ISSN = 1980-993X – doi:10.4136/1980-993X
www.ambi-aguawww.ambi-agua.net
E-mail: ambi-agua@agro.unitau.br
Tel. (12) 3625-4212

RECEIVED 05 Aug 2011
24 Nov 2011 ACCEPTED

Application of multivariate statistical techniques to evaluate organic pollution on a river in Argentina

(http://dx.doi.org/10.4136/ambi-agua.696)

Soledad Oliva González[1,2]; César A. Almeida[1,2]; Sylvia Quintar[1]; Miguel A. Mallea[1]; Patricia S. González[1]

[1]Departmento de Química Analítica, Facultad de Química, Bioquímica y Farmacia, Universidad Nacional de San Luis, San Luis, Argentina
[2]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina
e-mail: sololivagonzalez@gmail.com; cesarbioq@gmail.com; squintar@unsl.edu.ar; mmallea@unsl.edu.ar; spgon@unsl.edu.ar

ABSTRACT

The aim of this paper was the application of multivariate statistical techniques to evaluate spatial and temporal variations in the water quality of Potrero de los Funes River using physical, chemical and bacteriological parameters and select the most significant parameters of organic pollution in the river in order to implement in the future water quality monitoring. The river was monitored regularly at three sites: RP1, RP2 and RP3, over the period 2008–2009, for 16 parameters. The complex data matrix was treated with three multivariate statistical techniques: cluster analysis (CA), principal component analysis (PCA) and discriminant analysis (DA). CA generated three groups of sites, cluster 1 (RP1), cluster 2 (RP2) and cluster 3 (RP3) according to relatively low, very high and moderate pollution regions, respectively. PCA identified two components, which were responsible for the data structure explaining 73% of the total variance of the data matrix. Temporal DA (Wet season and Dry season) showed that turbidity, NO_3^- and COD were the discriminant variables. Spatial DA shows that there were significant differences between the three categorical classes, 1 (RP1, low pollution region), 2 (RP2, strongly polluted zone) and 3 (RP3, moderate polluted site) .The discriminating functions contained only eight parameters (EC, NO_3^-, turbidity, DO, BOD, COD, total coliform and fecal coliform) to discriminate between sites. The application of these techniques has achieved meaningful classification of physical, chemical and bacteriological variables and of river water samples, based on seasonal and spatial criteria. This study is essential for the future design of fast and effective monitoring programs of river water quality. That would include only parameters that are indicative of organic pollution.

Keywords: Water quality; organic pollution; multivariate techniques; cluster analysis; principal component; discriminant analysis.

Aplicação de técnicas estatísticas multivariadas para avaliar a poluição orgánica em um rio na Argentina

RESUMO

O objetivo deste trabalho foi avaliar a variação espacial e temporal da qualidade da água do rio Potrero de los Funes com base na aplicação de técnicas estatísticas multivariadas utilizando-se paramêtros físicos, químicos e bacteriológicos, sem perder informações importantes e obter uma metodologia rápida, fácil e barata para monitorar a qualidade da água em rios afetados por poluição orgânica. O rio foi monitorado mensalmente em locais diferentes (RP1, RP2 e RP3) durante um período de 2008-2009 com base em 16 parâmetros. O tratamento envolveu três técnicas estatísticas multivariadas: a análise de cluster (CA), análise de componentes principais (PCA) e análise discriminante (DA). A CA gerou três grupos: Grupo 1 (RP1), grupo 2 (RP2) e grupo 3 (RP3), que correspondem a um nível de contaminação relativamente baixo, alto e moderado, respectivamente. Por meio do PCA foram obtidos dois componentes que explicam 73% da variância total no conjunto de dados. O tempo na DA (estação chuvosa e estação seca) mostraram que a turbidez, NO_{3-} e CE foram as variáveis discriminantes. O DA espacial mostra que existem diferenças significativas entre as três categorias, 1 (RP1 região de baixa contaminação), 2 (RP2 área altamente contaminada) e 3 (RP3, local de poluição moderada). As funções discriminantes obtidas são compostas de apenas oito parâmetros (CE, NO_{3-}, turbidez, OD, DBO, DQO, coliformes totais e coliformes fecais) entre as regiões. A aplicação dessas técnicas tem permitido uma classificação significativa das diferentes regiões do rio baseada em parâmetros físicos, químicos e bacteriológicos, a partir de critérios de tempo e espaço. Este estudo é essencial para a futura concepção de programas de monitorização rápida e eficaz da qualidade da água do rio. Isso incluiria apenas os parâmetros que são indicativos de poluição orgânica

Palavras-chave: qualidade da água; poluição orgânica; técnicas de análise multivariada; análise de cluster; componentes principais; análise discriminante.

1. INTRODUCTION

During the recent decades there has been an increasing concern for the quality of surface waters (Chapman, 1992; Lambrakis et al., 2004; Zhou et al., 2007).

Surface waters are highly vulnerable to pollution due to their easy accessibility for disposal wastewaters. Both, natural processes, such as precipitation inputs, erosion, weathering of crustal materials, as well as anthropogenic influences (urban, industrial and agricultural activities) increase exploitation of water resources and determine the quality of surface waters in a region (Carpenter et al., 1998; Jarvie et al., 1998). Rivers play a major role in digestion or carrying off the municipal and industrial wastewater and agricultural run-off. Municipal and industrial wastewater discharges constitute constant pollution sources, whereas, surface runoff is a seasonal phenomenon, largely affected by weathering process. Seasonal variations in precipitation, surface runoff, interflow and groundwater flow and pumped in/out flows have a strong effect on river discharge and subsequently on the concentration of pollutants in river water (Vega et al., 1998). Since, rivers constitute the main inland water resources for domestic, industrial and irrigation purposes, it is imperative to prevent and control rivers pollution and to have reliable information on water quality for effective management.

Considering spatial and temporal variations in hydrochemistry of rivers, regular monitoring programs are required for reliable estimates of water quality. This result in enormous and complex data matrix comprised of a large number of physical, chemical and microbiological parameters, which are often difficult to interpret rendering meaningful conclusions (Kraft et al., 2003). The problem of data reduction and interpretation of multi-constituent physical, chemical and microbiological measurements can be approached through the application of multivariate statistical methods and exploratory data analysis (Astel et al., 2004; Bagur et al., 2009; Einax et al., 1997, 1998; Eriksson and Hermens, 1995; Marengo et al., 1995; Ouyanga et al., 2006; Simeonov et al., 2002).

