Assessment of seasonal variations in stream water by principal component analysis

M. M. Taboada-Castro, M. L. Rodríguez-Blanco
& M. T. Taboada-Castro
Faculty of Sciences, University of A Coruña, Spain

Abstract

Assessment of seasonal changes in surface water is an important aspect for the interpretation of hydrochemical data. Thirteen physical and chemical parameters monitored at four sampling stations along the Corbeira stream, NW Spain, were analyzed during a three-year period. The Corbeira stream drains a rural catchment (16 Km^2) with low population density. The land use consists mainly of forested and agricultural land. The geological material is basic schist.

A total of 51 samples were collected from each station. The principal component analysis (PCA) technique was used to evaluate the seasonal correlations of water quality parameters.

Four principal components, accounting for 88.7 and 83.3% of the total variances of information contained in the original dataset for spring and winter, respectively, were obtained. In summer and autumn, three principal components accounted for 80.3 and 81.1% of the total variance, respectively.

The results revealed that conductivity, chloride, magnesium, sulphate and nitrate were always the most important variables contributing to water physical-chemical properties in the stream for all seasons. The first three can be interpreted as a mineral component of the stream water. This clustering of variables points to a common origin for these minerals, most likely from an alteration of schist, whereas nitrates may be interpreted as representing influences from natural (decomposition of organic matter from soils) and anthropogenic inputs. Autumn and winter (periods with high water discharge) showed a strong influence of dissolved organic carbon (DOC) and total nitrogen (Kjeldahl), respectively. This finding could be due to surface runoff.

Keywords: hydrochemistry, surface water, principal component analysis, seasonal variation.

1 Introduction

Characterization of seasonal changes in surface water quality is an important aspect for evaluating temporal variations of river pollution due to natural or anthropogenic inputs of point and non-point sources. Long-term surveys and monitoring programs of water quality are an adequate approach to a better knowledge of river hydrochemistry and pollution, but they produce large sets of data, which are often difficult to interpret, and are not fully explored [1].

The problems of data reduction and interpretation, characteristic change in water quality parameters, and indicator identification can be approached through the use of the principal component analysis (PCA) and the closely associated method of factor analysis (PFA). PCA and PFA are multivariate statistical techniques used to identify important components or factors that explain most of the variances of a system. They are designed to reduce the number of variables to a small number of indices (i.e., principal components or factors), which make the data easier to understand. Details for mastering the arts of PCA and PFA are published elsewhere [2–4].

In recent years, the PCA and PFA techniques have been applied to a variety of environmental issues, including the evaluation of ground water monitoring wells and hydrographs, examination of spatial and temporal patterns of surface water quality, identification of chemical species related to hydrological conditions, and assessment of environmental quality indicators [5–11].

The aim of this study is to apply the PCA technique to evaluate the seasonal correlations of water hydrochemical parameters and to extract those parameters that are most important in assessing seasonal variations of the Corbeira water quality.

2 Material and methods

2.1 Study area

The Corbeira (Fig. 1) is a small tributary of the Mero River, the principal river supplying the city of A Coruña (NW Spain). It has a catchment area of approximately 16 Km^2, a river length of 10 Km and altitudes varying from 470 m.a.s.l. to 70 m.a.s.l. at its confluence with the Mero. Within the Corbeira catchment, land use is best described as rural (more forest than agricultural) and there is low population density. The underlying geology across the area is basic schist. The climate is humid temperate, and mean rainfall is in the range of 1022mm. Wet period precipitations are maximal between October and February. The dry period corresponds with the summer.

