Provided for non-commercial research and educational use only.
Not for reproduction or distribution or commercial use.

This article was originally published by IWA Publishing. IWA Publishing recognizes the retention of the right by the author(s) to photocopy or make single electronic copies of the paper for their own personal use, including for their own classroom use, or the personal use of colleagues, provided the copies are not offered for sale and are not distributed in a systematic way outside of their employing institution.

Please note that you are not permitted to post the IWA Publishing PDF version of your paper on your own website or your institution's website or repository.

Please direct any queries regarding use or permissions to hydrology@iwap.co.uk

Water quality assessment using multivariate statistical techniques in Río Tercero Reservoir, Argentina

Matias Bonansea, Claudia Ledesma, Claudia Rodriguez and Lucio Pinotti

ABSTRACT

Water quality monitoring programs generate complex multidimensional data sets. In this study, multivariate statistical techniques were employed as an effective tool for the analysis and interpretation of these water quality data sets. Principal component analysis (PCA) and cluster analysis (CA) were applied to evaluate spatial and temporal variation of water quality in Río Tercero Reservoir (Argentina). Six sampling sites were surveyed each climatic season for 21 parameters during 2003–2010. The results revealed that PCA showed the existence of four significant principal components (PCs) which account for 96.7% of the total variance of the data set. The first PC was assigned to mineralization whereas the other PCs were built from variables indicative of pollution. Hierarchical CA grouped the six monitoring sites into three clusters and classified the different climatic seasons into two clusters based on similarities in water quality characteristics.

Key words | cluster analysis, monitoring, multivariate statistical techniques, principal components, surface water, water quality

Matias Bonansea (corresponding author)
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET),
Ecología,
Facultad de Agronomía y Veterinaria,
Universidad Nacional de Río Cuarto (UNRC),
Ruta Nacional 36 Km 601,
(5800) Río Cuarto, Córdoba,
Argentina
E-mail: *mbonansea@ayv.unrc.edu.ar*

Claudia Ledesma
Claudia Rodriguez
Departamento Ciencias Básicas,
Facultad de Agronomía y Veterinaria,
Universidad Nacional de Río Cuarto (UNRC),
Córdoba,
Argentina

Lucio Pinotti
CONICET – Departamento de Geología,
Universidad Nacional de Río Cuarto (UNRC),
Córdoba,
Argentina

INTRODUCTION

Water quality is a matter of serious concern today, being considered the main factor controlling health and the state of disease in both humans and animals. Anthropogenic influences (urban, industrial and agricultural activities increasing consumption of water resources) as well as natural processes (change in precipitation inputs, erosion, weathering in crustal material) degrade surface waters and impair their use for drinking, industrial, agricultural, recreation or other purposes (Simeonov *et al.* 2003; Li *et al.* 2009).

Because lakes, reservoirs and rivers constitute the main inland water resources for different purposes, it is imperative to prevent and control water pollution and to have reliable information on water quality (Varol *et al.* 2012; Bonansea & Fernandez 2013). Therefore, regular monitoring programs are required for representative and reliable estimates of surface water quality (Kazi *et al.* 2009; Fernandez *et al.* 2012). However, due to spatial and temporal variations in water quality, such programs need to include a large number of physico-chemical parameters taken at different times and from many sites involving huge financial inputs and resulting in a large and complex data matrix which is often difficult to interpret towards drawing meaningful conclusions. Thus, there is a need to optimize the monitoring networks, reducing the number of water quality parameters, sampling sites and monitoring periods to representative ones without losing useful information (Singh *et al.* 2004).

Multivariate statistical techniques are presented as an effective tool for the interpretation of this complex data matrix, allowing a better understanding of water quality and ecological status of the studied reservoir (Wunderlin *et al.* 2001; Shrestha & Kazama 2007). The application of different multivariate statistical approaches, such as principal component analysis (PCA), factor analysis (FA), discriminate analysis and cluster analysis (CA) have been widely used to achieve great efficiency of data compression from the original data and to interpret natural associations between samples or variables, highlighting the information which is not available at first glance (Chen *et al.* 2007).

doi: 10.2166/nh.2014.174

Furthermore, the multivariate treatment of environmental data is widely successfully used to identify possible pollution sources that influence water systems and offers a valuable technique for reliable management of water resources as well as rapid solutions on pollution problems (De Bartolomeo *et al*. 2004; Kalamaras *et al*. 2010; Wang *et al*. 2012; Ledesma *et al*. 2013).

The objective of the present study was to analyse the spatial and temporal patterns of water quality of Rio Tercero Reservoir (Argentina). The large data matrix obtained during an 8-year (2003–2010) monitoring program was subjected to the PCA and CA multivariate techniques to obtain information about the similarities and dissimilarities between sampling sites and seasons, and to identify water quality variables responsible for variations in surface water quality. Additionally, the influence of the pollution sources on the water quality parameters was determined. Thus, we illustrated the usefulness of the multivariate statistical analysis to improve the understanding of the surface water system.

METHODOLOGY

Study area

Río Tercero Reservoir (32° 11′ S, 64° 23′ W) is located in the province of Córdoba, Argentina (Figure 1). This reservoir,

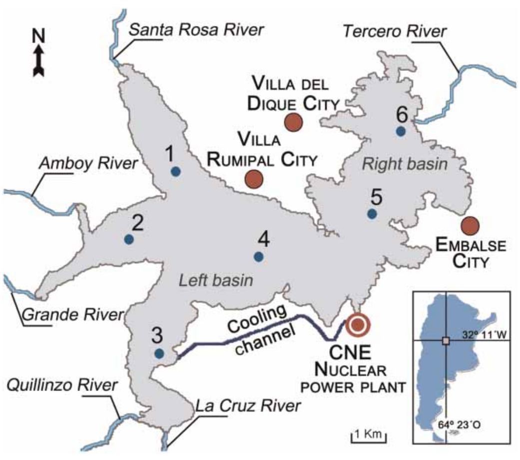

Figure 1 | Location of Río Tercero Reservoir in the central-west of Argentina and position of sampling sites.

which is the largest artificial reservoir in the province, has an area of 46 km^2, a volume of 10^7 m^3 and maximum and mean depths of 46.5 and 12.2 m, respectively (Mariazzi *et al*. 1992). The watershed area is 3,300 km^2 and rainfall is strongly seasonal, with dry winters and heavy rains during spring and summer (MacDonagh *et al*. 2009).