The usefulness of multivariate statistical tools in the treatment of analytical and environmental data is reflected by the increasing number of papers cited in Analytical Chemistry Reviews (Brown et al., 1996).

Multivariate statistical techniques such as cluster analysis (CA), principal component analysis (PCA) and discriminant analysis (DA) have been widely used as unbiased methods of water quality data analysis for drawing meaningful information (Bengraine and Marhaba, 2003; Helena et al., 2000; Simeonov et al., 2003; Reghunath et al., 2002; Voncina et al., 2002). Multivariate treatments of data are widely used to characterize and evaluate surface and freshwater quality and are useful to evidence temporal and spatial variations caused by natural and anthropogenic factors linked to seasonality.

In Argentina, there are few river monitoring programs. Water quality of all majors' rivers systems are usually monitored at several sites for a large number of physico-chemical, bacteriological and hydrological parameters with an outcome of large databases of high complexity. Such monitoring programs also involve huge financial inputs. Thus, there is a need to optimize the monitoring networks and the number of water quality parameters considered, including only the more representatives, without losing valuable information (Dixon and Chiswell, 1996). Multivariate statistical techniques and exploratory data analysis are the appropriate tools for a meaningful data reduction and interpretation.

The aim of this paper was:

To apply and evaluate multivariate statistical techniques (CA, PCA, DA) as analytical tools to study the spatial and seasonal variations of water quality of Potrero de los Funes river, using physical, chemical and bacteriological parameters without loosing important information.

To select the most significant parameters of organic pollution in the river, in order to implement future water quality monitoring programs.

2. MATERIALS AND METHODS

2.1. Monitoring area

Potrero de los Funes river is located in a hilly village with a homonymous name, in the province of San Luis, in the centre of Argentina. This river originates in the Sierras Pampeanas between the 33° and 34° parallels of south latitude and the 66° and 65° meridians of east longitude. Its tributaries are Los Molles and La Bolsa streams, whose flows depend on local rains. The river has a length of 2700 m. The river ends into the Potrero de los Funes dam. Its average flow is 7.5 $m^3 s^{-1}$, constituting the dam's main source, with a reservoir capacity of 6.8 hm^3. The basin area is 42 km^2 and has an irrigation potentiality of 4500 Ha (Figure 1).

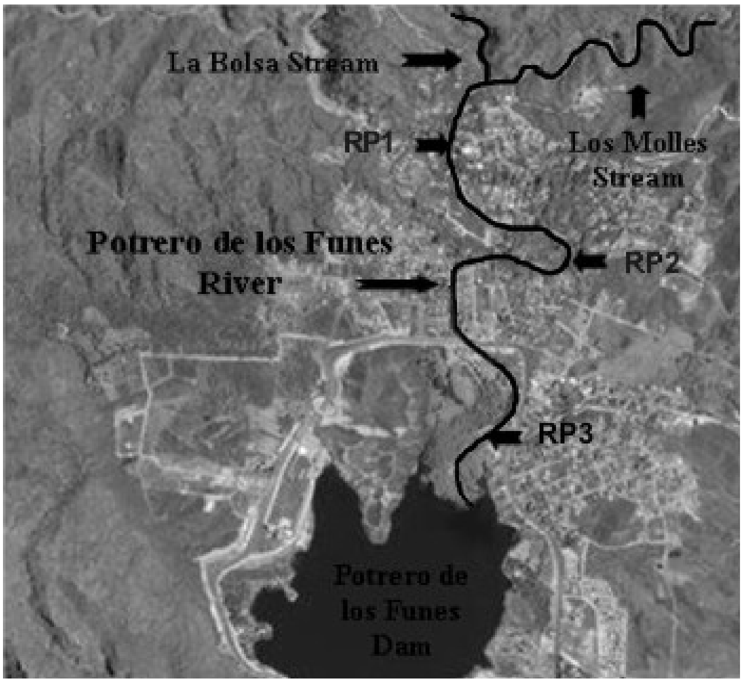

Figure 1. Location of sampling points on Potrero de los Funes River.

Potrero de los Funes village has a population of approximately 2,500 people. The urban settlement extends on the banks of the river and has increased its population by 100 % during the last years. In summer, this river receives a considerable number of tourists, since Potrero de los Funes village is one of the most visited places in the province and the most popular destination for recreational and vacation purposes. Its population increases considerably during the months of December, January, February and March in comparison to the rest of the year.

The main use of this river is for recreational purposes and for drinking water source.

The river discharge varies considerably at different sites, showing an increasing trend towards downstream course due to inputs of domestic wastewater drainage from Potrero de los Funes town. In summer months the river receives pollution load from anthropogenic sources due to both, tourism activities and drainage surface and domestic wastewater runoff.

Three sampling sites along the river (upper, middle and downstream) were chosen for the study. They were: RP1, representing background values, i.e. minor interference from anthropogenic activities. This station was located in the main intake of the drinking water plant of the Potrero de los Funes town, this zone reflect the "**environmental base line of the studied water course**". RP2 located 1112 m down stream from RP1; in a fully urbanized area nearby wastewater drainage from residential buildings and hotels. The whole area of this second sampling site was strongly influenced by tourism activities. A third site, RP3 was located 2082 m down stream from RP1, at the end of the urban area, 618 m before the river mouth discharge in Potrero de los Funes dam.

2.2. Monitored parameters and analytical methods

The monitoring network and sampling strategy were designed to cover a wide range of determinants at key sites, which reasonably represented the water quality of the river system, accounting for inputs from tributaries and from drains that may impact downstream water quality (Almeida et al., 2006, 2007; Garbagnati et al., 2005).

Samples were collected each month from two points (a–b) across the river width at all the three sampling sites in order to monitor changes caused by the seasonal hydrological cycle and anthropogenic influence during the 2008-2009 period. Sampling, preservation and transportation of water samples were performed according to Standard Methods for the Examination of Water and Wastewater (APHA, 2005).