2.2 Data and statistical methods

In this study, four sampling stations located in the main course and 13 physico-chemical parameters obtained from each station were used for analysis.
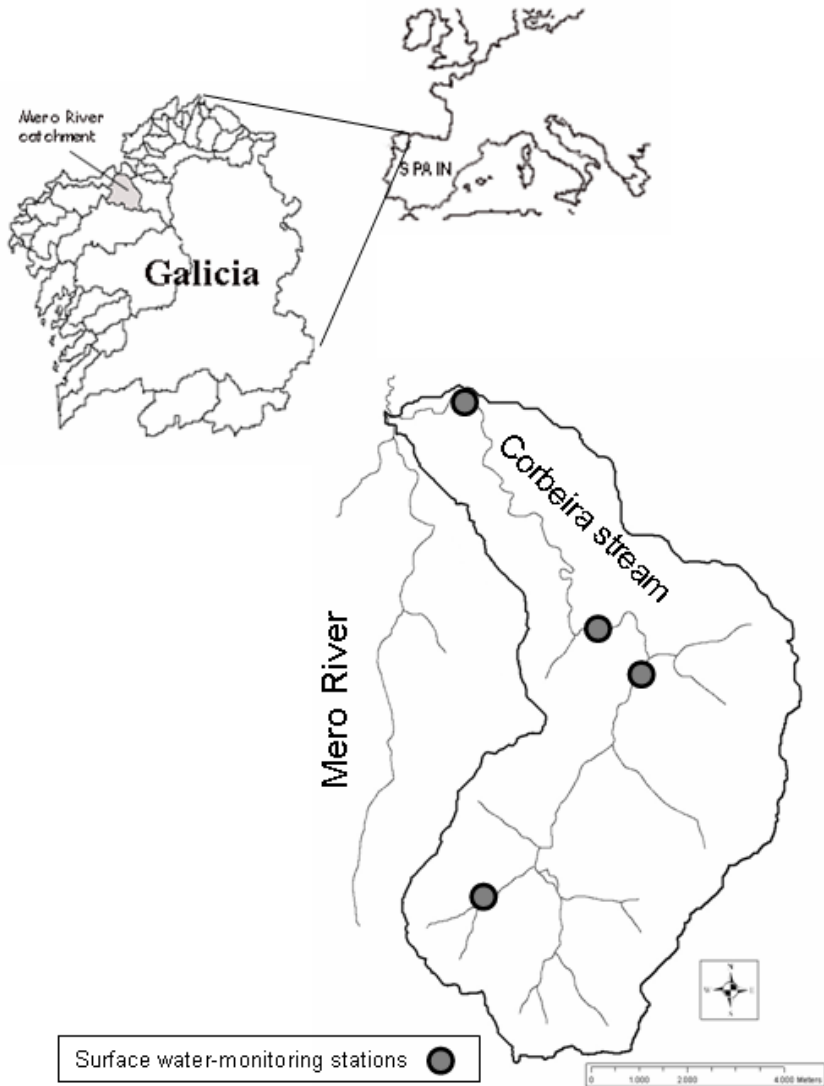
Figure 1:   Map showing the area studied and the location of sampling stations.

Fifty one samples were collected from each station during a 3-year time period (between December 2003 and December 2006). The timing of sample collection was routine and not intended to capture any specific flow or rainfall events.

Physico-chemical parameters were determined by following standard methods of analysis. Electrical conductivity, pH and temperature were measured on unfiltered samples using standard electrochemical techniques. After sample

filtration, cations were analyzed by ICP-MS, and anions and organic carbon by capillary electrophoresis. Nitrogen Kjeldahl was determined.

Principle component analysis was used to reduce the complexity of the datasets and to ascribe concentration seasonal variations. The fundamental goal of PCA is to find a set of uncorrelated linear combinations that is able to explain most of the variance of the original multivariate data [12].

Using the standard statistical software SPSS 11.5, significant principal components (PCs) with eigenvalues greater than unity (i.e. PCs explaining more than the variance of one parameter) were extracted [7, 8, 13]. The Varimax criterion, which ensures maximum correlation of each parameter with one single PC, was applied for optimal rotation of the PCs. For this study, a factor correlation coefficient greater than 0.85 (or 85%) was considered significant.

To distinguish the variations of each parameter for a given season, the data was divided into four distinct temporal databases (data from each season: winter, spring, summer and autumn). Therefore, four seasonal separation principal components or factors were performed.

3   Results and discussion

3.1 Seasonal correlation of hydrochemical parameters

For the four seasons, high and positive correlations were observed for sodium, magnesium, potassium, calcium, chloride, sulphate and nitrate, which are responsible for water mineralization. Electrical conductivity (EC) also showed correlation with these variables, as well as DOC and HCO_3, except in autumn.

Water temperature (T^a) showed no correlations with other parameters in spring. A good negative correlation between T^a and mineral parameters (EC, Cl, HCO_3, pH) was found in winter. This could have occurred due to the dilution effects during the winter season. Very strong positive correlations between T^a and EC, Mg, K, Cl and DOC were also found in summer. This could be attributed to the influence of baseflow in the concentrations of mineral parameters and the faster decomposition of organic matter with higher temperatures in summer. The latter could also explain the significant relation for DOC and T^a in autumn.

