International Journal of River Basin Management

ISSN: 1571-5124 (Print) 1814-2060 (Online) Journal homepage: http://www.tandfonline.com/loi/trbm20

Assessment of surface water quality of Ain Zada dam (Algeria) using multivariate statistical techniques

Ammar Bouguerne, Abderrahmane Boudoukha, Abdelkader Benkhaled & Abdel-Hafid Mebarkia

To cite this article: Ammar Bouguerne, Abderrahmane Boudoukha, Abdelkader Benkhaled & Abdel-Hafid Mebarkia (2016): Assessment of surface water quality of Ain Zada dam (Algeria) using multivariate statistical techniques, International Journal of River Basin Management, DOI: 10.1080/15715124.2016.1215325

To link to this article: http://dx.doi.org/10.1080/15715124.2016.1215325

Accepted author version posted online: 22 Jul 2016.
Published online: 23 Aug 2016.

INTL. J. RIVER BASIN MANAGEMENT, 2016
http://dx.doi.org/10.1080/15715124.2016.1215325

RESEARCH PAPER

Assessment of surface water quality of Ain Zada dam (Algeria) using multivariate statistical techniques

Ammar Bouguerne[a], Abderrahmane Boudoukha[b], Abdelkader Benkhaled[c] and Abdel-Hafid Mebarkia[d]

[a]Department of Hydraulic, University of Batna, Fesdis-Batna, Algeria; [b]Laboratory of Applied Research in Hydraulics, University of Batna, Fesdis-Batna, Algeria; [c]Research Laboratory in Subterranean and Surface Hydraulics, University of Mohamed Khider, Biskra, Algeria; [d]Laboratory of Geo-Environment FSTGAT/USTHB, Algiers, Algeria

ABSTRACT

Multivariate statistical techniques, such as cluster analysis (CA), principal component analysis (PCA), factor analysis (FA) and discriminant analysis, have been applied for the assessment of temporal variations of surface water quality in Ain Zada dam, Algeria, for 10 years by monitoring 16 parameters. The different parameters indicate that the data are homogeneous. As against this record an annual variation is more important than the monthly change in connection with climate change. The facies of these waters is Cl–Na especially in connection with human actions. Values of the Water Quality Index classified the surface water as medium to good quality. The Pearson correlation analysis revealed a significant positive relationship between salinity and all variables and negative relationship between water volume of dam and all variables. The CA in R mode grouped the 16 variables into 4 clusters of similar water quality characteristics and in Q mode, 160 sampling are grouped into 2 statistically groups where total dissolve solids and capacity seem to be major distinguishing factors between variables and years. The CA has classified the data into two groups, one formed by the dry years and the other formed by wet years. The PCA and the FA applied to the datasets have resulted in two significant factors which represent 69.92% of total variance. The first factor as salinization factor explained 58.68% of the total variance. The second factor, can be called organic pollution factor, explained 11.24% of the total variance. The results of discriminant analysis showed only 11 parameters were necessary in the temporal variations analysis, affording more than 90% correct assignations.

ARTICLE HISTORY
Received 5 March 2016
Accepted 17 July 2016

KEYWORDS
Surface water; cluster analysis; principal component analysis; factor analysis; discriminant analysis; Ain Zada dam; Algeria

1 Introduction

Surface water quality is affected by both anthropogenic activities and natural processes. Natural processes influencing water quality include precipitation rate, weathering processes and sediment transport, whereas anthropogenic activities include urban development, industrial expansion and agricultural practices. Therefore, effective long-term management of a dam's water requires a fundamental understanding of chemical and biological characteristics. However, due to temporal variations in water quality (which are often difficult to interpret), a monitoring programme, providing a representative and reliable estimation of the surface waters quality, is necessary (Shrestha and Kazama 2007). Several studies (Etchanchu and Probst 2006, Guillaud and Bouriel 2007, Boudoukha and Boulaarak 2013) have reported the effects of agricultural, industrial and urban effluents on the quality of surface waters. The Ain Zada dam is constructed across the Bousselem river, which is geographically located at 5°08′57″ E longitude and 36°10′28″ N latitude. The Ain Zada dam in eastern Algeria is among the dams that supply drinking and industrial water to different urban and industrial centres in the region. The river accounts more than 90 % of Setif's water supplies and about two million people depend on river water for their daily use. This dam is experiencing a degradation of water quality due to different sources of pollution in addition to the natural processes of erosion and leaching of various elements which can cause deterioration of the water quality.

The aim of the research is also determination of the surface water quality of Ain Zada dam, presumably affected by agriculture and community activity, using pollution. Water Quality Index (WQI) is a way to summarize the data on the water quality in simple language to be used more easily by the user (Akkaraboyina and Raju 2012, Effendi *et al.* 2015).

There are several water quality indices that have been developed to support water quality divisions in USA, Canada, Indonesia, Algeria and Malaysia. However, most of these indices are based on the WQI developed by the US National Sanitation Foundation (NSF). The NSF developed an index, called the NSF-Water Quality Index (NSF-WQI), to provide a standardized method for comparing the relative quality of various water bodies (Said *et al.* 2004). Several researchers use the WQI in the assessment of river water quality (Bai *et al.* 2009). The water quality status might be considered in water resources management (Najah *et al.* 2009).

Application of environmetric techniques such as cluster analysis (CA), principal component analysis (PCA), factor analysis (FA) and discriminant analysis (DA) has increased significantly in recent years for analysing environmental data and drawing meaningful information (Shihab and Abdul Baqi 2010, Batayneh and Zumlot 2012). These tools are effective means of manipulating, interpreting and representing data concerning groundwater pollutants and geochemistry. They are frequently employed to characterize the quality of groundwater. These tools are also used to resolve hydrological factors, such as aquifer boundaries, groundwater

CONTACT Abderrahmane Boudoukha boudoukha_abderrahmane@yahoo.fr

flow paths and hydrochemical parameters (Belkhiri *et al.* 2011), to identify geochemical controls on the composition (Batayneh and Zumlot, 2012), to separate anomalies such as anthropogenic impacts from background (Güler *et al.* 2002) and to formulate geochemical models on the basis of available data (Belkhiri *et al.* 2010). I also discuss advantages of multivariate statistical techniques in assessment of water quality in reservoirs especially in comparison with other methods.

The CA has been utilized by various researchers to detect a group of sampling points into clusters on the basis of similarities/homogeneity within a cluster and dissimilarities/heterogeneity between different clusters: for example, CA techniques have been applied by Varol (2013) in their studies in the Tigris River Basin, Turkey. The CA grouped the sampling points into three clusters, 1, 2 and 3 which corresponded to low-, moderate- and high-polluted regions, respectively. Similarly, Hellar-Kihampa *et al.* (2013) applied the CA on surface water quality data to assess spatial variation in the Pangani River Basin, in Tanzania and the 12 sampling stations studied were grouped into two clusters: ionic species and nutrient loads clusters. The PCA and/or FA have been utilized by researchers to reduce the number of variables that are necessary to describe the observed variation within the datasets (Mustapha *et al.* 2014). The combination of PCA and FA as a data reduction technique is widely used, being capable of detecting similarities among samples and/or variables (Wang *et al.* 2006, Mendiguchıa *et al.* 2007). Various researches conducted around the world used the DA to reveal the most significant variables that result in surface water quality variation. Mustapha *et al.* (2013) studied river water quality variation in the Jakara River (Nigeria): a spatial DA was performed on the raw datasets of physico-chemical parameters and the standard DA constructs a discriminant function (DF) including all the variables used in the study. Thus, application of different multivariate statistical techniques can facilitate the interpretation of complex data matrices, and the understanding of how to increase the water quality and ecological status of a freshwater system (Sheela *et al.* 2012). It also allows for identification of possible factors influencing the water systems, as well as being valuable tools for facilitating the reliable management of water resources.