Water pH and electrical conductivity were measured using a portable meter. All other parameters were determined in laboratory following standard protocols (APHA, 2005). The samples were analyzed for 16 parameters, which include pH, electrical conductivity (EC), turbidity (Tbd), total alkalinity(T-Alk), total hardness (T-Hard), dissolved oxygen (DO), 5-days biochemical oxygen demand (BOD), chemical oxygen demand (COD), chloride (Cl^-), sulfate (SO_4^{2-}), phosphate (PO_4^{3-}), nitrate (NO_3^-), sodium (Na^+), potassium (K^+), total coliform (T-Coli) and faecal coliform (F-Coli). All the water quality parameters were expressed in mgL^{-1}, except pH, EC ($\mu S\ cm^{-1}$), total coliform and faecal coliform (MPN/100 ml). Analytical quality of data was ensured through careful standardization, procedural blank measurements, spiked and duplicate samples. The ionic charge balance of each sample was within ±5%.

2.3. Data treatment and multivariate statistical methods

Multivariate analysis of the water quality data set was performed through CA, PCA and DA techniques (Simeonov et al., 2003; Reisenhofer et al., 1998; Wunderlin et al., 2001). Cluster analysis and PCA were applied on experimental data standardized through z-scale transformation in order to avoid misclassification due to wide differences in data dimensionality. Standardization tends to increase the influence of variables whose variance is small and reduce the influence of variables whose variance is large. Furthermore, the standardization procedure eliminates the influence of different units of measurement and renders the data dimensionless.

PCA was performed on a correlation matrix of rearranged data, so that it explained the structure of the underlying data set (Kowalkowski et al., 2006). All the mathematical and statistical computations were made using Infostat 2008 software.

2.3.1. Cluster analysis (CA)

Cluster analysis is an unsupervised pattern recognition technique that uncovers intrinsic structure or underlying behavior 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. Hierarchical agglomerative CA was performed on the normalized data set by means of the Ward's method, using Euclidean distances as a measure of similarity. Cluster analysis was applied to the water quality data set with view to group the similar sampling sites (spatial variability) spread over the river stretch and in the resulted dendrogram (Hagedorn et al., 1999; Librando, 2000).

2.3.2. Principal component analysis (PCA)

Principal Component (PC) is a linear combination of observable water quality variables. PCA of the normalized variables (water quality data set) was performed to extract significant PCs and to further reduce the contribution of variables with minor significance (Bagur et al., 2009; Helena et al., 2000; Zhou et al., 2007).

2.3.3. Discriminant analysis (DA)

Discriminant analysis is used to determine the variables, which discriminate between two or more naturally occurring groups. It operates on raw data and the technique constructs a discriminant function for each group (Ouyanga et al., 2006; Otto, 1998; Parveen et al., 1999)

In this case study, two groups for temporal (two seasons, **Wet:** October to February and **Dry:** March to September) and three groups for spatial (three sampling regions: **RP1**, **RP2** and **RP3**) evaluations were selected. Discriminant analysis was applied to raw data by using the forward stepwise mode. Site (spatial) and season (temporal) were the grouping (dependent) variables, while all the measured parameters constituted the independent variables.

3. RESULTS AND DISCUSSION

Water quality monitoring of Potrero de los Funes river was regularly conducted over a period of 2008-2009 at three different sites. All the samples were analyzed for 16 parameters. Basic statistics for the river water quality data set is summarized in Table 1.

Table 1. Basic descriptive statistic for sampling points in Potrero de los Funes River.

	RP1			RP2			RP3		
	$\bar{x}$	SD	CV(%)	$\bar{x}$	SD	CV(%)	$\bar{x}$	SD	CV(%)
EC	121,4	14,32	11,80	275,65	61,46	22,30	185,43	21,86	11,79
Tbd	0,55	0,51	93,44	4,77	0,87	18,21	2,62	0,50	18,97
SO_4^{2-}	1,95	0,50	25,50	7,58	2,36	31,16	5,44	1,08	19,79
NO_3^-	0,67	0,40	59,91	2,29	0,22	9,39	1,73	0,17	9,82
COD	1,75	0,59	33,96	5,45	2,62	48,07	2,33	0,56	24,16
BOD	1,41	0,42	29,72	3,56	0,94	26,51	2,72	2,12	77,89
DO	9,02	1,13	12,52	7,30	0,38	5,19	7,55	0,47	6,26
T-Coli	12,14	6,29	51,82	488,18	403,69	82,69	155,77	33,61	21,57
F-Coli	0,00	0,00	0,00	45,55	27,21	59,75	17,65	5,62	31,84
Cl^-	5,24	3,04	58,01	8,30	3,03	36,55	6,12	3,01	49,16
Na	5,87	1,58	26,84	23,25	14,14	60,82	10,58	1,67	15,74
K	1,38	0,31	22,14	6,14	1,71	27,84	3,25	0,74	22,90
pH	7,59	0,27	3,52	7,95	0,43	5,38	7,89	0,45	5,70
T-Hard	79,20	11,59	14,64	96,20	18,43	19,15	104,16	30,63	29,40
Alk	87,92	12,39	14,09	106,66	18,28	17,14	202,75	349,20	172,20

Water quality data were subjected to different multivariate statistical techniques to explore their temporal and spatial trends.

Cluster analysis rendered a dendrogram (Figure 2), where sampling sites of the river were grouped into three statistically significant clusters. The clustering procedure generated three groups of sites in a very convincing way, as the sites in these groups have similar characteristic features and natural background source types. **Cluster 1** (**RP1**), **cluster 2** (**RP2**) and **cluster 3** (**RP3**) correspond to a relatively low pollution, very high pollution and moderate pollution regions, respectively. It implies that for rapid assessment of water quality, only one site (**Cluster 1**) may serve as reference in the spatial assessment of water quality as the point reflecting the whole network of this work what it is known as “**environmental base line of the studied water course”**.