Although DOC has good correlations with inorganic parameters in spring, summer and winter, such correlations became very poor in autumn, except for the sulphate. It is interesting to observe that DOC showed positive correlations with TKN in autumn, NO_3 in winter and with both nitrogen compounds in the other two seasons. A positive correlation between DOC and HCO_3 was found in spring but the correlation was negative in autumn, most likely due to the dilution effect during the autumn.

Finally, pH showed a positive relation with Ca, Mg, Na, Cl, SO_4 and NO_3 in spring and negative in autumn. In addition, in winter pH was also correlated with Na, Cl, EC and HCO_3. Good correlations also occurred with EC and HCO_3 in spring. Limited correlations were obtained in summer.
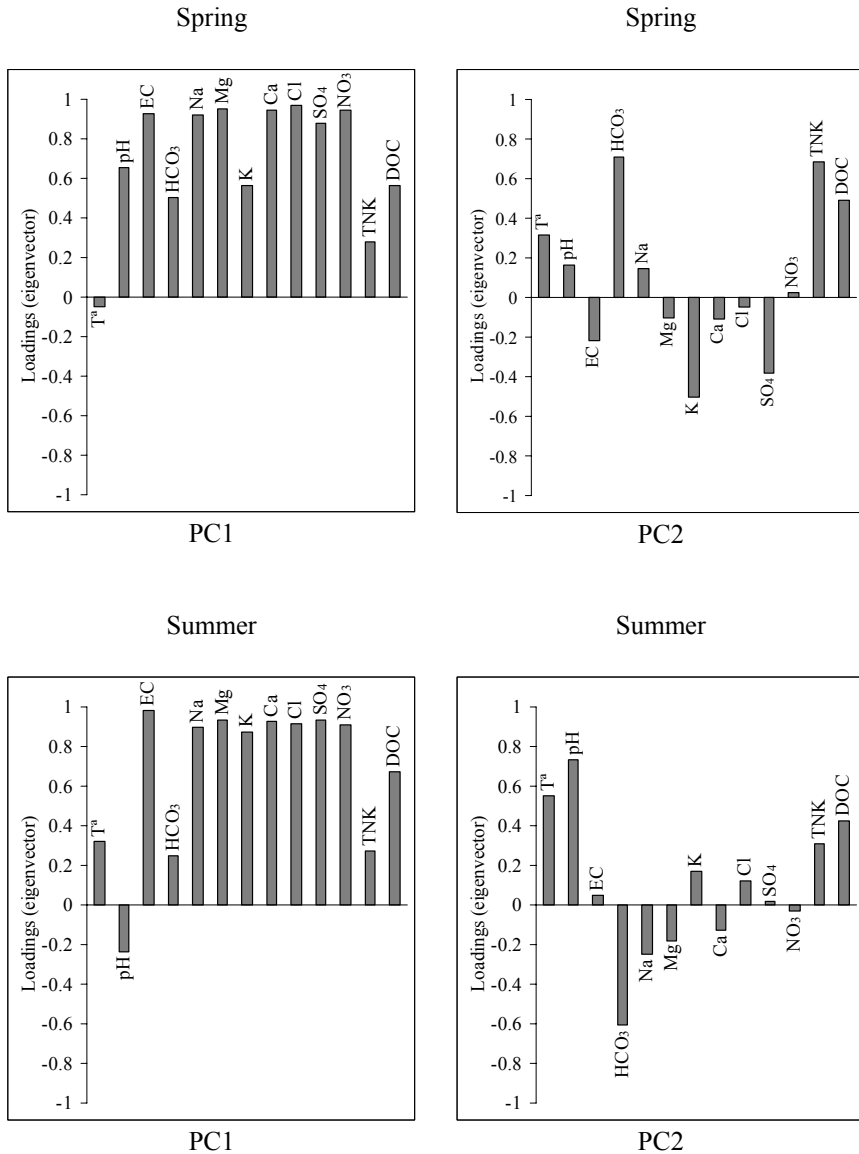
Figure 2:   Component loadings for the first component (PC1) and the second component (PC2) in spring and summer.
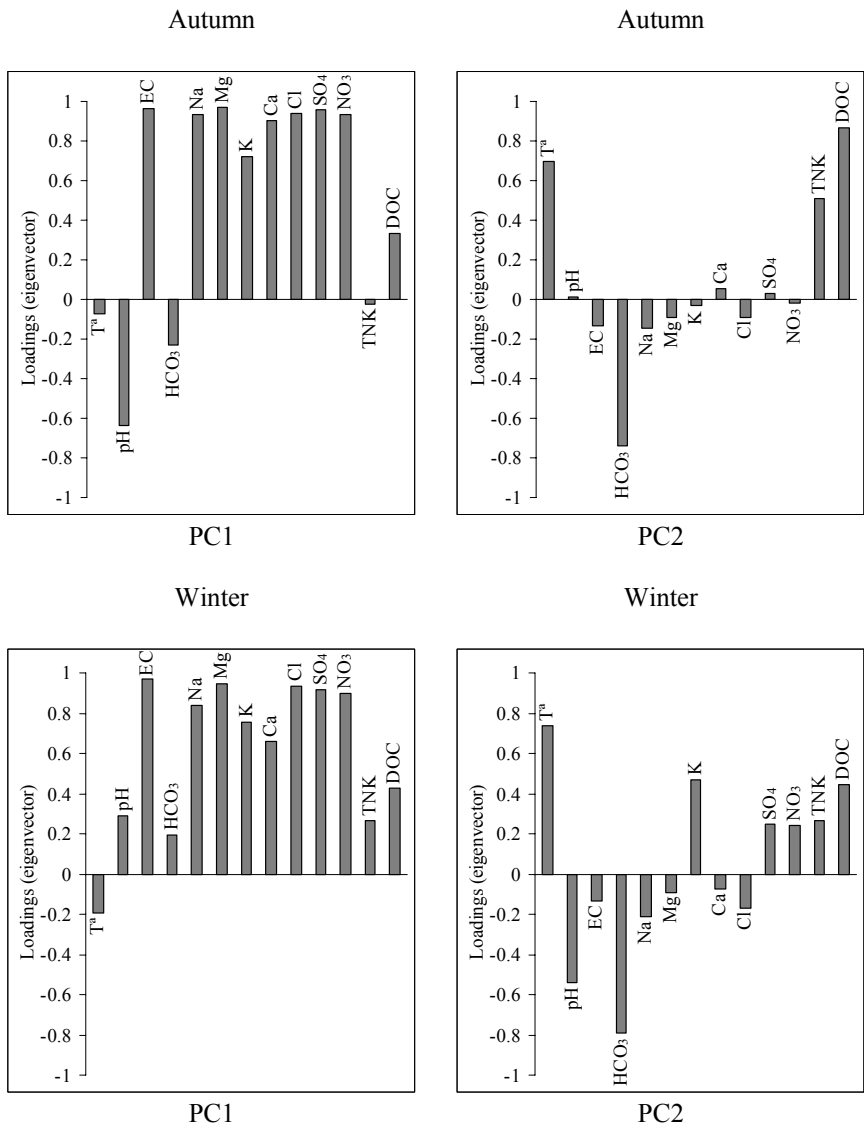
Figure 3:   Component loadings for the first component (PC1) and the second component (PC2) in fall and winter.

