# Assessment of water quality of Asopos River in central Greece using multivariate analysis

**Tsadilas C.[1], Tziouvalekas M.[1], Evangelou E.[1,*], Papadopoulos S.[1], Papadoulis A.[1], Petsoulas C[1] , Tsitouras A.[1], Grypari P.[1] , Smyrnis T[2], Peppas A.[2]**

[1]Institute of Industrial and Forage Crops, 1 Theophrastos street, Larissa, Greece 41335
[2]ETME Peppas & Associates Ltd, Sinopis 43 Goudi, Athens, Greece 11527

*corresponding author: e-mail: levagel@gmail.com

**Abstract**
The main objective of this study was to assess the quality of water from five monitoring stations in Asopos River (central Greece), and evaluate the main factors that affect water quality. Fifty one biochemical parameters measured both in situ and in the laboratory three times a year for two consecutive years. Multivariate analysis of Hierarchical cluster analysis and Principal component analysis were used to interpret water quality characteristics. Cluster analysis grouped the samples in two main clusters (classes) corresponding to different main activities which affect water quality characteristics. The first cluster includes the two sampling stations located near the industrial areas and the second one the three stations located in areas affected by agricultural activities. Principal Component Analysis identified two factors which explain 84,5% of the variability of the original mean data set. The first factor explaining 50,5 % of the whole data variability related mainly to "chemical quality parameters", while the second factor explaining 34% of total data variability related mainly to "biological quality parameters". This study showed that multivariate statistical techniques proved effective in river water quality classification based on large and complex water quality data sets.

**Keywords:** water quality, industrial pollution, agricultural practices, multivariate analysis

## 1. Introduction

The quality of river water at any point reflects several major influences, including the lithology of the basin, atmospheric inputs, and climatic conditions and governed by both natural process and anthropogenic effects. Wastewater from agricultural, industrial and urban activities and often natural processes such as erosion and weathering could degrades water quality and impair their use for drinking, industrial, agricultural, recreational or other purposes.

Seasonal variations in precipitation, surface runoff, interflow, groundwater flow and pumped in and outflows coupled with spatial distributions of point and non point source pollution have a strong effect on river discharge and, subsequently, on pollutants concentration in surface water. Due to the spatial and temporal variations in water quality, a monitoring program has to provide a representative and reliable estimation of the surface waters quality.

The application of different multivariate statistical techniques such as cluster analysis (CA), and principal component analysis (PCA), helps in the interpretation of complex data matrices to better understand water quality and ecological status of the rivers, allows the identification of possible factors/sources that influence water systems and offers a valuable tool for reliable management of water resources as well as rapid solution to pollution problems. The aim of this study was to use a set of multivariate statistical techniques in order to interpret a complex river quality data set and to identify factors affecting Asopos river's water pollution.

## 2. Material and Methods

The Asopos basin, covering approximately 680 km$^2$, is located in Central Greece and is characterized by different land-uses. The river crosses several towns, agricultural areas and industrial zones, the biggest one being near the town of Oinophyta and receive waste waters originating from more than 400 industries including textile, metal processing, chemical, food, fertilizer, paint, tan-nery and pharmaceutical industries (Massoura, 2008).The rest of the Asopos area is characterized by urban settlements, intensive agricultural activities like vineyards, vegetables (e.g., celery, carrots and potatoes) and olive oil trees, as well as livestock breeding facilities, mainly of cattle, pigs and poultry. (Agelidis and Aloupi 2000). During a two year monitoring program supported by ministry of Rural Development and Foods 51 biochemical parameters (Table 1) measured both in situ and in the laboratory three times a year for two consecutive years in five monitoring stations through the riverbed. Hierarchical CA and PCA were used in the average values of seasonal measurements of the water quality data set. CA is a group of multivariate techniques whose primary purpose is to assemble objects based on the characteristics they possess. Cluster analysis classifies objects, so that each object to be similar to the others in the cluster with respect to a predetermined selection criterion. The resulting clusters of objects should then exhibit high internal (within-cluster) homogeneity and high external (between Cluster) heterogeneity. PCA is designed to transform the original variables into new, uncorrelated variables (axes), called the principal components, which are linear combinations of the original variables. The new axes lie along the directions of maximum variance. PCA applied to the normalized data to compare the compositional pattern between the analyzed water samples and to identify the factors that influenced each sample. Rotation of the axis defined by PCA produced a new set of factors, involving primarily subset of the original variables is divided into groups. CA used Ward's method with squared Euclidian distances as a measure of similarity among the monitoring stations where PCA used to quantify the significant of the variables that explain the observed groupings. New orthogonal variables (factors) explained by a reduced set of calculated factors are called Principal components (PCs).