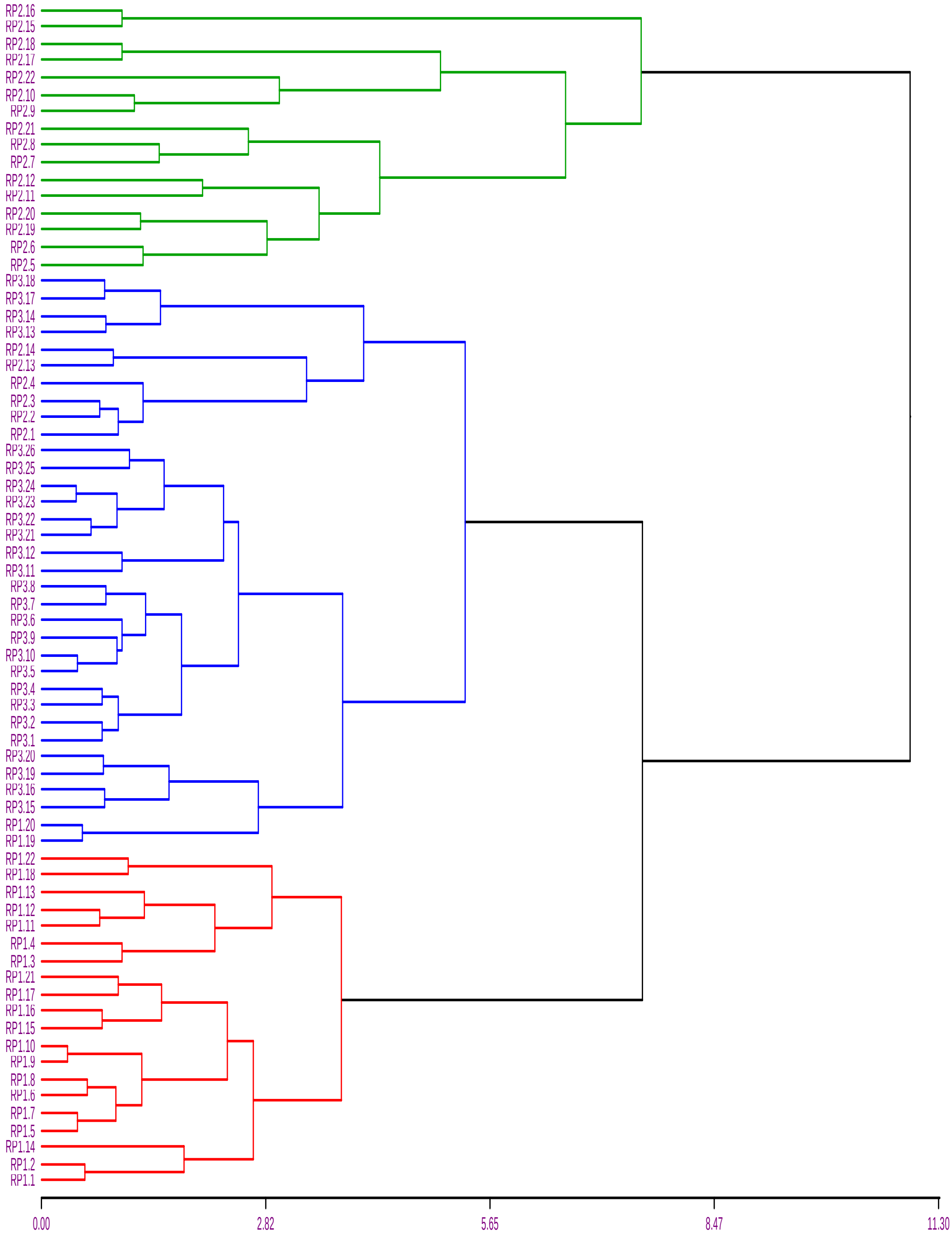

Figure 2. Dendrogram of the CA according to Ward. Monitoring locations (obtained on the basis of dataset cover the period of 2008-2009) of the Potrero de los Funes River: RP1 – RP3 – RP2. Euclidean Distance - Standardized variables.

PCA was applied to normalized data to compare the compositional patterns between water samples and to identify the factors that influence each one. PCA of the entire data set (Table 1) resulted in two PCs with eigenvalues explaining about 73% of the total variance in the water quality data set.

The Scree plot was used to identify the number of PCs to be retained, in order to comprehend the underlying data structure (Vega et al., 1998). This plot (figure not shown) showed a pronounced change of slope after the first two eigenvalues. Projections of the original variables on the subspace of the PCs are called loadings and coincide with the correlation coefficients between PCs and variables.

The first PC, accounting for 63% of the total variance, was correlated with electrical conductivity, turbidity, NO_3^-, SO_4^{2-}, COD, Ln T-Coli and Ln F-Coli, and can be interpreted as a mineral and organic component of the river water. The second PC, accounting for 10% of the total variance, was correlated with turbidity, dissolved oxygen, COD, BOD, Ln T-Coli and Ln F-Coli. This PC2 represents anthropogenic pollution sources and can be explained by the positive correlation between DO and COD and the negative correlation between DO, BOD and bacteria as indicators of anthropogenic influence (Tables 2 and 3).

Table 2. Cumulative variance for the PCs.

Component	Lambda	Proportion	Cumulative
CP1	7.55	0.63	0.63
CP2	**1.19**	**0.10**	**0.73**
CP3	0.95	0.08	0.81
CP4	0.66	0.05	0.86
CP5	0.53	0.04	0.91
CP6	0.39	0.03	0.94
CP7	0.22	0.02	0.96
CP8	0.20	0.02	0.98
CP9	0.13	0.01	0.99
CP10	0.09	0.01	1.00

Table 3. Loadings of experimental variables on the first three PCs for complete data set.

Variable	PC1	PC2
SO_4^{2-}	0.90	0.13
NO_3^-	0.92	0.09
COD	0.72	0.62
BOD	0.60	-0.78
EC	0.89	0.13
Turbidity	0.93	0.67
Cl^-	0.42	-0.15
Ln T-Coli	0.79	-0.90
Ln F-Coli	0.83	-0.93
Na	0.37	0.07
K	0.26	0.17
DO	-0.62	0.94
Cophenetic correlation: 0.98		

The first PC represents run-off from fields with high load of solids (turbidity and NO_3^-) and waste disposal activities (SO_4^{2-}, EC, COD, Ln T Coli, Ln F Coli). The second PC described anthropogenic pollution (turbidity, dissolved oxygen, COD, BOD, Ln T-Coli and Ln F-Coli). The results from the PCA suggested that most of the variation is explained by a set of variables that show that sewage and run off from urban area affect river water quality and another set of variables that are anthropogenic indicators. Therefore, water quality of Potrero de los Funes river is also affected by human activities.

