghrib Dam. All sampled parameters indicated significant temporal variability. From the chemical point of view, water in Ghrib Dam presents high mineralization in the water body. Nutrient concentrations, the organic matter, and total solid are high enough during the rainy season in the station S1 located upstream of Ghrib dam. All other stations undergo an almost constant distribution of all physicochemical parameters. From multivariate analysis, it could be construed that the dam water quality is primarily influenced by agricultural runoff and wastewater discharge. The PCA indicates that parameters responsible for water quality variation in the Ghrib Dam are mainly related to domestic and agricultural non-point sources or influenced by agricultural field runoff and waste transfer movements. Results of regression analysis clearly showed that runoff raises the concentration of most of the inorganic and organic parameters. Although the results the cluster analysis indicates that dam waters are vulnerable to pollution at the same time, after obtaining the two similar sets of water quality characteristics. Through these results, future strategies can be designed, including other stations at several points of the wadi Cheliff-Ghrib, to accurately identify sources of pollution that threaten the water quality of the Ghrib Dam.

References

Badillo-Camacho J, Reynaga-Delgado E, Barcelo-Quintal I, del Valle PFZ, López-Chuken UJ, Orozco-Guareño E, AlvarezBobadilla JI, Gomez-Salazar S (2015) Water quality assessment of a tropical Mexican lake using multivariate statistical techniques. J Environ Prot 6:215–224. https://doi.org/10.4236/jep.2015.63022

Bahroun S, Chaib W (2017) The quality of surface waters of the dam reservoir Mixa, northeast of Algeria. J Water Land Develop 34(1): 11–19 https://doi.org/10.1515/jwld-2017-0034

Barakat A, El Baghdadi M, Rais J, Aghezzaf B, Slassi M (2016) Assessment of spatial and seasonal water quality variation of Oum Er Rbia River (Morocco) using multivariate statistical techniques. J Soil Water Conserv 4:284–292. https://doi.org/10.1016/j.iswcr.2016.11.002

Bhat SA, Meraj G, Yaseen S, Pandit AK (2014) Statistical assessment of water quality parameters for pollution source identification in Sukhnag stream: an inflow stream of Lake Wular (Ramsar site), Kashmir Himalaya. Journal of Ecosyst 2014:18. https://doi.org/10.1155/2014/898054

Bhateria R, Jain D (2016) Water quality assessment of lake water: a review. Sustain Water Resour Manag 2:161–173. https://doi.org/10.1007/s40899-015-0014-7

Bouguerne A, Boudoukha A, Benkhaled A, Mebarkia AH (2016) Assessment of surface water quality of Ain Zada dam (Algeria) using multivariate statistical techniques. JRBM 15(2):133–143. https://doi.org/10.1080/15715124.2016.1215325

Boyd CE (2015) Dissolved solids. In: Water Quality. Springer, Cham. https://doi.org/10.1007/978-3-319-17446-4_4

Brodie JE, Mitchell AW (2005) Nutrients in Australian tropical rivers: changes with agricultural development and implications for receiving environments. Mar Freshw Res 56:279–302. https://doi.org/10.1071/MF04081

Cetin M (2016) Sustainability of urban coastal area management: a case study on Cide. J Sustain Forest 35(7):527–541. https://doi.org/10.1080/10549811.2016.1228072

Charulatha G, Srinivasalu S, Uma Maheswari O, Venugopal T, Giridharan L (2017) Evaluation of ground water quality contaminants using linear regression and artificial neural network models. Arab J Geosci 10:128. https://doi.org/10.1007/s12517-017-2867-6

Dixon W, Chiswell B (1996) Review of aquatic monitoring program design. Water Res 30(9):1935–1948. https://doi.org/10.1016/0043-1354(96)00087-5

Djelita B, Bouzid-Lagha S, Nehar KC (2016) Spatial and temporal patterns of the water quality in the Hammam Boughrara reservoir in Algeria. In: Grammelis P. (eds) Energy, transportation and global warming. Green energy and technology. Springer, Cham. https://doi.org/10.1007/978-3-319-30127-3_46

Guerraiche Z, Boudoukha A, Benkadja R (2016) Variation of the chemical composition of Grouz dam waters. Eastern Algeria. Desalin Water Treat 57(11):4878–4887. https://doi.org/10.1080/19443994.2014.995131

Gumbo JR, Dzaga RA, Nethengwe NS (2016) Impact on water quality of NandoniWater reservoir downstream of municipal sewage plants in Vhembe District, South Africa. Sustainability 8:597. https://doi.org/10.3390/su8070597

Harkat S, Arabi M, Taleb S (2011) Impacts of human activities on water erosion and pollution of surface water in the Algerian Cheliff watershed. LJEE 19:56–75 https://www.asjp.cerist.dz/en/article/37440