One of the main tasks of the National Agency of Water Resources (NAWR) is to ensure the qualitative and quantitative conservation of water resources. This paper presents details of fluctuations in the chemical composition of water of Ain Zada dam under the effect of a prolonged dryness between 2003 and 2012, and attempts to explain why the temporal variations have occurred in the concentrations of major dissolved components and the factors explaining the structure of the database and the influence of possible sources (natural and anthropogenic) on the water quality parameters of the Ain Zada dam. For this reason, the present work was carried out mainly for the waters of the dam where the pollution problem begins to grow.

2 Materials and methods

2.1 Study area

Sétif city is located at latitude 36°11′20″N and longitude 5°24′15″E with more than 2×10^6 inhabitants. The principal activity in this area is the production of cereals (wheat and barley) and a large diversified industry. The Ain Zada dam is located at 25 km west of Sétif on the Bousselem River which drains in a watershed presenting a total area of 1785 km^2 (Figure 1). The main river has a length of 65 km, with an exoreic flow. The dam has a capacity (V) of 125 million m^3, a maximum depth of 26 m and an area of 1.140 km^2. The region is characterized by a semi-arid climate and the rainfall station of the dam has recorded during the period 2003–2012 an average annual rainfall (P) of 327 mm and an average annual temperature (T_a) of 14.6°C. The annual evapotranspiration is 307 mm and the potential evapotranspiration is 840 mm showing an annual deficit of 513 mm. However, during the winter period, the annual excess is about 20 mm, which allows a constant flow during the period of high water (December to March) with an average water yield of 0.5 m^3/s (Bouguerne *et al.* 2010). During the dry season, the flow is tributary on wastewaters from urban (2×10^6 inhabitants) and industrial centres bordering the dam. The self-filtering of the various wastewaters becomes insufficient; thereby the enriching of the soluble and unsolvable material retention becomes very important. Agricultural land represents about 70% of the watershed area. Since the ground cover is weak, large amounts of suspended matter and other soluble components are mobilized towards the dam during heavy precipitation. The heavy rainfall of year 2003 (633 mm) has increased the volume of water in the dam (20–125 Mm3), which is a result of a very irregular rainfall regime. Then the volume of water from the dam has decreased gradually until 68 Mm^3 at the end of 2012.

2.2 Monitored parameters

Throughout the last 10 years, many hydrochemical data (prolonged study and ad hoc analyses) were acquired on surface water from the watershed Boussellem (Bouguerne 2001, Boudoukha *et al.* 2014) and the dam of Ain Zada. (Mebarkia 2012). In this context, these authors conducted a series of samples in order to monitor the quality of discharges into the receiving environment, and to study the temporal variation of chemical parameters from these waters with a view to understand the mechanisms of pollution. A total of 190 representative water samples were collected between January 2003 and December 2012. Samples were analysed for the major ions. The physico-chemical and organic data analysed are pH, temperature of water (T_w), total dissolve solids (TDS),

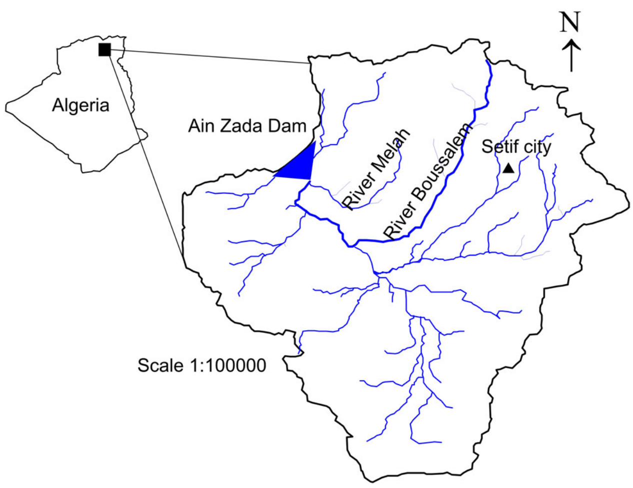

Figure 1. Map of study area.

turbidity (TU), calcium (Ca^{2+}), magnesium (Mg^{2+}), sodium (Na^{+}), potassium (K^{+}), chloride (Cl^{-}), sulphate (SO_4^{2-}), bicarbonates ($HCO3^{-}$), nitrate (NO_3^{-}), phosphate (PO_4^{3-}), dissolved oxygen (DO), chemical oxygen demand (COD), biochemical oxygen demand after 5 days (BOD_5), organic matter (OM) and suspended solids (SLS). The NAWR of Constantine city has analysed all the water quality parameters according to the standard methods, as suggested by the American Public Health Association (APHA 2005). The accuracy of the chemical analysis was verified by calculating ion-balance errors where the errors were generally within 5%. All statistical computations were made using Excel 2010 (Microsoft Office ®) and STATISTICA 6 (StatSoft, Inc. ®).

2.3 Water Quality Index

The NSF-WQI is used to determine the level of water quality, based on 9 parameters, namely BOD_5, DO, nitrate, total phosphate, temperature change (from 1 mile upstream), turbidity, TDS, pH and faecal coliform. This technic is developed by Brown *et al.* (1970) using the Delphi method was done by selecting parameters rigorously, developing a common scale and assigning weights to the parameters. NSF supported this index so it is also called as the NSF-WQI. It has been mentioned in many papers because it is the most comprehensive work (Pacini *et al.* 2013, Landwehr and Deininger 1976).

2.4 Statistical analysis

The application of different multivariate statistical techniques can facilitate the interpretation of complex data matrices, and can help to simplify and organize large datasets to provide meaningful insight (Laaksoharju *et al.* 1999). In our case, a matrix of 190 (19 variables × 10 years) can only be analysed then by statistical methods because other methods such as the various diagrams that can give only partial results. The main statistical techniques used are CA, PCA, FA and DA. Prior to commencing statistical analyses, the dataset should be assessed for normality and completeness of data (O'Shea and Jankowski 2006). These tests are to see if the data are suitable for use in the raw form or they require processing to result in more data to specify and therefore be used in a correct manner in the statistical analyses. Data producing a bell-shaped frequency distribution, with values clustered around a central point and the frequency of occurrence declining away from this point, are said to be normally distributed (Davis 1986). It is often assumed that variables are normally distributed and many statistical tests are based on this assumption. The FA and the PCA are an exception to this rule, as they are based only on Eigen analysis of the correlation/covariance matrix (Meglin 1991). However, the CA assumes that the data are either normal or log-normal (Güler *et al.* 2002); therefore, an assessment of normality was required before the CA could be undertaken.

When performing some statistical tests for homogeneity of variance it suffices to show that the test result is statistically significant ($p \leq .05$), it means that the data do not show homogeneity of variance. If the test is not significant ($p > .05$) then one can assume that the data show homogeneity of variance. This is possible using the Hartley's F_{max} (Hartley 1950) by dividing the larger variance by the smaller one. This gives us an F-ratio. If the variances are similar to each other, then the F-ratio will be close to 1: the more the variances differ, the larger the F-ratio will be. The Student t-test is also used to compare two sets of quantitative data when data in each sample set are related in a special way as in hydrochemical data.

Other statistical parameters used are mean, standard deviation (SD), coefficient of variation (Cv), kurtosis (Kurt), skewness (Skew) and analysis of variance (ANOVA). Mean explains average value; standard deviation gives a measure of 'spread' or 'variability' of the sample. Kurtosis and skewness are used for verifying whether the distribution of a sample is normal (Loether and McTavish 1988). It is considered that the sample follows a normal distribution to 95% when the value of kurtosis is between −2 and +2 and the value of the skewness is between −2 and +2 (Groeneveld and Meeden 1984). ANOVA is a statistical technique to test for significant differences between means by comparing variances.