Río Tercero Reservoir is divided into two basins linked by a passage. The left basin has three branches where tributary rivers flow. The right basin supplies water to the Tercero River.

The reservoir has multiple purposes, such as water supply for three cities located on its shores (Embalse, Villa del Dique and Villa Rumipal with a combined population of 20,000 inhabitants approximately), recreational activities, flood control, irrigation and power generation. In 1986 a nuclear power plant (Central Nuclear Embalse, (CNE)) was installed. Water for cooling the nuclear reactor is taken from the middle section of the reservoir. In the cooling system, water temperature (WT) increases 7.0 °C and is returned to the reservoir by a 5-km-long open-sky channel (Mariazzi *et al*. 1992; Lamaro *et al*. 2013).

Sampling and analytical procedures

Water sampling was conducted once during each climatic season (defined in the southern hemisphere as winter: July–September; spring: October–December; summer: January–March; and fall: April–June) from winter 2003 to spring 2010. Six sampling sites were surveyed (Figure 1). The locations of these sites, which were selected in order to reasonably represent the water quality of the reservoir, were recorded using a GPS device. Water samples were collected at 20 cm depth. Preservation and transportation of the water samples to the laboratory were performed according to *Standard Methods* (APHA-AWWA-WEF 2000). In total, the samples were analysed for 21 parameters. *In situ*, WT, pH, dissolved oxygen (DO) and electrical conductivity (EC) were measured using portable electronic instruments. Secchi disk transparency (SDT) was measured using a standard 20 cm diameter Secchi disk with alternating black and white quadrants. In the laboratory, chlorophyll-a concentration (Chl-a), total dissolved solids (TDS), total phosphorus (TP), total nitrogen (TN), total Kjeldahl nitrogen (TKN), nitrate nitrogen (NO_3-N), total hardness (T-Hard), total alkalinity (T-Alk), bicarbonates (HCO_3^-), sulphate (SO_4), chloride (Cl), sodium (Na), potassium (K), calcium (Ca), magnesium (Mg) and fluoride (F) were determined following standard analytical methods (APHA-AWWA-WEF 2000; WHO 2006).

Statistical analysis

All data analyses were performed using the InfoStat software package (Di Rienzo *et al.* 2009). One-way analysis of variance (ANOVA) was employed as a first approach to analyse the significant spatial and temporal differences of water quality variables ($p < 0.05$; least-significance difference, LSD test). To find relationships among these variables, a correlation matrix using the Pearson's correlation coefficient (r) was carried out with statistical significance set at $p < 0.05$.

Multivariate statistical analysis can help to simplify and organize large data sets to provide meaningful insight (Chen *et al.* 2007). In this study, two multivariate methods were applied to the water quality data set: the PCA and the CA. This data set was previously standardized using z-scale transformation in order to avoid misclassifications arising from the different orders of magnitude of both numerical value and variance of the parameters analysed (Wunderlin *et al.* 2001; Singh *et al.* 2004). According to Güler *et al.* (2002) and Cloutier *et al.* (2008), standardization is commonly used in multivariate statistical analysis.

PCA technique

The PCA is a data transformation technique used to find out associations between variables, thus reducing the dimensionality of the data set (Helena *et al.* 2000; Chen & Hong 2012). This technique extracts the eigenvalues and eigenvectors from the covariance matrix of original variables. The principal components (PCs) are the new uncorrelated (orthogonal) variables which are obtained by multiplying the original correlated variables with the eigenvector, which is a list of coefficients (loadings or weightings). Thus, the PCs are weighted linear combinations of the original variables (Varol *et al.* 2012). There are as many PCs as original variables. However, the first principal component loading explains the most variance and each subsequent component explains progressively less. As a result, a small number of factors usually account for approximately the same amount of information as the much larger set of the original observations do (Chen *et al.* 2007). Thus, PC provides information on the most meaningful parameters, which describe the whole data set affording data reduction with minimal loss of information. The PC loadings can be examined to provide further insight into the processes that are responsible for the similarities in the trace element concentrations in the water samples (Helena *et al.* 2000; Wunderlin *et al.* 2001; Farnham *et al.* 2003; Singh *et al.* 2004). PCA was applied to extract significant PCs summarizing the statistical correlation among components in the water samples.

CA technique

CA is an unsupervised pattern recognition technique that uncovers intrinsic structure or underlying behaviour of a data set without making a priori assumption about the data, in order to classify the objects of the system into categories or clusters based on their nearness or similarity (Vega *et al.* 1998). This technique classifies objects, so that each object is similar to the others in the cluster with respect to a predetermined selection criterion. The resulting clusters should then exhibit high internal (within-cluster) homogeneity and high external (between clusters) heterogeneity (Shrestha & Kazama 2007). Hierarchical clustering is the most common approach in which clusters are formed sequentially, by starting with the most similar pair of objects and forming higher clusters step-by-step. The result can be displayed as a dendrogram (tree diagram), which provides a visual summary of the clustering process, presenting a picture of the groups and their proximity, with a dramatic reduction in dimensionality of the original data (Wunderlin *et al.* 2001). The Euclidean distance usually gives the similarity between two samples and a distance can be represented by the difference between analytical values from the samples (Otto 1998). In this study, hierarchical CA was performed using the Ward's method and Euclidean distances as a measure of similarity. The Ward's method uses an analysis of variance approach to evaluate the distances between clusters, attempting to minimize the sum of squares of any two clusters that can be formed at each step. CA was applied to the water quality data set to group the similar sampling sites (spatial variability) and seasonal (temporal) similarity among the samples, resulting in spatial and temporal dendrograms. The linkage distance is reported as Dlink/Dmax, which represents the quotient of the linkage distances for a particular case divided by the maximal distance, multiplied by 100, as a way to standardize the linkage distance represented on the y-axis (Varol *et al.* 2012). Finally, ANOVA was performed ($p < 0.05$) to analyse the significant differences between the clusters obtained by the spatial CA. Whereas, as the temporal CA generated only two clusters, differences between clusters were tested for statistical differences using a Student T-test ($p < 0.05$).