Hawkes JA, Patriarca C, Sjöberg PJR, Tranvik LJ, Bergquist J (2018) Extreme isomeric complexity of dissolved organic matter found across aquatic environments. L&O Lett 3(2):21–30. https://doi.org/10.1002/lol2.10064

Kazi TG, Arain MB, Jamali MK, Jalbani N, Afridi HI, Sarfraz RA, Shah AQ (2009) Assessment of water quality of polluted lake using multivariate statistical techniques: a case study. Ecotox Environ Safe 72(20):301–309. https://doi.org/10.1016/j.ecoenv.2008.02.024

Khatri N, Tyagi S (2015) Influences of natural and anthropogenic factors on surface and groundwater quality in rural and urban areas. Front Life Sci 8:23–39. https://doi.org/10.1080/21553769.2014.933716

Liu CW, Linn KH, Kuon YM (2003) Application of factor analysis in the assessment of groundwater quality in a black foot disease area in Taiwan. Sci Total Environ 313(1–3):77–89. https://doi.org/10.1016/S0048-9697(02)00683-6

Liu WC, Chen WB, Kimura N (2009) Impact of phosphorus load reduction on water quality in a stratified reservoir-eutrophication modeling study. Environ Monit Assess 159:393–406. https://doi.org/10.1007/s10661-008-0637-3

Mebarkia A, Haouchine A, Boudoukha A, Nedjai R (2017) Assessment of nutrient contamination in surface water, case study of Ain Zada Dam (north-east of Algeria). J Fundam Appl Sci 9(3):1358–1377. https://doi.org/10.4314/jfas.v9i3.8

Moss B (2008) Water pollution by agriculture. Philos Trans R Soc Lond Ser B Biol Sci 363(1491):659–666. https://doi.org/10.1098/rstb.2007.2176

National Agency for Dams and Transfers (2017) Annual operating report for Ghrib Dam (2012–2017). Water Resources Ministry of Algeria

N.A.W.R. National Agency for Water Resources Alger (2009). Annu Hydrol de l'Algérie (Alger, Hydrological Yearbook of Algeria). Water Resources Ministry

Ndungu J, Augustijn DCM, Hulscher SJMH, Fulanda B, Kitaka N, Mathooko JM (2015) A multivariate analysis of water quality in Lake Naivasha, Kenya. Mar Freshw Res 66(2):177–186. https://doi.org/10.1071/MF14031

Otto M (1998) Multivariate methods. In: Kellner R, Mermet JM, Otto M, Widmer HM (eds) Analytical chemistry. Wiley/VCH, Weinheim, p 916

Pandit AK, Yousuf AR (2002) Trophic status of Kashmir Himalayan lakes as depicted by water chemistry. J Res Dev 2:1–12

Peat J, Barton B, Elliott E (2009) Statistics workbook for evidence-based health care Evidence-Based Medicine , John Wiley & Sons

Penn CJ, Bowen JM (2018) Introduction to phosphorus and water quality. In: Design and Construction of Phosphorus Removal Structures for Improving Water Quality. Springer, Cham. https://doi.org/10.1007/978-3-319-58658-8_1

R Development Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria Available at http://www.R-project.org

Rodier J (2004) L'analyse de l'eau: Eaux naturelles, Eaux résiduaires, Eaux de mer. Dunod, Paris

Sánchez E, Colmenarejo MF, Vicente J, Rubio A, García MG, Travieso L, Borja R (2007) Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution. Ecol Indic 7:315–328. https://doi.org/10.1016/j.ecolind.2006.02.005

Sevik H, Cetin M (2015) Effects of water stress on seed germination for select landscape plants, pol. J Environ Stu 24(2):689–693. https://doi.org/10.15244/pjoes/30119

Shrestha S, Kazama F (2007) Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. Environ Model Softw 22:464–475. https://doi.org/10.1016/j.envsoft.2006.02.001

Simeonov V, Stratis JA, Samara C, Zachariadisb G, Voutsac D, Anthemidis A, Kouimtzisc T (2003) Assessment of the surface water quality in northern Greece. Water Res 37:4119–4124. https://doi.org/10.1016/S0043-1354(03)00398-1

Singh KP, Malik A, Sinha S (2005) Water quality assessment and apportionment of pollution sources of Gomti River (India) using multivariate statistical techniques-a case study. Anal Chim Acta 538:355–374. https://doi.org/10.1016/j.aca.2005.02.006

Touhari F, Mehaiguene M, Meddi M (2015) Study of the quality of surface water in the upper Cheliff Basin. Int J Environ Ecol Eng 9(12). https://scholar.waset.org/1307-6892/10006763