For the CA, there are two types: R and Q modes. In clustering, the objects are grouped such that similar objects fall into the same class (Danielsson *et al.* 1999). A review by Belkhiri *et al.* (2010) suggests Ward's clustering procedure to be the best, because it yields a larger proportion of correct classified observations than do most other methods. Hence, Ward's clustering procedure is used in this study. As a distance measure, the squared Euclidean distance was used, which is one of the most commonly adopted measures (Fovell and Fovell 1993). In the present work, the CA was applied using the Euclidian distance as a distance measure between samples and Ward's method as a linkage rule for classification of the hydrochemical data of the Ain Zada area. The PCA method was also applied to the treatment of hydrochemical data. Although the PCA is an exploratory and descriptive method, the aim of the treatment is to identify the main factors that control the chemistry of the groundwater (Dagnelie 2006). This multivariate statistical method has been widely used to investigate the phenomena of the environment (Parmar and Bhardwaj 2014, Mustapha *et al.* 2014, Belkhiri *et al.* 2010, Tiri *et al.* 2016). These multivariate statistical tools were used successfully to study the hydrochemical processes and this work deals with the power of multivariate techniques to characterize hydrochemical variations in the area.

The FA provide more insight into the subjacent structure of a dataset, the use of these techniques might require further analyses to identify distinct groups (Belkhiri and Mouni 2012). The FA is related to the PCA, but the two are not identical (Zarei and Bilondi 2013). The PCA is far more commonly used than Principal Factor Analysis. However, it is common to use 'factors' interchangeably with 'components' in multivariate analysis (Belkhiri *et al.* 2010). The FA technique follows that of the PCA. The main purpose of the FA is to reduce the contribution of less significant variables to simplify even more the data structure coming from the PCA by rotating the axis defined by the PCA, according to well-established rules, and constructing new variables, also called varifactors (VFs). Principal Components (PC) is a linear combination of observable water quality variables, whereas VF can include unobservable, hypothetical, latent variables (Helena *et al.* 2000). The FA was performed on the normalized datasets, to compare the compositional pattern between analysed water samples and identify the factors that affect them.

The DA is a multivariate statistical technique which discriminates variables between two or more naturally occurring groups. The DA is a powerful technique that identifies the processes that control surface water chemistry (Shrestha and Kazama 2007), grouping samples of similar composition and origin (Juahir *et al.* 2011) and to predict the variables that differentiate the sampling stations temporally (Shrestha and Kazama 2007). Furthermore, the DA helps in grouping samples sharing common properties (Mustapha *et al.* 2014). A simple linear DF transforms an original set of measurements in a sample into a single discriminant score (Sanchez Lopez *et al.* 2004). The DA involves the determination of a linear equation that will predict to which group the case belongs (Asif *et al.* 2011). The form of the function is:

$$D = v1X1 + v2X2 + v3X3 \cdots viXi + a,$$

where
D = discriminate function;
v = the discriminant coefficient or weight for that variable;
X = respondent's score for that variable;
a = constant; and
i = the number of predictive variables.

3 Results and discussion

3.1 Statistics summary

Table 1 gives the detail of statistical analysis, including mean, standard deviation, kurtosis, skewness, ANOVA, Student and Hartley's F_{max} of variation for all water quality parameters. Examination of the standard deviation and the coefficient of variation show that EC, T_w, pH, DO and HCO_3^- are assigned to a low variation (<20%) around the mean. Ca^{2+}, Mg^{2+}, Na^+, Cl^- and $SO4^{2-}$ show a variation around the average between 20% and 30%. The remaining parameters, that is, K^+, NO_3^-, PO_4^{3-}, COD, BOD^5, SLS and OM, have a high variation around the mean (>50%). These large variations concern mainly the pollution parameters resulting from the effluents and the land leaching due to torrential heterogeneous rainfalls in time and space. All variables have values of kurtosis and skewness between −2 and +2 except Ca^{2+} and Mg^{2+} which, let us say, are consistent and follow a normal distribution as shown by the *t*-student tests. The examination of the monthly variance (Sm^2) and annual variance (Sy^2) shows that $Sy^2 > Sm^2$ in 63% of cases. This shows that the annual change is more important than the monthly change. Previous studies on the impact of climate change on the chemistry of the surface waters of this dam (Boudoukha *et al.* 2014) showed that the annual chemical variability is in close liaison with the annual variability of precipitation. The dam supply is largely ensured by the rains which leach sewage and agricultural spreading. This annual–monthly variability is confirmed by the Hartley's F_{max} test at 5% significance level.

3.2 General chemistry

The examination of the chemical analyses (Table 1) shows that the pH varies between 7.2 and 8.8, with a mean value of 8.3 ± 0.3 indicating slightly basic, it characterized waters where life develops optimally (Tsytsarin 1988). Water surface of the Ain Zada dam has pH ≥ 8 in 86% of cases thus highly alkaline due to the intense evaporation from the lake dam (MSDEWP 2013) and release of basic industrial and agricultural effluents (Dinkaa *et al.* 2015). TDS has a mean value of 614 ± 90.1 mg/l. According to Freeze and Cherry (1979), water in the Ain Zada dam was fresh water (TDS < 1000 mg/l). Alkalinity has a mean value of 168.1 ± 31.8 mg/l. The Na^+ ion has a mean value of 94.9 ± 23.1 mg/l. K^+ has a mean value of 6.6 ± 4.7 mg/l. These low concentrations are due to the fact that K minerals have weak migration (Kempe 1982), are resistant to decomposition by weathering (Pradhan and Pirasteh 2011) and are largely derived from the fertilizers (Guerraiche *et al.* 2016). The Ca^{2+} parameter has a mean value of 71.4 ± 17.5 mg/l. The Mg^{2+} ion has a mean value of 29.9 ± 8.5 mg/l. Cl^- has a mean value of 144.4 ± 33.4 mg/l. SO_4^{2+} has a mean value of 159.3 ± 33.9 mg/l. Therefore, the facies of this water is Cl>SO4 >HCO3 for anions and Na>Ca>Mg>K for cations in 87% of cases. This is in liaison with the evaporation phenomenon and wastewater (Guerraiche *et al.* 2016). The NO_3- parameter has a mean value of 4.8 ± 3.6 mg/l. The low concentration of nitrates in surface water is due to their reduction by bacteria (Martin 1979).

The COD parameter has a mean value of 46.8 ± 19.5 mg/l. BOD_5 has a mean value of 3.1 ± 1.7 mg/l. OM has a mean value of 9.5 ± 3.1 mg/l. SLS has a mean value of 55.8 ± 26.7 mg/l. These concentrations showed that the weight of material stored at the dam is considerable (Armah *et al.*

Table 1. Statistics of water quality parameters of the Ain Zada dam during 2003–2012.

Parameters	Min.	Mean	Max.	SD	Cv (%)	Sm^2	Sy^2	Kurt	Skew	*T*	*F*
V	19.7	87.5	125	26.2	30	37.9	594.9	−0.17	−0.71	0.44	15.70
T_w	4	16.7	29	6.4	03	32.1	0.95	−1.36	0.09	0.17	33.80
pH	7.2	8.3	8.8	0.3	03	0.02	0.01	1.40	−0.47	0.08	2.00
TDS	448	614	832	90.1	14	762.9	12201	−0.99	−0.02	0.29	15.99
Ca^{2+}	36	71.4	174	17.5	25	22.1	40.9	10.51	1.76	0.04	1.85
Mg^{2+}	11	29.9	73	8.5	28	4.6	53.7	2.25	1.36	0.14	11.67
Na^+	44	94.9	166	23.1	24	21.1	264.8	−0.37	0.27	0.21	12.55
K^+	1	6.6	13.5	4.7	63	0.6	6.27	−1.07	0.14	0.22	10.45
Cl^-	60	144.4	235	33.4	23	42.4	523.1	−0.20	0.05	0.17	12.34
SO_4^{2-}	54	159.3	258	33.9	22	70.3	243.7	0.63	0.01	0.08	3.47
HCO_3^-	79.3	168.1	238	31.8	19	266.3	182.2	−0.28	−0.16	0.10	1.46
NO_3^-	1	4.8	17	3.7	78	0.6	2.6	1.37	1.16	0.01	4.33
DO	5	9.1	14.1	1.75	19	21.5	75.7	0.46	0.01	0.64	29.11
COD	8	46.8	99	19.5	54	6.7	96.6	0.16	0.42	0.11	14.42
BOD_5	1	3.1	8.5	1.7	6	7.4	0.2	1.54	0.95	0.11	37.00
OM	18	95	188	49	52	7.5	1.1	0.80	0.42	0.11	6.80
SLS	10	55.8	120	26.7	58	127.4	882.6	−0.92	0.12	0.03	6.93
PO_4^{3-}	0.03	0.18	0.35	0.09	47	37.8	0.01	−0.77	0.17	0.35	0.37
TU	0.9	3.16	8.8	2.04	64	32.03	4.17	1.63	1.47	0.15	7.68

Notes: Concentrations in mg/l, T_w in °C, *V* in Mm^3 and TU in NTU.
Min, minimum; Max, maximum; SD, standard deviation; Cv, coefficient of variation; Sm^2, intermensual variance; Sy^2, interannual variance; Kurt, kurtosis; Skew, skewness; *T*, value of the student *t*-test at 5% significance level; *F*, value of the Hartley's F_{max} at 5% significance level.