PCA did not result in much data reduction, we still needed 9 parameters (about 56% of the 16 parameters) to explain 73% of the data variance. However, PCA was useful to identify those parameters which had the greatest contribution to anthropogenic and organic variation in the river water quality (Tables 2 and 3).

Temporal variations in water quality were further evaluated through DA. Temporal DA was performed on raw data after dividing the whole data set into two seasonal groups (**Wet season** and **Dry season**). The discriminant function (DF) and classification matrix (CT) obtained from the forward stepwise mode of DA are shown in Table 4. In forward stepwise mode, variables are included step-by-step beginning with the more significant until no significant changes are obtained. The forward stepwise mode using 16 predictor variables was used to develop a model to discriminate among the two seasons (**Wet** and **Dry)**. Using a stepwise selection algorithm, three variables were identified as significant predictors of the Wet-Dry classification. The discriminating function with P-value less than 0.05 was statistically significant at the 95% confidence level (Table 4).

Table 4. Classification matrix and analysis summary for discriminant analysis of temporal variation in Potrero de los Funes River.

Classification Matrix					
Group	**Dry**	**Wet**	**Total**	**Error (%)**	
Dry	29	9	38	23.68	
Wet	5	27	32	15.63	
Total	**34**	**36**	**70**	**20.00**	
Analysis Summary					
Discriminant Function	**Eigenvalue**	**Relative Percentage**	**Canonical Correlation**		
1	19.91	92.69	0.98		
Functions Derived	**Wilks Lambda**	**Chi-Square**	**DF**	**P Value**	
1	0.70	23.36	3	0.00	

Forward stepwise DA showed that turbidity, NO_3^- and COD were the discriminant variables (Table 5). Thus, temporal DA results suggested that they were the most significant variables to discriminate between seasons; which implies that these three parameters account for most of the expected temporal variations in river water quality. This also suggests that anthropogenic pollution, mainly due to discharge of wastewater into the river and tourist activities, discriminates between seasons throughout the year, since these variables increase during wet season coincidently with the increasing tourist activities (Bagur et al., 2009; Wunderlin et al., 2001; Zhou et al., 2007).

Table 5 Classification standardized coefficients for Forward stepwise DA mode of temporal variation in Potrero de los Funes River water.

Standardized Coefficients	
	1
Turbidity	2.16
NO_3^-	-1.17
COD	-0.67

The summer values of COD, turbidity and NO_3^- (wet season) increased compared to winter values, because of the tourist influx and the extra input from drainage basin surface runoff due to improved precipitation, but these increase was not so remarkable because of the fold dilution attributed to enhanced weathering process (Figure 3).

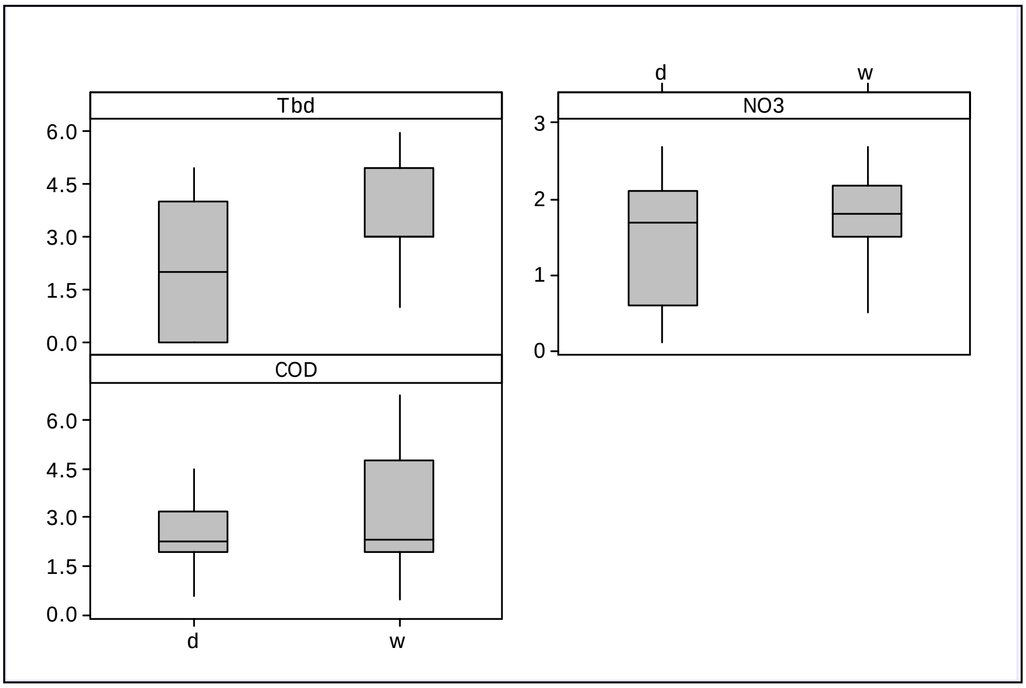

Figure 3. Temporal variations: (Tbd) turbidity, (NO3) nitrate and (COD) Chemical Oxygen Demand in Potrero de los Funes River water. **d-** Dry season **w-** Wet season.

Spatial DA was performed with the same raw data set comprised of 16 parameters, after grouping them into three categorical classes, **1** (**RP1**, low pollution region), **2** (**RP2**, strongly polluted zone) and **3** (**RP3**, moderate polluted site) as obtained from CA. The site (clustered) was the grouping (dependent) variable, while all measured parameters constituted independent variables. Discriminant functions and classification matrices obtained from the forward stepwise mode of DA are shown in Tables 6 and 7. Similar to temporal DA, forward stepwise DA mode constructed two DFs with P-values less than 0.05 statistically significant at the 95% confidence level (Table 6). The correct assignments (94%) by DA for three different site clusters (**RP1**, **RP2** and **RP3**) further confirmed the adequacy of DA. The grouping pattern was coincident with our previous spatial CA (Figure 2 and 4). Both CA and DA predicted important differences in water quality due to impacts from anthropogenic influence. DA showed that there were significant differences between these three regions (low, very high and moderate polluted), which were expressed in terms of eight discriminating parameters conforming the discriminating functions (Table 7). Hence, DA rendered a considerable data reduction. In this manner, DA provides additional information over PCA in the assessment of spatial differences between regions as expressed by the corresponding classification matrix (Table 6).