RESULTS AND DISCUSSION

Chemistry of reservoir water

The basic statistics of the 21 water quality parameters collected from six samples sites during the study period are presented in Table 1. The results, which were based on 180 total water samples, were compared with the maximum limits recommended by the World Health Organization (WHO 2006), revealing that the concentrations of all analysed parameters were below the prescribed maximum limits by WHO guidelines for drinking water. This confirmed that all water samples collected in Río Tercero Reservoir corresponded to low polluted regions.

Most water quality parameters except WT and SDT did not show significant spatial variations ($p > 0.05$). Higher values of WT and lower values of SDT were recorded in site 3, which was related to the cooling channel of CNE and rivers influxes. The lowest values of SDT registered in the left basin can be explained by river inputs which provide greatest loads of suspended materials and dissolved solids that decrease the penetration of light. In contrast, sites 5 and 6, located in the right basin of the reservoir where deeper waters and sedimentation processes were more active, exhibited the highest SDT values. Similar results were found by Giardino *et al.* (2001) and Guan *et al.* (2011) studying different water bodies. On the other hand, WT, pH, EC, TP, T-Hard, T-Alk, Cl, K and Mg showed significant temporal differences ($p < 0.05$). Among these, WT and TP showed higher values in summer and spring, while pH, EC, T-Hard, T-Alk, Cl, K and Mg showed higher values in fall and winter.

Correlation between variables

The Pearson correlation matrix of the 21 analysed variables was calculated (Table 2). Only those correlation values higher than 0.5 were considered. High and positive correlations were observed between EC, TDS, T-Hard, T-Alk, HCO_3^-, SO_4, Cl, Na, K, Ca, Mg and F ($r = 0.502$ to 0.998). Most of these associations are responsible for water mineralization, being directly related to hydrochemical characteristics of the zone. Chl-a was positively correlated with PT, TN and TKN, indicating that these variables are responsible for phytoplankton growth.

Data structure determination and source identification

PCA was performed on the normalized data to compare the compositional pattern between the water samples and to identify the factors influencing each one. Table 3 summarized the PCA results including the loadings and the eigenvalues of each PC. There are several criteria to identify the number of PCs to be retained in order to understand the underlying data structure (Chen *et al.* 2007). In the present study, all the PCs with eigenvalues equal to or greater than 1.0 were used. Four PCs were retained explaining about 96.7% of the total variance in the water quality data set. The first component (PC1), which accounted for over 54.8% of the total variance, had positive loadings for EC, TDS, HCO_3^-, SO_4, Cl, Na, K, Ca, Mg, F, T-Hard and T-Alk. This PC can be interpreted as the mineral component of the surface water of the reservoir, being related to natural sources of the ionic groups of salts in the basin from inflows, soil weathering and runoff. Some of these results agree with those obtained with the Pearson test (Table 2). The high level of major cations (Na, Ca and Mg) and concentration of major anions (Cl, F, CHO_3^- and SO_4) in reservoir water increased EC and TDS, being consistent with other studies (Reghunath *et al.* 2002; Kazi *et al.* 2009). Furthermore, results indicated that EC and major ions were higher in winter and lower in summer, due to the accumulation on

Table 1 | Range and mean values of water quality parameters at different sampling sites of Río Tercero Reservoir

			Sample sites					
Parameters	**WHO limits**		**1**	**2**	**3**	**4**	**5**	**6**
WT (°C)	–	Mean	19.0	20.4	23.5	19.3	18.4	18.3
		Range	12.0–26.5	10.8–29.0	16.0–31.0	12.1–28.0	11.5–26.0	12.5–26.0
pH	6.5–8.5	Mean	7.87	7.68	7.67	7.93	7.74	7.82
		Range	7.04–9.15	6.83–8.37	6.82–8.76	6.98–9.90	6.95–8.80	6.63–9.14
DO (mg/L)	–	Mean	8.3	8.4	7.8	8.7	8.7	8.7
		Range	5.0–11.1	6.7–11.5	4.5–10.3	5.0–12.1	4.7–12.6	5.0–12.3
EC (μS/cm)	1500	Mean	251	254	305	272	244	255
		Range	127–580	126–580	101–800	117–580	127–580	128–580
SDT (m)	–	Mean	2.1	2.0	1.4	2.6	3.4	3.5
		Range	1.3–3.0	0.7–4.0	0.6–2.8	2.0–3.4	2.0–5.3	1.8–6.1
Chl-a (μg/L)		Mean	13.5	20.9	10.3	9.2	11.2	13.3
		Range	1.4–72.0	0.1–133.5	0.4–31.0	0.6–25.4	0.3–56.8	0.8–68.2
TDS (mg/L)	1000.0	Mean	141.6	146.5	194.0	142.8	139.1	137.5
		Range	111–183	108–179.0	92–560	110–184	101–180	102–184
TP (μg/L)	–	Mean	26.3	26.7	29.1	30.3	23.6	24.2
		Range	16.0–56.3	16.0–51.0	16.0–39.0	16.3–48.0	16.0–51.0	16.0–54.0
TN (μg/L)	–	Mean	1116.7	1280.0	1377.8	1400.0	980.0	1200.0
		Range	700–1500	800–2300	500–2300	600–2300	500–1600	600–3400
TKN (μg/L)	–	Mean	1080.0	1122.2	1307.5	1300.0	911.1	1263.6
		Range	600–1500	500–2500	500–2300	700–2200	500–1400	500–3400
NO_3-N (μg/L)	50,000.0	Mean	2.4	2.9	3.8	3.5	3.4	3.4
		Range	0.0–5.7	0.0–7.2	0.0–11.5	0.0–8.9	0.0–8.5	0.0–9.0
T-Hard (mg/L)	500.0	Mean	1.2	1.3	1.6	1.2	1.2	1.2
		Range	1.0–1.4	0.9–2.3	0.8–4.4	0.9–1.4	0.9–1.4	1.0–1.4
T-Alk (mg/L)	200.0	Mean	1.6	1.4	1.6	1.4	1.4	1.4
		Range	1.3–1.9	1.0–1.9	0.8–4.5	1.0–1.8	1.0–1.9	1.0–1.9
HCO_3^- (mg/L)	–	Mean	97.5	87.5	101.5	90.0	88.3	88.8
		Range	82.5–116.3	65.0–120.0	47.5–282.5	65.0–115.0	65.0–117.5	62.5–117.5
SO_4 (mg/L)	250.0	Mean	18.2	18.2	39.8	19.0	18.7	16.3
		Range	7.5–26.4	10.6–27.5	11.8–190.1	11.0–27.9	10.6–25.2	6.3–22.8