2010). The increase in this polluting load over time is marked by a positive slope COD (0.02), BOD_5 (0.01), OM (0.02) and SLS (0.01) and negative slopes of DO (−0.007) and pH (−0.0005) reflecting an oxidation of material organic material according to the reactions 1 and 2 (Kempe 1982).

$$CH_2O + 0_2 \leftrightarrow CO_2 + H_2O, \quad (1)$$

$$CO_2 + H_2O \leftrightarrow H_2CO_3 \leftrightarrow H^+ + HCO_3^-. \quad (2)$$

3.3 Water quality

For water quality, in this study, seven parameters were applied without faecal coliform and temperature change between 2003 and 2012, the period during which these parameters were measured. The weight score (Wi) was multiplied by the sub-index value (Li) of parameter-I obtained by Calculator NSF-WQI Online. (http://www.water-research.net/watrqualindex/ index.htm). The WQI ranges have been defined according to Brown *et al.* (1970) (Table 2). Based on the analysis using the NSF-WQI, surface water of Ain Zada dam classification in each year is presented in Table 3. The water quality at every year was almost the same. Those WQI values ranged 69–81 and can be classified as medium to good quality; this characteristic is dependent on the low activity intensity of the surrounding area (Valeriani *et al.* 2015, Boudoukha *et al.* 2014). That might due to decomposition process of OM from decomposing plant by microbes consuming DO.

Table 3. WQI value of surface water of the Ain Zada dam.

Year	WQI	Score classification
2003	73	Good
2004	69	Medium
2005	73	Good
2006	77	Good
2007	78	Good
2008	80	Good
2009	79	Good
2010	81	Good
2011	81	Good
2012	79	Good

3.4 Statistical analysis

3.4.1 Pearson correlation coefficients

The Pearson correlation coefficient matrix for the Ain Zada area is given in Table 4. The results show that the parameters in the water samples are correlated at $p < .05$ level. The pH shows a significant positive correlation (0.55–0.60) with carbonate elements (HCO_3, Ca and Mg). This is in liaison with the calc-carbonic equilibrium where the pH influences the dissolution of carbonate rocks. The pH values show significant negative correlation (−0.46 to −0.58) with anthropic elements (K^+, Cl^-, SO_4^{2-}, PO_4^{3-} and NO_3^-) linked to industrial effluent, and/or agricultural activities and with pollution parameters (COD, BOD_5 and OM) related to domestic sewage. Salinity, represented by TDS, shows a positive correlation (0.58–0.74) with all elements except K^+ and NO_3^-. The positive correlation indicated that the water was mainly controlled by Ca^{2+}, Mg^{2+}, Na^+, Cl^-, SO_4^{2-} and HCO_3^- ions, which depend upon the mineral dissolution, mineral solubility, ion exchange, evaporation and anthropogenic activities (Mahtab *et al.* 2015). The negative correlation between TDS and V indicates a reverse change by concentration of the water salinity following a decrease in water volume in the dam due to intense evaporation from the lake (Boudoukha *et al.* 2014). This analysis shows also a positive correlation (0.61–0.79) between SO_4^{2-}, Ca^{2+}, Mg^{2+}, Na^+, K^+ and Cl^-. This is attributed to the effect of anthropogenic action of waste water (Armah *et al.* 2010). The NO_3^- shows a positive correlation (0.57–0.60) with SO_4^{2-} and K^+, indicating that agricultural practices and human economic activities are significant. This analysis shows also a positive correlation (0.49–0.72) with parameters pollution (COD, BOD_5, OM, TU and SLS). The positive relationship between these parameters indicates the oxidation of OM by microorganism in the water (Armah *et al.* 2010). Municipal and industrial wastewaters are polluted with solids originating from domestic waste, run-off, urban development and uncontrolled land use through urbanization increasing SLS and TU with the degree of water pollution (Mustapha *et al.* 2014). Many researchers globally reported that the nonlinear relationship between DO and organic compound in water is due to anaerobic conditions in the river from the high concentration of dissolved OM (Wu *et al.* 2009, Mustapha *et al.* 2014). The volume of water of the dam, in turn, is negatively correlated (−0.52 to −0.72) with all parameters. This indicates dilution and concentration phenomena with respect to time.

Table 2. Weight score (Wi) for nine parameters on the NSF-WQI and classification criteria.

Parameters	Weight factor	NSF-WQI	Criteria score
DO	0.17	90–100	Excellent
Faecal coliform	0.16	70–90	Good
pH	0.11	50–70	Medium
BOD5	0.11	25–50	Bad
Change *T*	0.10	0–25	Very bad
Total phosphate	0.10	90–100	Excellent
Nitrates	0.10		
Turbidity	0.08		
TDS	0.07		

3.4.2 Cluster analysis

In this CA, the R mode is used with only 16 hydrochemical measured variables (TDS, pH, SLS, T_w, V, Ca, Mg, Na, K, Cl, SO4, HCO3, NO3, COD, BOD5 and OM). For statistical purpose, all the variables were log-transformed and more closely correspond to normally distributed data. Subsequently, they were standardized to their standard scores (*z*-scores), as described by Güler *et al.* (2002). As there is no test to determine the optimum number of groups in the dataset, the visual control is the only criterion to select the groups in the dendrogram (Figure 2). The defined phenon line (Sneath and Sokal 1973) was chosen at a linkage distance of 900. At this distance, the groups could be distinguished in terms of their hydrochemical variables. As shown in Figure 2, 16 variables were classified based on visual examination into 4 groups of water where TDS and V seem to be major distinguishing factors between variables. Group 1 (G1) comprises TDS and V. V controls significantly salinity and therefore soluble chemical elements. Group 2 (G2) is formed by Na, Cl and SO4, elements which come mainly from wastewater and secondarily from dissolving the saliferous minerals. Group 3 (G3) covers pH, Ca, Mg, HCO3, K and NO3. This group is influenced by carbonate and agricultural practices. Group 4 (G4) covers T_w, OM, COD, BOD5 and SLS. It is influenced

Table 4. Pearson coefficients correlation for physico-chemical parameters at the .05 level and for *n*=190.