3.2 Identification of principal components

After applying a PCA to the original database for spring, summer, fall and winter, several components were distinguished. Four principal components explaining 88.7% of the total variance for spring and 83.3% for winter were obtained. In summer and autumn, three principal components accounted for 80.3 and 81.1% of the total variance, respectively.

Component loadings can be used to determine the relative importance of a variable as compared to other variables in a PC and do not reflect the importance of the component itself. They express the correlation between the original variables and the newly formed component.

The component loadings of the first two retained PCs for each season are presented in Figs 2 and 3. In spring, the principal component 1 (PC1) explained 57.8% of the variance and showed high contribution of most variables: chloride, electrical conductivity, magnesium, calcium, sulphate, nitrate, sodium, and to a lesser extent, bicarbonate, potassium and dissolved organic carbon. These variables were shown to correlate (data not shown). The second component (PC2) explained 14.1% of the variance and included total nitrogen (Kjeldahl) and dissolved organic carbon, and to a lesser extent, bicarbonate and potassium. The third component (PC3) (8.7% of the variance) included TKN (positive loading) and potassium (negative participation). Finally, PC4 explained 8.1% of the total variability of the original data and was highly influenced by temperature.

Similar component loading patterns were obtained for PC1 in summer (Fig. 2). That is, PC1 (which explained 58% of the total variance) was positively contributed to by EC, Mg, SO_4, Ca, Cl, NO_3, Na, K and DOC, whereas PC2 (which explained 12.6% of the total variance) was positively participated to by TKN, DOC and T^a and negatively impacted by HCO_3. Furthermore, PC3 explained 9.6% of the variance and included T^a, pH (positive loading) and TKN (negative participation).