(*continued*)

Table 1 | continued

Parameters	WHO limits		Sample sites 1	2	3	4	5	6
Cl (mg/L)	250.0	Mean	7.6	8.2	8.7	6.6	7.0	6.6
		Range	4.3–14.3	2.9–14.3	2.9–18.6	2.9–14.3	2.9–11.4	2.9–11.4
Na (mg/L)	200.0	Mean	14.2	13.2	22.4	14.3	12.8	13.2
		Range	10.7–17.7	9.8–16.7	5.9–74.8	10.1–21.4	9.1–18.2	10.1–19.2
K (mg/L)	12.0	Mean	3.1	2.8	3.4	2.8	2.7	2.7
		Range	2.6–3.7	2.2–3.5	1.8–10.0	2.1–3.7	2.1–3.7	2.1–4.0
Ca (mg/L)	100.0	Mean	16.5	16.6	21.9	17.1	16.2	16.4
		Range	13.6–18.8	12.0–20.8	10.4–55.6	14.4–19.2	13.6–19.6	13.6–20.0
Mg (mg/L)	50.0	Mean	4.6	4.3	6.0	4.4	4.4	4.4
		Range	3.7–6.3	2.4–6.8	2.9–19.8	2.4–6.8	2.4–5.4	3.4–5.9
F (mg/L)	1.5	Mean	0.3	0.3	0.5	0.4	0.3	0.3
		Range	0.2–0.4	0.1–0.4	0.2–1.0	0.2–1.0	0.1–0.4	0.1–0.4

these ions in the deeper water during summer stagnation (Varol *et al.* 2012).

The second component (PC2) explained 21.2% of the total variance and was characterized by loadings of Chl-a, TP, NO_3-N, TN and TKN. This PC represents the effects of stream input, agricultural runoff and erosion from upland areas of the basin. The results reveal that nutrients were highest in summer and lowest in winter according to rainfall recorded in the region. Córdoba is a province where farming and cattle raising are common activities. Thus, fertilizers are widely used and contribute to high levels of soil nitrogen and phosphorus that can be incorporated into the reservoir by streams and runoff after rainfall, increasing eutrophication processes which often lead to a reduction in the supply of ecosystem services (Hein 2006; Granéli *et al.* 2008; Saaremäe *et al.* 2013). PC2 also showed the effect of nutrients on Chl-a, which is one of the most important parameters since it is used to estimate the primary productivity of water ecosystems (Dall'Olmo *et al.* 2005). This parameter registered the highest values in spring of 2009 (133.5 μg/L) agreeing with Mancini *et al.* (2010) who detected a bloom of *Anabaena spiroides* y *Microcystis aeruginosa* in Piedras Moras Reservoir, located 11 km downstream of Río Tercero Reservoir. These harmful cyanobacterial blooms, which have increased worldwide in fresh and marine waters, are a primary concern for water management and may cause a variety of water quality problems, including fish kills, aesthetic nuisances (e.g., odours, scums, unsightliness), and unpalatable and possibly unsafe drinking water (Torbick *et al.* 2008). Such problems can severely limit aquatic habitat quality, recreational activities, fisheries and use of a water body as a potable water resource (Vincent *et al.* 2004). Subsequent nutrient abatement strategies implemented by environmental agencies or water authorities could be largely successful resulting in a reduction in algal biomass and greater reservoir transparency.

The third component (PC3) showed 12.6% of the total variance and had positive loadings on WT, Chl-a and TP, whereas PC4, explaining the lowest variance (8.0%), had positive loadings on pH, NO_3^-N, TN and TKN and negative loading on DO and SDT. Both these PCs point to sources of nutrient of the reservoir being related to stream inputs, agricultural runoff, soil erosion, municipal effluents and