	Ca	Mg	Na	K	Cl	SO_4	HCO_3	NO_3	PO_4	TDS	V	Tw	COD	BOD	OM	DO	SLS	pH	TU
Ca	1.00																		
Mg	**0.67**	1.00																	
Na	**0.53**	0.06	1.00																
K	0.18	0.11	**0.54**	1.00															
Cl	0.19	0.26	**0.78**	**0.73**	1.00														
SO_4	**0.70**	**0.78**	**0.61**	**0.79**	**0.62**	1.00													
HCO_3	**0.70**	**0.71**	0.16	0.02	0.06	0.13	1.00												
NO_3	0.05	0.10	−0.01	**0.57**	0.17	**0.60**	0.04	1.00											
PO4	0.04	0.14	−0.02	**0.58**	0.18	**0.55**	0.03	**0.65**	1.00										
TDS	**0.62**	**0.74**	**0.64**	0.17	**0.70**	**0.65**	**0.58**	0.25	0.01	1.00									
V	**−0.72**	**−0.69**	**−0.63**	**−0.52**	**−0.64**	**−0.63**	**−0.59**	**−0.64**	**−0.56**	**−0.62**	1.00								
Tw	−0.02	0.08	0.02	0.14	0.06	0.14	0.19	0.07	0.07	0.10	**−0.49**	1.00							
COD	0.09	−0.01	0.16	0.02	0.24	0.12	0.02	0.10	0.01	0.22	**−0.62**	0.07	1.00						
BOD	0.09	−0.01	0.10	0.03	0.07	0.07	0.05	0.03	0.02	0.04	**−0.60**	0.13	**0.54**	1.00					
OM	0.13	−0.04	0.11	0.14	0.19	0.19	0.01	0.08	0.07	0.15	**−0.55**	0.02	**0.72**	**0.57**	1.00				
DO	0.08	0.11	0.12	0.13	0.17	0.08	0.05	**−0.63**	0.02	0.04	**0.60**	0.13	**−0.54**	**−0.55**	**0.90**	1.00			
SLS	0.12	0.09	0.14	0.34	−0.05	0.11	0.01	0.18	0.14	0.14	**−0.49**	0.01	**0.52**	**0.49**	**0.56**	**−0.59**	1.00		
pH	**−0.66**	**−0.60**	**−0.50**	**−0.44**	**−0.54**	**−0.49**	**0.55**	**−0.46**	**−0.47**	0.01	−0.16	−0.06	**−0.55**	**−0.50**	**−0.58**	**−0.42**	0.14	1.00	
TU	0.31	0.19	0.24	0.24	−0.15	0.21	0.11	0.28	0.27	0.24	**−0.59**	0.01	**0.55**	**0.51**	**0.57**	0.20	**0.56**	0.12	1.00

Notes: The bold values indicate the correlated variables.

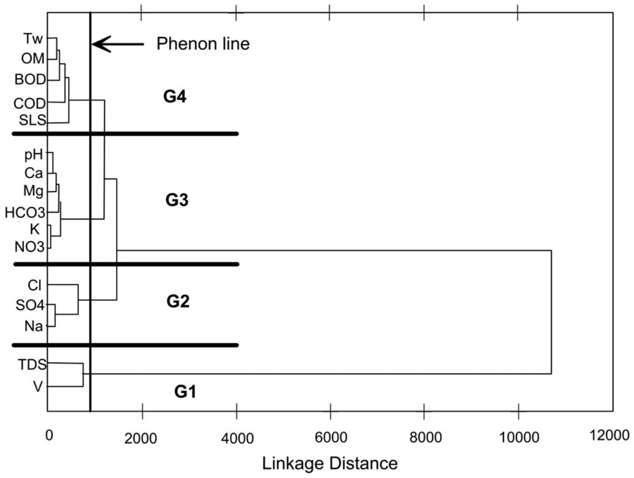

Figure 2. Dendrogram showing grouping of variables.

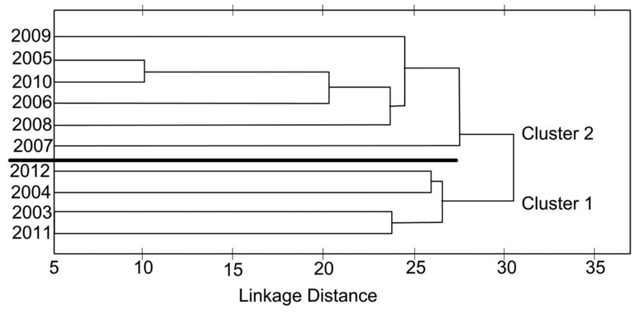

Figure 3. Dendrogram showing grouping of studied samples.

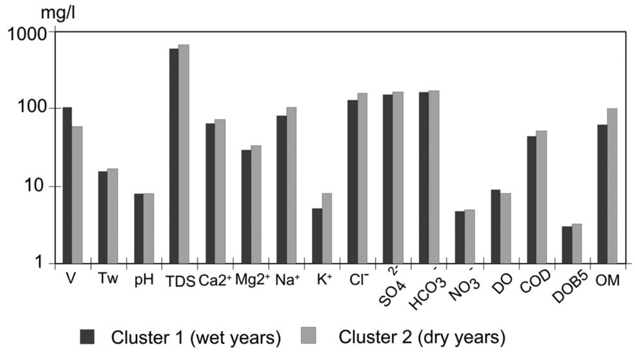

Figure 4. Comparison of physico-chemical parameters between clusters 1 and 2.

by industrial activities, wastewater discharge and natural processes.

On the other hand, the Q mode reveals the interactions between the studied samples (Figure 3). This CA grouped all 160 sampling for 10 years into 4 statistically groups. The result was synthesized according to the years in both groups (clusters 1 and 2) according to the TDS and V. Cluster 1 (C1) with high median values of V and low median values of TDS by against cluster 2 (C2) is characterized by a low median values of V and high median values of TDS. C1 covers the samples of years: 2005, 2006, 2007, 2008, 2009 and 2010. This period is characterized by dryness (precipitation < 300 mm) which resulted in the decrease in median values of V of water in the dam which was accompanied by an increase in the median values of salinity by concentration. C2 covers the samples of 2003, 2004, 2011 and 2012 years which coincide with the beginning and end of the study period. It is characterized by relatively higher rainfall (precipitation > 300 mm) which resulted in the increase in median values of V of water in the dam. This is accompanied by a reduction in median values of salinity by dilution (Figure 4). However the precipitations permit a cooling water, a decrease in median values of T_w (17.3°C to 15.5°C) and oxygenation of dam water which resulted in a slight increase in median values of OD (8.03 to 9.25 mg/l). The difference in median values between all parameters excepted DO in C1 and C2 is the result of heavy rainfall in the rainy period. So the CA has classified the data into two groups, one formed by the dry years and the other by wet years.

3.4.3 Principal component analysis/factor analysis

The Kaiser–Meyer–Olkin (KMO) is a measure of sampling adequacy that provides an index between 0 and 1 reflecting the proportion of variance among the variables (Wang *et al.* 2012). The KMO result obtained in this study is 0.75 and the Bartlett test of sphericity is very significant (0.0001, $p < .05$), indicating that the PCA and FA could be considered as an appropriate and useful tool to provide significant reduction in the dimensionality of the data (Hinkle *et al.* 2003). Based on these criteria, one PCA was performed for the set of 160 samples and 16 variables. Table 5 shows the eigenvalues of the extracted factors and the proportion of total sample variance which are explained by the factors. The analysis generates 10 factors and only 2 significant factors which have Eigenvalues >1 and explain 65.1% of total variance. Results from the PCA are more significant when the number of factors is small (Kraiem *et al.* 2014). The parameters loading for the 2 components from the PCA of the dataset are given in Table 5.

The Ca, Mg, Na, Cl, SO4, HCO_3, TDS, pH and V parameters are marked as factor 1 (F1) and explain 58.68% of the variance. Factor 1 has a strong to moderate positive loading in Ca, Mg, Na, Cl, SO4, HCO3 and TDS which were 0.62, 0.56, 0.70, 0.71, 0.68, 0.63 and 0.84, respectively. High positive loadings indicated strong linear correlation between the factor and parameters. Thus, F1 can be termed as salinization factor. This indicates a first opposition between the parameters of the mineralization and V by dilution (Figure 5). The second opposition is between pH and carbonate elements that require an acidic environment to allow dissolution. F1 is the factor of the natural contamination which progresses

Table 5. Variance explained and component matrixes.

Variables	F1	F2
Ca	**0.62**	−0.15
Mg	**0.56**	−0.15
Na	**0.70**	0.18
K	0.35	**0.67**
Cl	**0.71**	0.49
SO4	**0.68**	−0.13
HCO_3	**0.63**	0.28
NO_3	0.02	**0.68**
TDS	**0.84**	0.13
V	**−0.81**	0.21
Tw	0.16	−0.18
COD	0.29	**−0.57**
BOD	0.12	**−0.61**
OM	0.28	**−0.74**
SLS	0.32	**−0.52**
pH	**−0.62**	0.17
Eigenvalue	2.98	1.79
% Total variance	58.68	11.24
Cumulative %	58.68	69.92

Notes: The bold values indicate the correlated variables.