In autumn, PC1, which accounted for 56.4% of the variance, had reliable positive contributions from Mg, EC, SO_4, Cl, NO_3, Ca and Na, in contrast to a significant negative contribution from pH. PC2, which represented 16.2% of the variance, was positively loaded by T^a and organic-related parameters (TKN and DOC) and negatively by HCO_3. Furthermore, PC3, which accounted for 8.4% of the variance, was positively influenced by T^a and pH and negatively by TKN.

In winter, PC1, which accounted for 49.9% of the total variance, was positively influenced by EC, Mg, Cl, SO_4, NO_3, Na, K and Ca. The second component, which explained 16.7% of the total variance, was positively correlated with DOC and negatively with pH and HCO_3. The third (8.8% of the variance) included TKN. Finally, PC4 explained 7.9% of the total variability and it was contributed by pH and DOC.

Vega et al. [6] and Ouyang et al. [11] had used exploratory data analysis to investigate the influence that seasonality and pollution have on the water quality of rivers. The first ones, in a study of two and half years (applying PCA to overall database), found that PC1 mostly contributed mineral parameters, similar to our results. However, PC1 in the study of Ouyang et al. [11], which applied PCA to four temporal databases (i.e., one database for each season), was largely participated by organic and physical parameters. These discrepancies could be due to characteristics of each catchment.

As seen in Fig. 2, PC1 is highly participated by most variables for all four seasons, thus hindering its hydrochemical interpretation, and therefore cannot explain which parameters are more important than others in influencing water

physical-chemical variations within a given season. A rotation of principal components can achieve a simpler and more meaningful representation of the underlying factors by decreasing the contribution to PCs of variables with minor significance and increasing the more significant ones.

Table 1 shows rotated factor correlation coefficients for the first factors in each season. In this study, parameters with correlation coefficients larger than 0.85 were taken into consideration in the interpretation. It must be noted that rotation has resulted in smaller groups of variables which can be now associated to individual rotated factors. The data in table 1 reveal that Cl, EC, Mg, SO_4, NO_3, Ca and Na were identified in the Corbeira stream for all four seasons, except the latter two variables in winter. On the other hand, K was an important parameter for water chemical variation in summer. In relation to organic parameters, only TKN in spring and winter and DOC in autumn contributed to water quality variation. Finally, temperature (physical parameter) just showed importance in spring.

Table 1: Rotated factor correlation coefficients for each season.

	T^a	pH	EC	HCO_3	Na	Mg	K	Ca	Cl	SO_4	NO_3	TKN	DOC
Spring													
Factor 1	-0.05	0.67	0.96	0.36	0.89	0.93	0.55	0.92	0.97	0.91	0.90	0.01	0.35
Factor 2	0.00	0.08	0.04	0.56	0.26	0.21	0.08	0.24	0.15	0.02	0.28	0.93	0.78
Factor 3	0.03	0.52	-0.02	0.54	0.23	-0.1	-0.78	-0.18	0.07	-0.24	-0.01	0.06	-0.09
Factor 4	0.98	-0.14	-0.09	0.23	0.07	0.02	-0.03	0.07	0.03	-0.21	0.07	-0.13	0.10
Summer													
Factor 1	0.3	-0.26	0.98	0.25	0.90	0.94	0.87	0.93	0.91	0.93	0.91	0.29	0.66
Factor 2	0.77	0.74	0.08	-0.28	-0.13	-0.10	0.10	-0.09	0.20	0.03	-0.12	-0.17	0.42
Factor 3	-0.16	0.20	-0.01	-0.67	-0.22	-0.14	0.19	-0.06	-0.06	0.02	0.15	0.83	0.16
Autumn													
Factor 1	-0.14	-0.61	0.98	-0.13	0.95	0.97	0.71	0.88	0.95	0.95	0.93	-0.12	0.22
Factor 2	0.82	0.11	0.04	-0.72	-0.02	0.02	-0.01	0.12	0.10	0.16	0.10	0.24	0.85
Factor 3	-0.26	-0.50	-0.12	-0.24	-0.01	0.06	0.22	0.17	-0.18	0.01	0.02	0.81	0.32
Winter													
Factor 1	-0.13	0.05	0.93	0.06	0.72	0.90	0.84	0.58	0.91	0.97	0.91	0.09	0.37
Factor 2	-0.74	0.56	0.27	0.82	0.35	0.25	-0.34	0.22	0.30	-0.11	-0.11	-0.16	-0.4
Factor 3	0.26	0.13	0.06	0.10	0.32	0.24	0.10	0.54	-0.02	0.01	0.08	0.89	0.05
Factor 4	-0.04	0.72	0.17	-0.06	0.21	0.03	-0.01	-0.17	0.16	0.06	0.23	0.14	0.72