Table 2 | Pearson correlation matrix of the 21 variables determined

	WT	pH	DO	SDT	EC	Chl-a	TP	NO3-N	TN	TKN	TDS	CHO3-	SO4	Cl	Na	K	Ca	Mg	F	T-Hard	T-Alk
WT	1																				
pH	–0.132	1																			
DO	–0.156	–0.086	1																		
SDT	–0.204	–0.024	–0.089	1																	
EC	–0.186	0.274[b]	–0.135	0.133	1																
Chl-a	0.100	0.173	0.025	–0.133	–0.195	1															
TP	0.450[a]	–0.117	–0.206	–0.125	0.039	**0.683**[b]	1														
NO_3-N	0.131	–0.046	0.238	0.115	0.383[b]	–0.112	0.135	1													
TN	0.214	–0.018	–0.185	0.040	–0.431[b]	**0.573**[b]	0.044	0.170	1												
TKN	0.166	–0.185	–0.120	0.030	–0.366	**0.512**[b]	0.085	0.130	**0.963**[a]	1											
TDS	0.074	0.251[b]	–0.213	–0.292[b]	**0.604**[a]	–0.109	0.107	–0.033	0.082	–0.002	1										
CHO_3^-	0.027	0.261[b]	–0.166	–0.262[b]	0.410[a]	–0.088	–0.127	–0.194	0.070	0.001	**0.877**[a]	1									
SO4	0.037	0.207	–0.197	–0.220	**0.546**[a]	–0.111	0.022	–0.056	–0.017	–0.122	**0.931**[a]	**0.874**[a]	1								
Cl	0.151	0.261[b]	–0.316[b]	–0.231	0.321[b]	0.140	–0.165	–0.185	0.158	0.108	**0.582**[a]	**0.373**[a]	0.409[a]	1							
Na	–0.031	0.348[a]	–0.181	–0.338[a]	**0.614**[a]	–0.097	0.042	0.019	0.064	–0.030	**0.965**[a]	**0.861**[a]	**0.925**[a]	**0.544**[a]	1						
K	0.096	0.211	–0.211	–0.332[b]	0.436[a]	–0.071	–0.157	–0.128	0.060	0.025	**0.916**[a]	**0.934**[a]	**0.892**[a]	**0.525**[a]	**0.890**[a]	1					
Ca	0.007	0.334[a]	–0.236	–0.343[a]	**0.502**[a]	–0.039	0.061	–0.066	0.178	0.073	**0.939**[a]	**0.899**[a]	**0.935**[a]	0.483[a]	**0.943**[a]	**0.920**[a]	1				
Mg	–0.030	0.389[a]	–0.276[b]	–0.210	0.474[a]	–0.014	–0.222	–0.258[b]	0.004	–0.127	**0.884**[a]	**0.861**[a]	**0.905**[a]	**0.551**[a]	**0.864**[a]	**0.883**[a]	**0.897**[a]	1			
F	–0.178	0.209	0.029	–0.275[b]	0.201	0.219	–0.194	0.055	0.040	–0.043	0.270[b]	0.120	0.241[b]	0.250[b]	0.346[a]	0.190	0.313[a]	0.200	1		
T-Hard	0.001	0.382[a]	–0.262[b]	–0.276[b]	**0.601**[a]	–0.064	–0.026	–0.126	0.119	–0.001	**0.899**[a]	**0.852**[a]	**0.903**[a]	**0.510**[a]	**0.893**[a]	**0.870**[a]	**0.924**[a]	**0.904**[a]	0.253[b]	1	
T-Alk	0.032	0.258[b]	–0.175	–0.283[b]	0.430[a]	–0.100	–0.135	–0.191	0.060	–0.011	**0.876**[a]	**0.998**[a]	**0.875**[a]	0.369[a]	**0.860**[a]	**0.931**[a]	**0.898**[a]	**0.859**[a]	0.118	**0.854**[a]	1

[a]Correlation is significant at the 0.01 level (2-tailed).
[b]Correlation is significant at the 0.05 level.
Bold indicates correlation values higher than 0.500.

Table 3 | Loadings of experimental variables (21) on the first four PCs for complete data set

Parameters	PC1	PC2	PC3	PC4
WT	0.07	– 0.23	0.49	0.14
pH	– 0.13	0.16	– 0.15	0.38
DO	– 0.07	0.18	– 0.35	0.10
EC	0.27	0.17	0.07	0.11
SDT	– 0.13	0.19	– 0.11	– 0.23
Chl-a	– 0.08	0.32	0.17	0.16
TDS	0.29	0.08	– 0.01	– 0.09
TP	0.02	0.30	0.42	– 0.02
TN	– 0.03	0.28	0.17	0.48
TKN	– 0.02	0.34	0.08	0.53
NO_3-N	– 0.01	0.43	0.15	0.23
T-Hard	0.28	0.02	0.15	– 0.02
T-Alk	0.26	– 0.07	– 0.15	0.08
HCO_3^-	0.26	– 0.07	– 0.18	0.11
SO_4	0.29	0.06	0.04	– 0.13
Cl	0.23	– 0.15	0.20	– 0.05
Na	0.29	0.08	– 0.01	– 0.02
K	0.28	– 0.04	– 0.16	0.05
Ca	0.29	0.09	0.05	– 0.04
Mg	0.29	0.06	– 0.06	– 0.04
F	0.27	0.15	– 0.01	0.07
Eigenvalue	11.514	4.456	2.656	1.675
% Total variance	54.8	21.2	12.6	8.0
Cumulative % variance	54.8	76.0	88.7	96.7

atmospheric deposition. The contribution of WT to PC3 can be considered to be a result of the constant input of heat from the cooling channel of CNE nuclear power plant. Thus, WT, which showed a very characteristic annual cycle with highest values in summer and lowest in winter, showed the highest values in site 3 and lowest in the right basin of the reservoir. From the temperature difference, changes are expected in the DO. According to Shrestha & Kazama (2007), the inverse relationship between WT and DO is a natural process because warmer water becomes more easily saturated with oxygen. This presumption is strengthened by the fact that site 3 showed the lowest DO concentrations. The monitoring of oxygen concentration in an aquatic system is an important subject, as the biological, chemical and physical processes involved in the increase or decrease of oxygen in a lake are so numerous and complex that there is no model that can be used without a careful analysis of local characteristics (Kazi *et al*. 2009).

The results from the PCA suggested that most of the variation in water quality was explained by a set of soluble salts (natural). A substantial contribution came from the nutrient group of pollutants (point sources such as municipal effluents and non-point such as erosion and agricultural runoff) and organic pollutants (anthropogenic). The rest of the contributions arose from physical parameters (natural and the effect of nuclear power plant on WT). PCA served as a means to identify those parameters that had the greatest contribution to variation in water quality and suggested possible sets of pollution sources in the watershed.