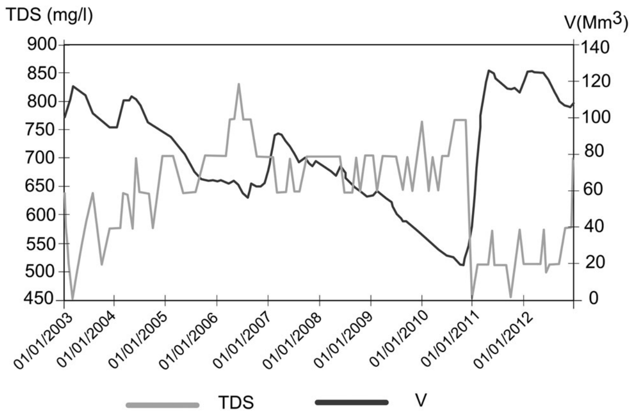

Figure 5. Variability of EC vs V during the study period.

inversely proportional to the water volume stored in the dam. Heavy rainfall during the wet years causes CO_2 dilution that allows the formation of carbonic acid responsible for the dissolution of carbonate formations. Loading values on F1 shows pH as a controlling factor of the alkaline earth elements (Ca^{2+} and Mg^{2+}). Researchers have shown that the quality of surface water in a region is governed by the natural factors which may include precipitation rate, weathering processes and soil erosion largely influenced by the climate in the basin (Singh *et al.* 2004).

Factor 2 (F2) explains 11.24 % of the total variance of the dataset and shows a strong to moderate positive loading in K (0.67), NO3 (0.68) and a moderate negative loading in COD (−0.57), BOD5 (−0.61) OM (−0.74) and SLS (−0.52). Therefore, F2 is associated with an opposition between chemical fertilizers and organic pollution from domestic wastewater. This factor could be linked to surface run-off that may be carried by soil erosion from the non-point source of water pollution and finally dissolved all inorganic particles in the water.

Surface water pollution in the Ain Zada is generated by domestic waste through dumping of biodegradable organic pollutants, nutrient and industrial effluent discharge. This is done by the release of organic and inorganic pollutants coming from the areas of drainage containing fertilizer, agricultural pesticides and suspended materials.

There is clear evidence from the FA of two major factors in the study areas that sources apportionment of water pollution could either be natural or anthropogenic (Huang *et al.* 2010). Figure 6 shows the score plot for the two PCs explaining 69.92% of the total variance. Four groups of water samples defined by high nutrient content, high salinity, polluted OM and low salinity, respectively, were classified through the diagram of the scores for PC 1 versus PC 2. The results show more variations depending on the concentration of the physico-chemical parameters and pollution elements. This distribution is similar to the CA but it was better detailed by the PCA/FA. The new distribution mentions details in the two groups, salted and unsalted. This new distribution indicates probably changes in the water chemistry due to a source of contamination or processes such as dilution or concentration (Belkhiri *et al.* 2011).

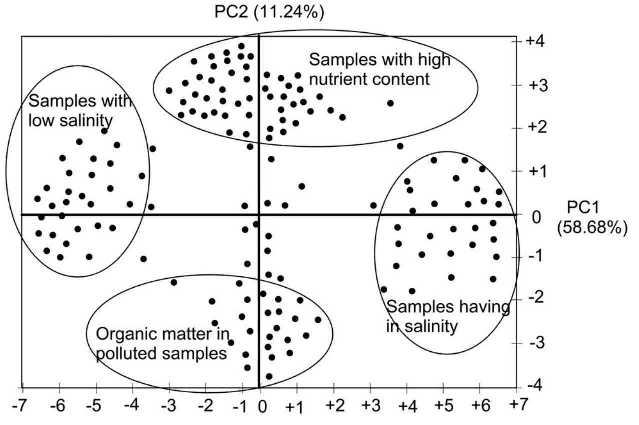

Figure 6. Plots of PC scores for PC1 versus PC2.

3.4.4 Discriminant analysis

Variation in water quality parameters was evaluated through the DA. The DA technique applied on raw data consisted of 16 parameters used to find one or many functions (linear combinations) of the observed data (called discriminates functions). To better assess the water quality of Ain Zada dam after grouping the data into two major classes of dry years and wet years as obtained by the CA method. The standard mode DA was applied in the present study. The DA was applied on raw data. Two discriminate functions (DFs) were found to discriminate the two major periods of surface water of the Ain Zada dam, as shown in Table 6. The Wilk's Lambda test show that both functions are statistically significant, as indicated in Table 7. Consequently, 90 % of the total variance is explained by the two DFs. The first DF explained 74.5% of the total spatial variance, and the second DF explained 15.5 %. The relative contribution of each parameter to both functions is given in Table 8. The relative contribution for each parameter is given in Eqs (3) and (4). A total of 11 parameters among 16 were determined by the two functions:

$$\begin{aligned} \mathrm{DFs1} &= 0.54\mathrm{Ca} + 0.73\mathrm{Na} + 0.85\mathrm{Cl} + 0.72\mathrm{SO_4} \\ &\quad + 0.58\mathrm{HCO_3} + 0.90\mathrm{TDS} - 0.88\mathrm{V}, \end{aligned} \tag{3}$$

$$\begin{aligned} \mathrm{DFs2} &= -0.55\mathrm{NO_3} - 0.87\mathrm{COD} - 0.71\mathrm{BOD} \\ &\quad - 0.91\mathrm{OM} + 0.75\mathrm{SLS}. \end{aligned} \tag{4}$$

The first function takes into account the most of chemical parameters (TDS, V, Cl, Na, SO4, HCO3 and Ca) and exhibited strong contribution in discriminating the two periods. It reports on the total temporal variation in the surface quality of Ain Zada dam water. The second function takes into account the pollution parameters, such as OM, COD, SLS, BOD and NO3. The relative contribution for water quality parameters can be arranged in the order:

- First function: TDS > V > Cl > Na > SO_4 > HCO_3 > Ca.
- Second function: OM > COD > SLS > BOD > NO_3.

Table 6. Eigenvalues for two DFs.

Function	Eigen value	% Variance	Cumulative %
1	4.16	74.5	74.5
2	3.55	15.5	90.0

Table 7. Wilk's Lambda test of DFs for temporal variation of water quality.

Test of function	Wilk's Lambda	Chi-square	*p* Level
1	0.081547	412.3319	.000
2	0.461340	127.2604	.000

Table 8. Discriminant function coefficients of temporal variation of water quality.

Parameters	Function 1	Function 2
Ca	**0.54**	−0.09
Mg	0.33	0.11
Na	**0.73**	−0.19
K	0.07	−0.18
Cl	**0.85**	0.47
SO4	**0.72**	−0.07
HCO3	**0.58**	0.06
NO3	0.24	**−0.55**
TDS	**0.90**	−0.17
V	**−0.88**	0.27
T_w	0.12	0.30
COD	0.38	**−0.87**
BOD	0.27	**−0.71**
OM	0.11	**−0.91**
SLS	0.14	**0.72**
pH	0.30	0.10

Note: Boldfaces are significant parameters at $p < .05$.

Table 9. Classification results for the DA of the water dam.

		Predicted group membership	
Period	% Correct	1	2
Wet years	90	16	0
Dry years	90	0	16

Note: 90.0% of original grouped cases correctly classified.