Table 6. Classification matrix and analysis summary for discriminant analysis of spatial variations in Potrero de los Funes River water.

Classification Matrix					
Group	**1**	**2**	**3**	**Total**	**Error(%)**
1	20	0	1	21	4.54
2	0	22	1	23	4.54
3	2	0	24	26	9.09
Total	**22**	**22**	**26**	**70**	**6.02**

Analysis Summary			
Discriminant Function	**Eigenvalue**	**Relative Percentage**	**Canonical Correlation**
1	19.91	92.69	0.98
2	1.57	7.31	0.78

Functions Derived	**Wilks Lambda**	**Chi-Square**	**DF**	**P-Value**
1	0.02	245.04	24	0.00
2	0.39	58.06	11	0.00

Table 7. Classification functions 1 and 2 for Forward stepwise DA mode of spatial variations in Potrero de los Funes River.

Standardized Coefficients		
	1	2
SO_4^{2-}	0.07	-0.78
NO_3^-	0.54	-0.53
BOD	0.35	0.37
T-Coli	-0.26	0.24
F-Coli	0.54	0.37
Cl^-	0.02	0.42
EC	-0.79	-0.03
Turbidity	1.17	0.00
COD	0.12	1.23
DO	-0.20	0.33

Box and whisker plots of some selected discriminating parameters identified by spatial DA were constructed to evaluate different patterns associated with spatial variations in river water quality (Figure 5). For example, the concentration of dissolved oxygen in RP2 and RP3 were lower than in RP1. At the same time, COD and BOD were higher than in RP2, thus pointing to a deterioration of the water quality caused by the anthropogenic influence. Similar trends of spatial variations observed for Turbidity, NO_3^- and EC were due to pollution load from several urban sources at RP2. High levels of bacteriological contamination (total coliform and fecal coliform) in RP2 are indicative of anthropogenic influence due to urban sources and tourist activities (Figure 5).

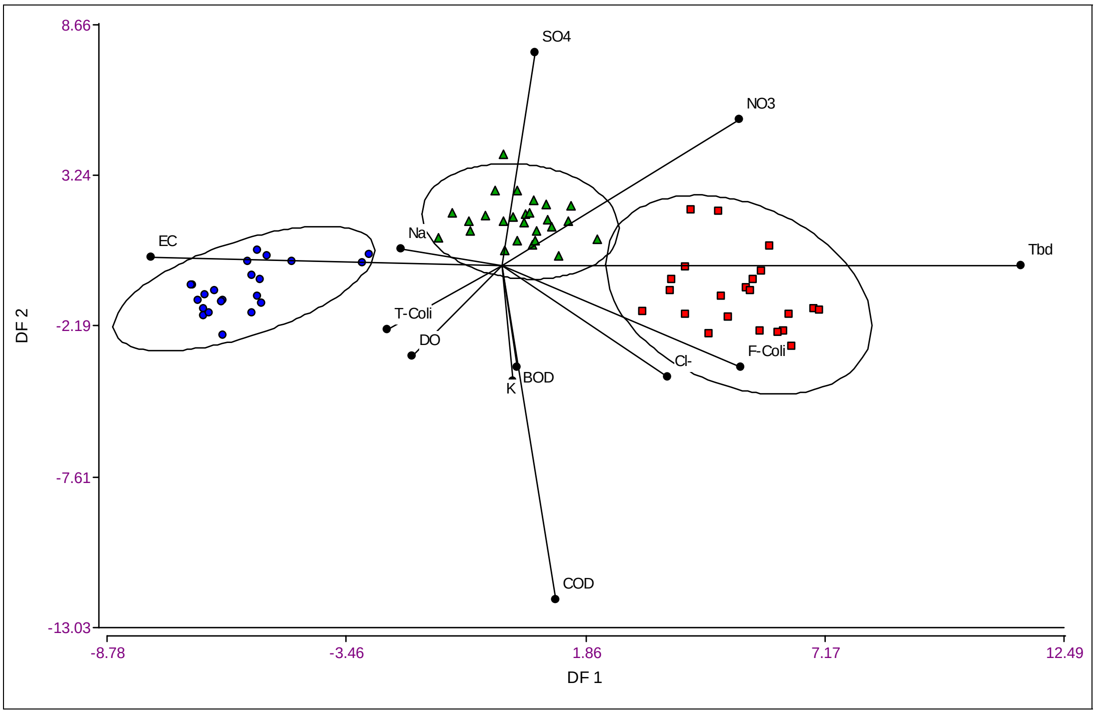

Figure 4. Biplot showing three classes for spatial variation for Forward stepwise DA mode. Class 1: [●] RP1 (Low polluted river region). Class 2: [■] RP2 (Very high polluted river region). Class 3: [▲] RP3 (Moderate polluted river region).