These results reveal the influence of minerals (Cl, EC, Mg), inorganic (SO_4, NO_3) and organic parameters (TKN, DOC) in stream water quality variations. Minerals and inorganic parameters are present in all seasons. The mineral parameters may be interpreted as representing influences from natural inputs. The inorganic variables are likely the result of natural and anthropogenic sources. The natural origin of nitrates is related to the decomposition of organic matter, which is abundant in the catchment soils, whereas the anthropogenic origin is probably due to the septic tank systems and use of organic and inorganic fertilizers in agricultural areas of the catchment. These practices could

also explain the importance of TKN in spring, whereas in winter it is probably a consequence of surface runoff. In autumn, stream water showed strong influence of DOC, which could be transported with runoff caused by autumn precipitations. However, the results of this study should be interpreted with caution because most of the parameters studied could be affected simultaneously by other factors (e.g., flow) such as found Dieguez *et al.* [14] in a river catchment nearly to the study area.

4   Conclusions

The principal component analyses found a reduced number of principal components that explain most of the variance of the experimental dataset. A varimax rotation of these PCs revealed that chloride, conductivity, magnesium, sulphate and nitrate were always the most important variables contributing to water physical-chemical variations in the stream for all four seasons. The analyses also showed that there are other parameters (Ca, Na, K, DOC, TKN, Tª) that are important in contributing to water quality variation for one season but may not be for another.

The most important parameters obtained to evaluate variations in the stream water quality studied are mineral, inorganic and organic parameters. The mineral parameters may be interpreted as representing influences from natural inputs (alteration of schist and soils of catchment). The inorganic and organic parameters could be attributed to the influence of both natural and anthropogenic inputs.

Acknowledgements

References

[1]   Dixon, W. & Chiswell, B., Review of aquatic monitoring program design, Water Resources, **30**, pp. 1935-1948, 1996.
[2]   Davis, J.C., *Statistical and Data Analysis in Geology*, ed. John Wiley & Sons, New York, 1986.
[3]   Manly, B.F.J., *Multivariate Statistical Methods: A Primer*, Chapman & Hall, London, 1986.
[4]   Tabachnick, B.G. & Fidell, L.S., *Using Multivariate Statistics*, Allyn and Bacon, Boston, London, 2001.
[5]   Shine, J.P., Ika, R.V. & Ford, T.E., Multivariate statistical examination of spatial and temporal patterns of heavy metal contamination in New

Bedford Harbor marine sediments, *Environment Science of Technology*, **29**, pp. 1781-1788, 1995.

[6]   Vega, M., Pardo, R., Barrado, E. & Debán, L., Assessment of seasonal and pollution effects on the quality of river water by exploratory data analysis, *Water Research*, **32** (12), pp. 3581-3592, 1998.

[7]   Voutsa, D., Manoli, E., Smara, G., Sofoniou, M. & Stratis, I., A study of surface water quality in Macedonia, Greece: speciation of nitrogen and phosphorus, *Water Air and Soil Pollution*, **129**, pp. 13-32, 2001.

[8]   Gangopadhyay, S., Gupta, A.D. & Nachabe, M.H., Evaluation of ground water monitoring network by principal component analysis, *Groundwater*, **39**, pp. 181-191, 2001.

[9]   Bengraine, K. & Marhaba, T.F., Using principal component analysis to monitor spatial and temporal changes in water quality, *Journal of Hazardous Material*, **B100**, pp. 179-195, 2003.

[10] Ouyang, Y., Evaluation of river water quality monitoring stations by principal component analysis, *Water Research*, **39**, pp. 2621-2635, 2005.

[11] Ouyang, Y., Tkedi-Kizza, P., Wu, Q.T., Shinde, D. & Huang, C.H., Assessment of seasonal variations in surface water quality, *Water Research*, **40**, pp. 3800-3810, 2006.

[12] Jolliffe IT. *Principal Component Analysis,* Springer Series in Statistics, Springer-Verlag, New York, 1986.

[13] Haag, I. & Westrich, B., Processes governing river water quality identified by principal component analysis, *Hydrological Processes*, **16**, pp. 3113-3130, 2002.

[14] Dieguez Villar A., Taboada Castro, M.M: & Taboada Castro, M.T., Ionic concentrations in surface waters of an agricultural catchment during storm events, *Communications in Soil Science and Plant Analysis,* **33(15-18)**, pp. 2673-2691, 2002.