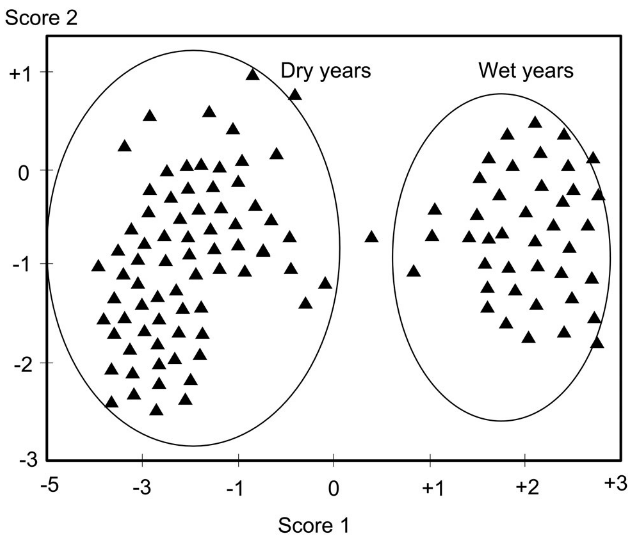

Figure 7. Canonical scores plot.

The classification matrix showed that 90% of the cases are correctly classified to their respective groups, as shown in Table 9. The DA method gave the best results for both the temporal analyses for two period years. For 16 parameters of the dam water, it yielded an important data reduction, as it used only 11 parameters (TDS, V, Ca, Cl, Na, SO4, HCO3, NO3, OM, COD, BOD5 and SLS) to discriminate between the two period years. Scores of two functions were plotted and the result clearly distinguishes the quality of surface water of the two periods, as shown in Figure 7.

4 Conclusion

In this study, statistical methods were used to evaluate temporal variations in the surface water quality of Ain Zada dam. Water quality-monitoring programmes generate complex data that need multivariate statistical treatment. The sources of hydrochemical in the dam water might be most likely derived from rock, urban and industrial wastewaters, and sediment loadings. The summary statistics show low (<30%) variation for physico-chemical parameters and high (>50%) organic parameters. These large variations concern mainly the pollution parameters resulting from the effluents and the land leaching due to torrential heterogeneous rainfalls in time and space. The different parameters indicate that the data are homogeneous. An annual variation is more important than the monthly change in connection with climate change. The facies of these waters is Cl–Na especially in connection with human actions. Values of the WQI classified the surface water as medium to good quality. The Pearson correlation coefficient shows a significant positive correlation with carbonates elements and pH, a significant positive correlation with TDS and all parameters. This indicates that these elements are the main components of salinity. A significant negative correlation with V and all parameters indicates dilution and concentration phenomena with time. The CA technique in R mode grouped the 16 sampling stations into 4 clusters of similar water quality characteristics, and in Q mode 160 sampling are grouped into 4 statistically groups where TDS and V seem to be major distinguishing factors between variables and years.

The PCA and FA methods helped to identify that the parameters responsible for water quality variations were mainly related to physico-chemical and organic. The analysis generates 10 factors and only 2 significant factors represent 69.92% of total variance. The first factor termed as salinization factor, explained 58.68 % of the total variance. The second factor which can be defined as organic pollution factor, explained 11.24% of the total variance. This analysis revealed that the dam water quality was mainly controlled by domestic wastewater, industrial discharges and agricultural run-off.

The DA method indicates the 11 significant parameters, which discriminate the dam water quality over 10 years (wet and dry years). It also classified the 10 years 90% correctly. Therefore, the DA method allowed a reduction in the dimensionality of the large dataset and indicated in particularly a few significant parameters that are responsible for large variations in water quality that could reduce the number of sampling parameters. Hence, this study illustrates that multivariate statistical methods are an excellent exploratory tool for interpreting surface water quality datasets and for understanding temporal variations, which are useful for water quality management.

Acknowledgements

The authors greatly appreciate the constructive and thoughtful comments of the anonymous reviewer. They also thank Professor James E. Ball of School of Civil and Environmental Engineering-University of Technology, Sydney, and Editor in Chief of the *International Journal of River Basin Management* for his kind cooperation.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCiD

Abderrahmane Boudoukha http://orcid.org/0000-0001-7889-3863

References

Akkaraboyina, M.K. and Raju B.S.N., 2012. Assessment of water quality index of River Godavari at Rajahmundry. *Universal Journal of Environmental Research and Technology*, 2 (3), 161–167.

APHA (American Public Health Association), AWWA (American Water Works Association) and WEF (Water Environment Federation). 2005. *Standard methods for the examination of water and wastewater*. 2nd ed. APHA-AWWA-WEF: Washington, DC.

Armah, F.A., *et al.*, 2010. Anthropogenic sources and environmentally relevant concentrations of heavy metals in surface water of a mining district in Ghana: a multivariate statistical approach. *Journal of Environmental Science and Health*, 45 (13), 1804–1813.

Asif, M., *et al.*, 2011. Application of multivariate statistical techniques for the characterization of ground water quality of Lahore, Gujranwala and Sialkot (Pakistan). *Pakistan Journal of Analytical & Environmental Chemistry*, 12 (1–2), 102–112.

Bai, R.V., Bouwmeester, R., and Mohan, S., 2009. Water quality index and importance of water quality parameters. *Air, Soil, and Water Research*, 2, 51–59.

Batayneh, A. and Zumlot, T., 2012. Multivariate statistical approach to geochemical methods in water quality factor identification; application to the shallow aquifer system of the Yarmouk Basin of North Jordan. *Research Journal of Environmental and Earth Sciences*, 4 (7), 756–768.

Belkhiri, L., *et al.*, 2010. Application of multivariate statistical methods and inverse geochemical modeling for characterization of groundwater, A case study, Ain Azel plain (Algeria). *Geoerma*, 159, 390–398.

Belkhiri, L., Boudoukha, A., and Mouni, L. 2011. A multivariate statistical analysis of groundwater chemistry data. *International Journal of Environmental Research*, 5 (2), 537–544.

Belkhiri, L. and Mouni, L., 2012. Hydrochemical analysis and evaluation of groundwater quality in El Eulma area, Algeria. *Applied Water Science*, 2, 127–133.

Boudoukha, A. and Boulaarak, M., 2013. Water pollution Grouz dam by nutriments (Eastern Algeria). *Bulletin Natural Geological Service*, 24, 139–149.

Boudoukha, A., Bouguerne, A., and Mebarkia, A., 2014. Impact of climate change and human activity on water quality of the dam Ain Zada. Algeria. *International Journal of Global Change*, 2 (3), 74–75.

Bouguerne, A., 2001. *Comparative study by hydrological and hydrochemical analysis of surface waters watersheds of Oued Bousselem and Oued Rhumel*. Magister Dissertation. University of Batna.

Bouguerne, A., Boudoukha, A., and Tiri, A., 2010. Evaluation of surface water quality in watershed with HBV code and pollution risk by heavy metals: case of two regions is the eastern Algerian. *Journal International Network Environmental Management Conflicts, Santa Catarina, Brazil*, 1 (1), 120–127.

Brown, R.M., *et al.*, 1970. A water quality index. Do we dare? *Water and Sewage Works*, 117 (10), 339–343.

Dagnelie, P., 2006. *Theoretical and applied statistics*, Tome 2, *In*: De Boeck and Larcier, eds. *Inferences and a two dimensional*, Brussels: Edis, Boeck & Larcier, Bruxelles-Univ.

Danielsson, A., *et al.*, 1999. Spatial clustering of metals in the sediments of the skagerrak/kattegat. *Applied Geochemistry*, 14, 689–706.

Davis, J.C., 1986. *Statistics and data analysis in geology*. 2nd ed. New York: Wiley.

Dinkaa, M.O., *et al.*, 2015. Hydrochemical characterization of various surface water and groundwater resources available in Matahara areas, Fantalle Woreda of Oromiya region. *Journal of Hydrology: Regional Studies*, 3, 444–456.

Effendi, H., Romanto, W., and Wardiatno, Y. 2015. Water quality status of Ciambulawung River, Banten province, based on pollution index and NSF-WQI. *Procedia Environmental Sciences*, 24, 228–237.

Etchanchu, D. and Probst, J.L., 2006. Evolution of the chemical composition of the Garonne river water during the period 1971–1984. *Journal of Hydrologique Science*, 33 (3), 243–256.