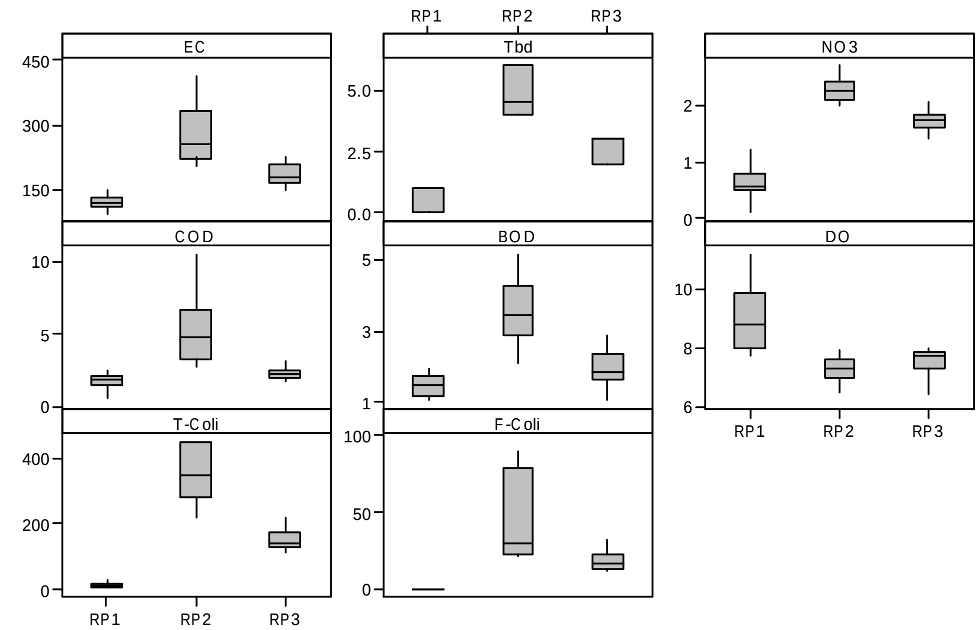

Figure 5. Spatial variations: (EC) Electrical Conductivity, (COD) chemical Oxygen Demand, (T-Coli) Total Coliform, (Tbd) Turbidity, (BOD) Biochemical Oxygen Demand, (F-Coli) Fecal Coliform, (NO3) Nitrate and (DO) Dissolved Oxygen in Potrero de los Funes River water. Spatial points: RP1 – Low polluted river point**,** RP2 – High polluted river point**,** RP3 – Moderate polluted river point.

Quality of Potrero de los Funes river surface waters impoverish down stream from point RP2. This loss of quality coincides spatially with the more urbanized part of the town and has also a temporal pattern linked to the increase of tourist activities during the summer and wet season.

The river enters to Potrero de los Funes dam with high organic load and enriched with nutrients, accordingly this input could seriously disturb the trophic status of the reservoir.

Up to present, eutrophization studies of Potrero de los Funes dam have not been done, therefore, it is important to know river water quality and the significant parameters that are indicatives of organic and anthropogenic water pollution, in order to indirectly, assess the water quality of the dam.

4. CONCLUSIONS

Hierarchical CA helped to group the sampling sites into three clusters of similar characteristics pertaining to water quality characteristics and pollution sources.

CA helped established RP1 as reference site. The water quality in this site will allow a future control of river water quality as a benchmark of good quality when to want to implement monitoring network programs of water quality, allowing from these "**environmental base line**" take corrective actions when the river will be affected by sewage and anthropogenic influence.

PC 1 and PC 2 showed that the parameters responsible for water quality variations are mainly related to organic pollution loads (anthropogenic). These parameters corresponded to those obtained in DA, highlighting their importance as indicatives of organic pollution under the influence of urbanization in river water quality.

Discriminant analysis gave the best results for both temporal and spatial analysis. It rendered an important data reduction as it used only three parameters (turbidity, NO_3^-, COD) to discriminate between seasons and only eight parameters (EC, NO_3^-, turbidity, DO, BOD, COD, total coliform and fecal coliform) to discriminate between the three spatial areas.

The discriminant functions obtained by DA will allow classifying new observations into the pre-established groups. In order to design future monitoring programs, it is imperative to make a study of river behavior. From this assessment and including powerful statistical tools, we obtained functions that will categorize future new samples according to the pollution status or water quality condition. This makes the study significant to design easy and effective monitoring programs of rivers water quality, using parameters that indicate organic pollution.

5. ACKNOWLEDGMENTS

This work was supported by Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) and Universidad Nacional de San Luis (Argentina).

6. REFERENCES

ALMEIDA, C.; QUINTAR, S.; GONZÁLEZ, P.; MALLEA, M. A. Assessment of irrigation water quality. A proposal of a quality profile. **Environmental Monitory and Assessment**, v. 114, p. 377-389, 2006.

ALMEIDA, C.; QUINTAR, S.; GONZÁLEZ, P.; MALLEA, M. A. Influence of urbanization and tourist activities on the water quality of the Potrero de los Funes River (San Luis–Argentina). **Environmental Monitory and Assessment**, v. 133, p. 1007-1014, 2007.

AMERICAN PUBLIC HEALTH ASSOCIATION - APHA. **Standard methods for the examination of water and wastewater**. 21st Edition. Washington DC, 2005.

ASTEL, A.; MAZERSKI, J.; POLKOWSKA, Z.; NAMIES´NIK, J. Application of PCA and time series analysis in studies of precipitation in Tricity (Poland). **Advances in Environmental Research**, v. 8, p. 337–349, 2004. http://dx.doi.org/10.1016/S1093-0191(02)00107-7

BAGUR, M. G.; MORALES, S.; LÓPEZ-CHICANO, M. Evaluation of the environmental contamination at an abandoned mining site using multivariate statistical techniques—The Rodalquilar (Southern Spain) mining district. **Talanta**, v. 80, n. 1, p. 377–384, 2009. http://dx.doi.org/10.1016/j.talanta.2009.06.075

BENGRAINE, K.; MARHABA, T. F. Using principal component analysis to monitor spatial and temporal changes in water quality. **Journal of Hazard Mater**, v. B, n. 100, p. 179–195, 2003.

BROWN, S. D.; SUM, S. T.; DESPAGNE, F.; LAVINE, D. K. Chemometrics. **Analytical Chemistry**, v. 68, n. 12, p. 21-62, 1996. http://dx.doi.org/10.1021/a1960005x

CARPENTER, S. R.; CARACO, N. F.; CORRELL, D. L.; HOWARTH; R. W.; SHARPLEY, A. N.; SMITH, V. H. Nonpoint pollution of surface waters with phosphorus and nitrogen. **Ecology Application**, v. 8, p. 559–568, 1998. http://dx.doi.org/10.1890/1051-0761(1998)008[0559:NPOSWW]2.0.CO;2

CHAPMAN, D. **Water quality assessment.** In: CHAPMAN, D.; UNESCO; WHO; UNEP. London: Chapman & Hall, 1992. 585 p.

DIXON, W. AND CHISWELL, B. Review of aquatic monitoring program design. **Water Research**, v. 30, p. 1935–1948, 1996. http://dx.doi.org/10.1016/0043-1354(96)00087-5

EINAX, J. W.; ZWANZIGER, H. W.; GEISS, S. **Chemometrics in environmental analysis.** Weinheim: Wiley, 1997. http://dx.doi.org/10.1002/352760216X

EINAX, J. W.; TRUCKENBRODT, D.; KAMPE, O. River pollution data interpreted by means of chemometric methods. **Microchemistry Journal**, v. 58, p. 315–324, 1998. http://dx.doi.org/10.1006/mchj.1997.1560

ERIKSSON, L.; HERMENS, J. L. M. A multivariate approach to QSAR. In: EINAX, J. W. (Ed.). **Chemometrics in environmental chemistry applications.** Berlin; Heidelberg: Springer, 1995. p. 135–169. Vol. 2, Part H.