**Table 1.** Biochemical parameters measured

| | |
|---|---|
| pH, EC, Alcalicity | $BOD_5$, COD, Disolved/saturated $O_2$ |
| $NO_2^-$, $NO_3$, $NH_4^+$ | Suspended Solids, Turbidity |
| $Cl^-$, $F^-$, | Redox, SAR |
| $Cr^{+6}$, $Cr^{+3}$, $Cr^{tot}$ | Water depth, Temperature |
| $Na^+$, $Mg^{+2}$, $Ca^{+2}$, | Total Dissolved Solids |
| Sn,Br,Al | Hardness(Total,Perm,Temp) |
| P, $K^+$, B, $Li^+$ | Chlorophyl, Secchi |
| Zn, Mn, Fe, Cu | $PO_4^{-3}$, $HCO_3^-$, $CO_3^{-2}$, $SO_4^{-2}$ |
| As, Pb, Hg, Ni,Co | |

## 3. Results and Discussion

The dendogram of the location pattern resulting from the CA (figure 1) shows that the monitoring stations may be grouped in two main clusters (groups). Cluster I formed by the 1st and 4th stations and cluster II by 2nd, 3d, and 5th, stations.

The 1st monitoring station is located in the "Oinofyta" industrial zone and water quality affected by waste water of a number of industries. The 4th station located near to the town of Thiva and water quality affected by agricultural activity and waste water of some industries near the town. All stations of the 2nd cluster located in areas affected mainly by agricultural activity. Results of PCA shows that up to 84,5% of the original data set variability is on the two first components. Thus almost all information about pollution in the five monitoring stations collected in the original 51 variables can be performed in the reduced space. The first factor explaining 50,5 % of the whole data variability named "Chemical Quality" factor as it is significant correlated ($r>0.8$, $p<0.05$) with mainly chemical parameters. In particular the first factor correlated with Cl-, Zn, TDS, EC, $Cr^{+6}$, Pb, $Cr^{+3}$ ,F- ,Mn, Total Hardnes, Ni, $Mg^{+2}$ $Ca^{+2}$ ,Co, Fe, Cu, Sn, Hg. The second factor explaining 34% of total data variability and named "Biological Quality" factor as it is correlated mainly to biological parameters. In particular the second factor correlated significantly with $BOD_5$, COD, Dissolved $O_2$, Saturation oercent $O_2$, Suspended Solids, and Turbidity. PCA provides information on the most meaningful parameters which describe the whole data set interpretation, data reduction and summarize the statistical correlation among constituent in the water with minimum loss of original information. CA and PCA provide a useful tool for optimizing future spatial network in Asopos river basin with lower cost. For example the number of monitoring sites could be reduced by selecting one site from cluster 2 in agricultural related monitoring stations. Moreover, from the 51 water quality parameters of the original data set almost all data variability could be represented by the 24 water quality parameters of the first two factors. The study illustrated the usefulness of multivariate statistical techniques for analysis and interpretation of complex data sets, in water quality assessment, identification of pollution sources/factors and understanding spatial variations in water quality for effective river water quality monitoring management.

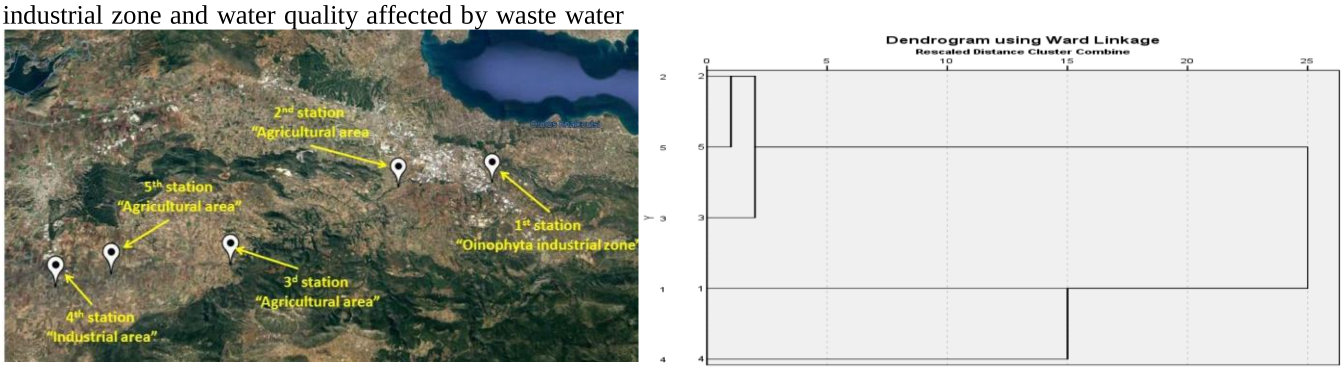


**Figure 1.** Monitoring stations of Asopos river and dendrogram showing spatial similarities of monitoring stations produced by cluster analysis. Monitoring stations 2-5-3 refers to agricultural activity areas. Monitoring stations 1and 4 refers to industrial activity areas


**References**

G. Massoura (2008). *Pollutant Hydrochemistry in Asopos River Basin.* (Dissertation) Agricultural University of Athens (in Greek with summery in English).

Angelidis, M.O., M. Aloupi (2000). *Geochemical study of coastal sediments influenced by river-transported pollution: Southern Evoikos Gulf, Greece.* Mar. Pollut. Bull. 77-82



**Acknowledgments**: This study was financed by the Ministry of Rural Development and Foods -Ministerial Decision No 2512/55958/19.05.2015.