Fovell, R. and Fovell, M.Y., 1993. Climate zones of the conterminous United States defined using cluster analysis. *Journal of Climate*, 6, 2103–2135.

Freeze, R.A. and Cherry, J.A., 1979. *Groundwater*. Englewood Cliffs, NJ: Prentice Hall.

Groeneveld, R.A. and Meeden, G., 1984. Measuring skewness and kurtosis. *The Statistician*, 33, 391–399.

Guerraiche, Z., Boudoukha, A., and Benkadja, R. 2016. Variation of the chemical composition of Grouz dam waters. Eastern Algeria. *Desalination and Water Treatment*, 57 (11), 4878–4887.

Guillaud, J.F. and Bouriel, L., 2007. Concentration-flow and time evolution of the nitrate I 25 rivers in Brittany (France). *Journal of Water Science*, 20 (2), 213–226.

Güler, C., *et al.*, 2002. Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrology Journal*, 10, 455–474.

Hartley, H.O., 1950. The use of range in analysis of variance. *Biometrika*, 37, 271–280.

Helena, B., *et al.*, 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Research*, 34 (3), 807–816.

Hellar-Kihampa, H., *et al.*, 2013. Water quality assessment in the Pangani River basin, Tanzania: natural and anthropogenic influences on the concentrations of nutrients and inorganic ions. *International Journal of River Basin Management*, 11 (1), 55–75.

Hinkle, D.E., Wiersma, W., and Jurs, S.G., 2003. *Applied statistics for the behavioral sciences*. Boston, MA: Houghton Mifflin.

Huang, F., *et al.*, (2010). Spatial variation and source apportionment of water pollution in Qiantang river (China) using statistical techniques. *Water Research*, 44 (5), 1562–1572.

Juahir, H., *et al.*, 2011. Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. *Environmental Monitoring and Assessment*, 173 (1), 625–641.

Kempe, S., 1982. Long-term records of the C07 pressure fluctuations in fresh water. *Mediterranean Geology Paleaontology Institute*, 52, 291–332.

Kraiem, Z., *et al.*, 2014. Geochemical characteristics of arid shallow aquifers in Chott Djerid, south-western Tunisia. *Journal of Hydro-environment Research*, 8, 460–473.

Laaksoharju, M., Skarman, C., and Skarman, E., 1999. Multivariate mixing and mass balance (M3) calculations, a new tool for decoding hydrogeochemical information. *Applied Geochemistry*, 14, 861–871.

Landwehr, J.M. and Deininger, R.A. 1976. A comparison of several water quality indexes. *Journal Water Pollution Control Federation*, 48 (5), 954–958.

Loether, H.L. and McTavish D.G., 1988. *Descriptive and inferential statistics: an introduction*. 3rd ed. Boston, MA: Allyn & Bacon.

Mahtab, M., *et al.*, 2015. Surface and groundwater quality assessment based on multivariate statistical techniques in the vicinity of Mohanpur, Bangladesh. *International Journal of Environmental Health Engineering*, 4 (1), 1–9.

Martin, G., 1979. *The nitrogen problem in the waters*. Paris: Technical and documentation Lavoisier ed.

Mebarkia, A. 2012. *Studies of physico-chemical characteristics of surface waters. Case of the dam Ain Zada. North-eastern Algeria*. Magister Dissertation. University of Annaba, Algeria.

Meglin, R.R., 1991. Examining large databases: a chemometric approach using principle components analysis. *Journal of Chemometrics*, 5, 163–179.

Mendiguchıa, C., Moreno, C., and Garcıa-Vargas, M., 2007. Evaluation of natural and anthropogenic influences on the Guadalquivir River (Spain) by dissolved heavy metals and nutrients. *Chemosphere*, 69, 1509–1517.

MSDEWP (Ministry of Sustainable Development, Environment, Wildlife and Parks). 2013. *Criteria for water surface quality*. 3rd ed. Quebec: Directorate for Monitoring the Environment.

Mustapha, A., *et al.*, 2013. Riverwater quality assessment using environmentric techniques: case study of Jakara River Basin. *Environmental Science and Pollution Research*, 20 (8), 5630–5644.

Mustapha, A., *et al.*, 2014. Statistical approach in determining the spatial changes of surface water quality at the upper course of Kano River, Nigeria. *Water Quality, Exposure and Health*, 6 (3), 127–142.

Najah, A., *et al.*, 2009. Prediction of Johor River water quality parameters using artificial neural networks. *European Journal of Scientific Research*, 28 (3), 422–35.

O'Shea, B. and Jankowski, J., 2006. Detecting subtle hydrochemical anomalies with multivariate statistics: an example from 'homogeneous' groundwaters in the Great Artesian Basin, Australia. *Hydrology Process*, *20*, 4317–4333.

Pacini, N., *et al.*, 2013. Water-quality management in a vulnerable large river: the Nile in Egypt water-quality management in a vulnerable large river: the Nile in Egypt. *International Journal of River Basin Management*, 11 (2), 205–219.

Parmar, K.S. and Bhardwaj, R., 2014. Water quality management using statistical analysis and time-series prediction model. *Applied Water Science*, 4 (2014), 425–434.

Pradhan, B. and Pirasteh, S., 2011. Hydro-chemical analysis of the ground water of the basaltic catchments: upper Bhatsai region, Maharashtra. *Open Hydrology Journal*, 2011 (5), 51–57.

Said, A., Stevens, D.K., and Sehlke, G. 2004. Environmental assessment an innovative index for evaluating water quality in streams. *Environmental Management*, 34 (3), 406–414.

Sanchez Lopez, J.F., *et al*., 2004. Assessment of metal contamination in Doñana National park (Spain) using Crayfish (*Procamburus clarkii*). *Environmental Monitoring and Assessment*, 93, 17–29.

Sheela, A.M., *et al*., 2012. Water quality assessment of a tropical coastal lake system using multivariate cluster, principal component and factor analysis. *Lakes & Reservoirs: Research and Management*, 17, 143–159.

Shihab, A.S. and Abdul Baqi, Y.T., 2010. Multivariate analysis of ground water quality of Makhmor plain/north Iraq. *Damascus University Journal*, 26 (1), 19–26.

Shrestha, S. and Kazama, F., 2007. Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji River Basin, Japan. *Environmental Modelling & Software – Journal*, 22 (4), 464–475.

Singh, K.P., *et al*., 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India): a case study. *Water Research*, 38 (18), 3980–3992.

Sneath P.H.A. and Sokal, R.R., 1973. *Numerical taxonomy: the principles and practice of numerical classification*. San Francisco, CA: W.H. Freeman.

Tiri, A., Lahbari, N., and Boudoukha, A., 2016. Hydrochemical characterization of surface water in the Timgad watershed, East Algeria. *Desalination and Water Treatment*, 57 (2), 562–571.

Tsytsarin, Y.V., 1988. *Introduction to water chemistry*. Moscow: Moscow University.

Valeriani, F., *et al*., 2015. River water quality assessment: comparison between old and new indices in a real scenario from Italy. *International Journal of River Basin Management*, 13 (3), 325–331.

Varol, M., 2013. Dissolved heavy metal concentrations of the Kralkizi, Dicle and Batman dam reservoirs in the Tigris River Basin, Turkey. *Chemosphere*, 93 (6), 954–62.

Wang, X., *et al*., 2012. Evaluation of spatial and temporal variation in stream water quality by multivariate statistical techniques, a case study of the Xiangxi River Basin, China. *Quaternary International – Journal*, 282, 137–144.

Wang, Y.S., *et al*., 2006. Multivariate statistical analysis of water quality and phytoplankton characteristics in Daya Bay, China, from 1999 to 2002. *Oceanologia*, 48 (2), 193–213.

Wu, B., *et al*., 2009. Multivariate statistical studies of organic pollutants in Nanjing reach of Yangtze River. *Journal of Hazardous Materials*, 169 (1), 1093–1098.

Zarei, H. and Bilondi, M.P., 2013. Factor analysis of chemical composition in the Karoon River basin, southwest of Iran. *Applied Water Science*, 3, 753–761.