GARBAGNATI, M. A.; GONZÁLEZ, S. P.; ANTÓN, R. I.; MALLEA, M. A. Physicist–chemistry characterization, capacity buffer and establishment of environmental background of the Grande River, San Luis–Argentina. **Ecología Austral**, v. 15, p. 59–71, 2005.

HAGEDORN, C.; ROBINSON, S. L.; FILTZ, J. R.; GRUBBS, S. M.; ANGIER, T. A.; RENEAU JR., R. B. Determining sources of fecal pollution in a rural Virginia Watershed with antibiotic resistance patterns in fecal streptococci. **Application of Environmental of Microbiology**, v. 65, p. 5522–5531, 1999.

HELENA, B.; PARDO, R.; VEGA, M.; BARRADO, E.; FERNANDEZ, J. M.; FERNANDEZ, L. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis. **Water Research**, v. 34, p. 807–816, 2000. http://dx.doi.org/10.1016/S0043-1354(99)00225-0

JARVIE, H. P.; WHITTON, B. A.; NEAL, C. Nitrogen and phosphorus in east-coast British rivers: speciation, sources and biological significance. **Science of the Total Environmental**, v. 210–211, p. 79–109, 1998. http://dx.doi.org/10.1016/S0048-9697(98)00109-0

KOWALKOWSKI, T.; ZBYTNIEWSKI, R.; SZPEJNA, J.; BUSZEWSKI, B. Application of chemometrics in river water classification. **Water Research**, v. 40, p. 744- 752, 2006. http://dx.doi.org/10.1016/j.watres.2005.11.042

KRAFT, J.; KOWALIK, C.; EINAX, J. W. Statistical evaluation of river pollution data Examplified by the Elbe river system. In: PARCZEWSKI, A. (Ed.). **Chemometrics. Methods and Applications.** II Conference, 16–19 October 2003. Zakopane, 2003.p. 40–49.

LAMBRAKIS, N.; ANTONAKOS, A.; PANAGOPOULOS, G. The use of multicomponent statistical analysis in hydrological environmental research. **Water Research**, v. 38, p. 1862–1872, 2004. http://dx.doi.org/10.1016/j.watres.2004.01.009

LIBRANDO, V. Chemometric evaluation of surface water quality at regional level. **Fresenius Journal of Analytical Chemistry**, v. 339, p. 613-619, 2000. http://dx.doi.org/10.1007/BF00325547

MARENGO, E.; GENNARO, M. C.; GIACOSA, D.; ABRIGO, C.; SAINI, G.; AVIGNONE, M. T. How chemometrics can helpfully assist in evaluating environmental data Lagoon water. **Analytical Chimical Acta**, v. 317, p. 53–63, 1995. http://dx.doi.org/10.1016/0003-2670(95)00402-5

OTTO, M. Multivariate methods. IN: KELLNER, R.; MERMET, J. M.; OTTO, M.; WIDMER, H. M. **Analytical chemistry.** Weinheim: Wiley-VCH, 1998. 916p.

OUYANGA, Y.; NKEDI-KIZZAB, P.; WUC, Q. T.; SHINDEB, D.; HUANGO, C. H. Assessment of seasonal variations in surface water quality. **Water Research**, v. 40, p. 3800–3810, 2006. http://dx.doi.org/10.1016/j.watres.2006.08.030

PARVEEN, S.; PORTIER, K. M.; ROBINSON, K.; EDMISTON, L.; TAMPLIN, M. L. Discriminant analysis of ribotype profiles of Escherichia coli for differentiating human and nonhuman sources of fecal pollution. **Application of Environmental Microbiology**, v. 65, p. 3142–3147, 1999.

REGHUNATH, R.; MURTHY, T. R. S.; RAGHAVAN, B. R. The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. **Water Research**, v. 36, p. 2437–2442, 2002. http://dx.doi.org/10.1016/S0043-1354(01)00490-0

REISENHOFER, E.; ADAMI, G.; BARBIERI, P. Using chemical and physical parameters to define the quality of karstic freshwaters (Timavo River, North-eastern Italy): a chemometric approach. **Water Research**, v. 32, p. 1193-1203, 1998. http://dx.doi.org/10.1016/S0043-1354(97)00325-4

SIMEONOV, V.; EINAX, J. W.; STANIMIROVA, I.; KRAFT, J. Environmetric modeling and interpretation of river water monitoring data. **Analytical and Bioanalalytical Chemistry Journal**, v. 374, p. 898–905, 2002. http://dx.doi.org/10.1007/s00216-002-1559-5

SIMEONOV, V.; STRATIS, J.A.; SAMARA, C.; ZACHARIADIS, G.; VOUTSA, D.; ANTHEMIDIS, A. et al. Assessment of the surface water quality in northern Greece. **Water Research**, v. 37, p. 4119–4124, 2003. http://dx.doi.org/10.1016/S0043-1354(03)00398-1

VEGA, M.; PARDO, R.; BARRADO, E.; DEBAN, L. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. **Water Research**, v. 32, p. 3581–3592, 1998. http://dx.doi.org/10.1016/S0043-1354(98)00138-9

VONCINA, D. B.; DOBCNIK, D.; NOVIC, M.; ZUPAN, J. Chemometric characterisation of the qualityof river water. **Analytical Chimica Acta**, v. 462, p. 87–100, 2002. http://dx.doi.org/10.1016/S0003-2670(02)00298-2

WUNDERLIN, D. A.; DIAZ, M. P.; AME, M. V.; PESCE, S. F.; HUED, A. C.; BISTONI, M. A. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia river basin (Cordoba-Argentina). **Water Research**, v. 35, p. 2881–2894, 2001. http://dx.doi.org/10.1016/S0043-1354(00)00592-3

ZHOU, F.; HUANG, G. H.; GUO, H.; ZHANG, W.; HAO, Z. Spatio-temporal patterns and source apportionment of coastal water pollution in eastern Hong Kong. **Water Research**, v. 41, p. 3429–3439, 2007. http://dx.doi.org/10.1016/j.watres.2007.04.022