Spatial and temporal similarity

CA was used to detect the similarity groups between the sampling sites and the climatic seasons. The spatial CA rendered a dendrogram where the six sampling sites were

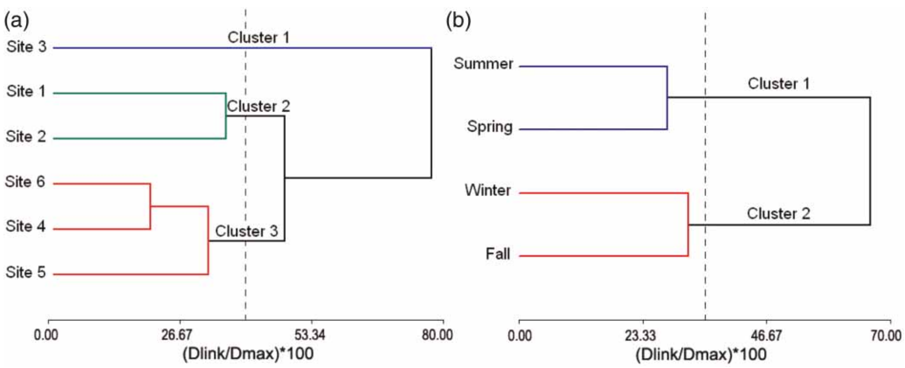

Figure 2 | Dendrogram showing hierarchical clustering of (a) sampling sites and (b) climatic seasons. In both cases the CA was performed according to Ward's method with Euclidean distance.

grouped into three statistically significant clusters at $(\text{Dlink}/\text{Dmax}) \times 100 < 50$ (Figure 2(a)), whereas the temporal dendrogram grouped the four climatic seasons into two clusters (Figure 2(b)). As we expected, in both cases the clustering procedure generated groups in a very convincing way, as the elements in these groups had similar characteristic features and natural background source types.

The spatial dendrogram, which represents geographical locations of sampling sites, generated clusters that correspond to regions of relatively high (cluster 1), moderate (cluster 2) and low pollution (cluster 3). The term 'relatively polluted' was used because, as previously mentioned, the parameters measured were always below the prescribed maximum limits by WHO guidelines for drinking water (WHO 2006). Cluster 1 (site 3) receives polluted effluents from different sources such as stream inputs, agricultural runoff and soil erosion by the Quillinzo and La Cruz rivers as shown in Figure 1. This cluster is also directly influenced by the effluent of CNE which, as we have previously mentioned, raises the temperature of water. Sites 1 and 2 make one group as cluster 2. Although this cluster receives effluents by the Grande, Amboy and Santa Rosa rivers, their watersheds are related to areas of natural vegetation and low anthropic activities, corresponding to a relatively moderate polluted region. However, in the last years these watersheds have reported an increase in agricultural activities and a decrease in native forest. Hence it is possible that in the near future this cluster will be more related to cluster 1. To fully understand water chemistry, a study of variability in river discharge patterns needs to be considered for an optimal management plan. Cluster 3 (sites 4 to 6) was generated by low relatively polluted sampling sites located in the central region of the reservoir and in the right basin of the dam, where general water quality conditions were better. The results from one-way ANOVA used to compare these clusters showed that WT, TDS, SO_4, Na, Ca, F and T-Hard of cluster 1 were significantly higher ($p < 0.05$) than clusters 2 and 3 (Table 4), whereas SDT was significantly higher in cluster 3 than in the rest of the groups.

The temporal dendrogram obtained by CA generated two clusters. Cluster 1 corresponded to the wet season including spring and summer. Fall and winter were grouped in cluster 2 corresponding to the dry season. Although we could expect this group combination, CA technique was able to corroborate this assumption. The results from Student's T-test showed significant difference ($p < 0.05$) in WT, pH, EC, TDS, Cl, Na, K, Ca, Mg T-Hard and T-Alk. Except for WT, these parameters were higher in the dry season.

The results obtained with the spatial CA suggested that for rapid assessment of water quality, only one site in each cluster may serve as a good spatial assessment of the water quality of the whole network. While the temporal CA implied that sampling during only two seasons (wet and dry) in a year may suffice for assessment of temporal variations in water quality. According to Singh *et al.* (2004) and Kazi *et al.* (2009), the CA technique is also useful in offering reliable classification of surface waters in the whole region and will make it possible to design a future sampling strategy in an optimal manner. Thus, sampling frequency, number of sites and cost in the monitoring network will be reduced without missing much information. The same aspects have been successfully applied in different water quality programmes (Wunderlin *et al.* 2001; Kim *et al.* 2005; Varol *et al.* 2012). The use of similar methodology in other reservoirs, either natural lakes or artificial dams, could be of great interest for further investigations.

CONCLUSIONS

Water quality monitoring programs generate complex multidimensional data sets that need multivariate statistical treatment for their analysis and interpretation. In this paper, different multivariate statistical techniques were used to evaluate temporal and spatial variations in surface water quality of Río Tercero Reservoir. PCA was used to identify the factors or sources responsible for water quality variations. The main parameters responsible for water quality variation are mainly related to soluble salts (natural), organic pollution and nutrients (point and non-point), and physical parameters (natural). The nuclear power plant was also among the major factors responsible for reservoir water quality. Based on similarities in water quality characteristics, hierarchical CA grouped the six sampling sites into three clusters (relative high, moderate and low pollution) and classified climatic seasons in two clusters (wet and dry season). This technique could also facilitate the design of an optimal future monitoring strategy in the reservoir,

Table 4 | Mean values with standard error (S.E.) for water quality parameters in clusters of the sampling sites and climatic seasons

	Spatial CA						Temporal CA			
	Cluster 1		Cluster 2		Cluster 3		Cluster 1		Cluster 2	
	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
WT	23.5 (a)	1.2	19.7 (b)	0.9	18.7 (b)	0.7	21.3 (a)	0.8	18.7 (b)	0.6
pH	7.67	0.15	7.78	0.09	7.83	0.08	7.51 (a)	0.07	7.99 (b)	0.08
DO	7.8	0.4	8.3	0.3	8.7	0.2	8.7	0.3	8.4	0.2
EC	305	85	253	39	256	30	165 (a)	6	335 (b)	36
SDT	1.4 (a)	0.2	2.0 (b)	0.1	3.2 (c)	0.1	2.7	0.2	2.4	0.2
Chl-a	10.3	2.8	17.2	5.2	11.4	1.9	10.7	2.5	15.2	3.1
TDS	194.0 (a)	41.9	144.1 (b)	5.1	139.7 (b)	3.5	128.9 (a)	3.0	166.3 (b)	14.2
TP	29.1	2.6	26.5	2.8	25.5	1.8	27.8	2.4	22.7	1.4
TN	1377.8	189.9	1218.8	134.9	1193.3	142.8	1333.3	94.5	1287	171.3
TKN	1307.5	189.4	1107.1	132.4	1158.2	137.1	1213.3	82.7	1194.1	198.8
NO_3-N	3.8	1.1	2.7	0.6	3.5	0.5	3.5	0.5	3.1	0.5
T-Hard	1.6 (a)	0.3	1.3 (ab)	0.1	1.2 (b)	0.1	1.1 (a)	0.1	1.4 (b)	0.1
T-Alk	1.6	0.3	1.5	0.1	1.4	0.1	1.3 (a)	0.1	1.6 (b)	0.1
HCO_3^-	101.5	19.5	91.2	3.8	89.0	2.7	83.8	3.2	98.8	6.5
SO_4	39.8 (a)	15.2	18.2 (b)	1.4	17.9 (b)	0.9	17.2	0.9	26	5.3
Cl	8.7	1.3	8.0	0.9	6.8	0.5	5.7(a)	0.4	9.0 (b)	0.6
Na	22.4 (a)	6.2	13.5 (b)	0.5	13.4 (b)	0.5	11.7 (a)	0.4	18.0 (b)	2.1
K	3.4	0.7	2.9	0.1	2.8	0.1	2.6 (a)	0.1	3.2 (b)	0.2
Ca	21.9 (a)	3.7	16.6 (b)	0.5	16.5 (b)	0.3	15.6 (a)	0.4	19.2 (b)	1.2
Mg	6.0	1.4	4.4	0.3	4.4	0.1	4.0 (a)	0.1	5.3 (b)	0.5
F	0.5 (a)	0.1	0.3 (b)	0.1	0.3 (b)	0.1	0.3	0.1	0.3	0.1

Spatial CA: cluster 1 (site 3), cluster 2 (sites 1 and 2), cluster 3 (sites 4 to 6).
Temporal CA: cluster 1 (summer and spring), cluster 2 (fall and winter).
The different letters indicate statistical difference among clusters ($p < 0.05$).

which could reduce the number of sampling sites, monitoring frequency and associated costs. The multivariate statistical techniques used in this study served as an excellent tool for the analysis and interpretation of complex data sets, identifying pollution sources and understanding temporal and spatial variations in water quality for effective water quality management.

ACKNOWLEDGEMENT

The authors thank the reviewers for their useful comments and helpful suggestions to improve the paper.

REFERENCES

APHA-AWWA-WEF (American Public Health Association – American Water Works Association – Water Environment Federation) 2000 *Standard Methods for the Examination of Water and Wastewater*, 18th edn. APHA-AWWA-WEF, Washington, DC.

Bonansea, M. & Fernandez, R. L. 2013 Remote sensing of suspended solid concentration in a reservoir with frequents wildland fires on its watershed. *Water Sci. Technol.* **67** (1), 217–223.

Chen, L. & Hong, Y. 2012 Regional Taiwan rainfall frequency analysis using principal component analysis, self-organizing maps and L-moments. *Hydrol. Res.* **43** (3), 275–285.

Chen, K., Jiao, J. J., Huang, J. & Huang, R. 2007 Multivariate statistical evaluation of trace elements in groundwater in a coastal area in Shenzhen, China. *Environ. Pollut.* **147** (3), 771–780.

Cloutier, V., Lefebvre, R., Therrien, R. & Savard, M. M. 2008 Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system. *J. Hydrol.* **353**, 294–313.

Dall'Olmo, G., Gitelson, A., Rundquist, D., Leavitt, B., Barrow, T. & Holz, J. 2005 Assessing the potential of SeaWiFS and MODIS for estimating chlorophyll concentration in turbid productive waters using red and near-infrared bands. *Remote Sens. Environ.* **96** (2), 176–187.

De Bartolomeo, A., Poletti, L., Sanchini, G., Sebastiani, B. & Morozzi, G. 2004 Relationship among parameters of lake polluted sediments evaluated by multivariate statistical analysis. *Chemosphere* **55** (10), 1323–1329.

Di Rienzo, J. A., Casanoves, F., Balzarini, M., Gonzalez, L., Tablada, M. & Robledo, C. W. 2009 *InfoStat version 2009*. InfoStat Group, FCA, National University of Córdoba, Argentina.

Farnham, I. M., Johannesson, K. H., Singh, A. K., Hodge, V. F. & Stetzenbach, K. J. 2003 Factor analytical approaches for evaluating groundwater trace element chemistry data. *Anal. Chim. Acta* **490**, 123–138.

Fernandez, R. L., Bonansea, M., Cosavella, A., Monarde, F. & Bresciano, J. 2012 Effects of bubbling operations on a thermally stratified reservoir: Implications for water quality amelioration. *Water Sci. Technol.* **66** (12), 2722–2730.

Giardino, C., Pepe, M., Brivio, P., Ghezzi, P. & Zilioli, E. 2001 Detecting chlorophyll, Secchi disk depth and surface temperature in a sub-alpine lake using Landsat imagery. *Sci. Total Environ.* **268** (1–3), 19–29.

Granéli, E., Weberg, M. & Salomon, P. 2008 Harmful algal blooms of allelopathic microalgal species: The role of eutrophication. *Harmful Algae* **8** (1), 94–102.

Guan, X., Li, J. & Booty, W. G. 2011 Monitoring lake Simcoe water clarity using Landsat-5 TM images. *Water Resour. Manag.* **25** (8), 2015–2033.

Güler, C., Thyne, G. D., McCray, J. E. & Turner, A. K. 2002 Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeol. J.* **10**, 455–474.

Hein, L. 2006 Cost-efficient eutrophication control in a shallow lake ecosystem subject to two steady states. *Ecol. Econ.* **59** (4), 429–439.

Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J. M. & Fernandez, L. 2000 Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Res.* **34**, 807–816.

Kalamaras, N., Michalopoulou, H. & Byun, H. R. 2010 Detection of drought events in Greece using daily precipitation. *Hydrol. Res.* **41** (2), 126–133.

Kazi, T. G., Arain, M. B., Jamali, M. K., Jalbani, N., Afridi, H. I., Sarfraz, R. A., Baig, J. A. & Shah, A. Q. 2009 Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. *Ecotoxicol. Environ. Saf.* **72** (2), 301–309.

Kim, J. H., Kim, R. H., Lee, J., Cheong, T. J., Yum, B. W. & Chang, H. W. 2005 Multivariate statistical analysis to identify the major factors governing groundwater quality in the coastal area of Kimje, South Korea. *Hydrol. Process.* **19** (6), 1261–1276.

Lamaro, A. A., Mariñelarena, A., Torrusio, S. E. & Sala, S. E. 2013 Water surface temperature estimation from Landsat 7 ETM+ thermal infrared data using the generalized single-channel method: Case study of Embalse del Río Tercero (Córdoba, Argentina). *Adv. Space Res.* **51** (3), 492–500.

Ledesma, C., Bonansea, M., Rodriguez, C. & Sánchez Delgado, A. R. 2013 Determination of trophic indicators in Rio Tercero reservoir, Cordoba (Argentina) *Rev. Cienc. Agron.* **44** (3), 419–425.

Li, S., Gu, S., Tan, X. & Zhang, Q. 2009 Water quality in the upper Han River basin, China: the impacts of land use/land cover in riparian buffer zone. *J. Hazard. Mater.* **165**, 317–324.

MacDonagh, M. E., Casco, M. A. & Claps, M. C. 2009 Plankton relationships under small water level fluctuations in a subtropical reservoir. *Aquat. Ecol.* **43** (2), 371–381.

Mancini, M., Rodríguez, C., Bagnis, G., Liendo, A., Prosperi, C., Bonansea, M. & Galizia Tundisi, J. 2010 Cianobacterial bloom and animal mass mortality in a reservoir from Central Argentina. *Braz. J. Biol.* **70**, 841–845.

Mariazzi, A. A., Donadelli, J. L., Arenas, P., Di Siervi, M. A. & Bonetto, C. 1992 Impact of a nuclear power plant on water quality of Embalse del Río Tercero Reservoir (Córdoba, Argentina). *Hidrobiologia* **246**, 129–140.

Otto, M. 1998 Multivariate methods. In: *Analytical Chemistry* (R. Kellner, J. M. Mermet, M. Otto & H. M. Widmer, eds). Wiley-VCH, Weinheim, Germany.

Reghunath, R., Sreedhara Murthy, T. R. & Raghavan, B. R. 2002 The utility of multivariate statistical techniques in hydrogeochemical studies: An example from Karnataka, India. *Water Res.* **36** (10), 2437–2442.

Saaremäe, E., Liira, M., Poolakese, M. & Tamm, T. 2013 Removing phosphorus with Ca-Fe oxide granules – a possible wetlands filter material. *Hydrol. Res.* doi:10.2166/nh.2013.101.

Shrestha, S. & Kazama, F. 2007 Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji River basin, Japan. *Environ. Modell. Softw.* **22**, 464–475.

Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M. & Kouimtzis, T. 2003 Assessment of the surface water quality in Northern Greece. *Water Res.* **37** (17), 4119–4124.

Singh, K. P., Malik, A., Mohan, D. & Sinha, S. 2004 Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) – a case study. *Water Res.* **38** (18), 3980–3992.

Torbick, N., Hu, F., Zhang, J., Qi, J., Zhang, H. & Becker, B. 2008 Mapping chlorophyll-a concentrations in West Lake, China using Landsat 7 ETM + . *J. Great Lakes Res.* **34** (3), 559–565.

Varol, M., Gökot, B., Bekleyen, A. & Şen, B. 2012 Spatial and temporal variations in surface water quality of the dam reservoirs in the Tigris River basin, Turkey. *Catena* **92**, 11–21.

Vega, M., Pardo, R., Barrado, E. & Deban, L. 1998 Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Res.* **32** (12), 3581–3592.

Vincent, R., Qin, X., McKay, R., Miner, J., Czajkowski, K., Savino, J. & Bridgeman, T. 2004 Phycocyanin detection from LANDSAT TM data for mapping cyanobacterial blooms in Lake Erie. *Remote Sens. Environ.* **89** (3), 381–392.

Wang, M., Markert, B., Chen, W., Peng, C. & Ouyang, Z. 2012 Identification of heavy metal pollutants using multivariate analysis and effects of land uses on their accumulation in urban soils in Beijing, China. *Environ. Monit. Assess.* **184** (10), 5889–5897.

WHO (World Health Organization) 2006 *Guidelines for drinking-water quality*. 1st Addendum to 3rd Edition Vol. 1. Recommendation. World Health Organization, Geneva, Switzerland.

Wunderlin, D. A., Diaz, M. P., Ame, M. V., Pesce, S. F., Hued, A. C. & Bistoni, M. A. 2001 Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia River basin (Cordoba, Argentina). *Water Res.* **35** (12), 2881–2894.

First received 29 October 2013; accepted in revised form 22 December 2013. Available online 28 January 2014