*RESEARCH INVENTY: International Journal of Engineering and Science*
*ISSN: 2278-4721, Vol. 1, Issue 4 (October 2012), PP 6-12*
*www.researchinventy.com*

# Evaluation of Cauvery River Water Quality at Srirangapatna in Karnataka using Principal Component Analysis

Sudevi Basu[1] and Lokesh K. S[2]
*1 Lecturer, Department of Biotechnology, Sir M Visvesvaraya Institute of Technology, Bangalore – 562157, Karnataka, India,*
*2 Professor and Head, Department of Environmental Engineering, Sri Jayachamarajendra College of Engineering (Autonomous), Mysore – 560006, Karnataka, India.*

***Abstract:*** *Multivariate statistical techniques like Principal Component Analysis are used to interpret large and complex water quality data in Cauvery River at Srirangapatna. The data sets of 6 water quality monitoring stations which comprised of 10 water quality parameters monitored monthly over 12 years (2000-2011) are used for this study. PCA helped to identify the important water quality parameters responsible for the deterioration of water quality in river Cauvery. The main parameters included BOD, COD, TC, FC, Conductivity, TDS and Chlorides across baseline, impact and trend stations. PCA rendered a significant reduction in water quality data from 120 in each station to three or four principal components in baseline, impact and trend stations. Thus this study illustrates the significance of multivariate statistics in simplifying complex data sets of monitoring stations in river Cauvery at Srirangapatna and also helps in understanding the monitoring strategy for effective water quality management.*

***Keywords**: Multivariate* statistical techniques, Principal Component Analysis, Baseline station, Impact station, Trend station

## I. Introduction

Rivers form the lifeline of human society. These are vital freshwater systems of strategic importance across the world, providing main water resources for domestic, industrial, agricultural and recreational purposes. Rivers play a major role in assimilating or carrying industrial and municipal wastewater and runoff from agricultural fields. In recent years, rivers are amongst the most vulnerable water bodies to pollution as a consequence of unprecedented development. Thus the water quality of these water resources is a subject of ongoing concern and has resulted in an increasing demand for monitoring river water quality. The quality of water is described by its physical, chemical and microbiological characteristics. Therefore a regular monitoring of river water quality not only prevents outbreak of diseases and checks water from further deterioration, but also provides a scope to assess the current investments for pollution prevention and control.

In this study, seasonal variations of physico-chemical and bacteriological characteristics of water quality in Cauvery river was assessed in Srirangapatna town in Karnataka. The data was analyzed using Principal Component Analysis (PCA).

## II. Materials and Methods

### 2.1 Study Area

Srirangapatna is an island town, situated between the North and South branches of river Cauvery. It is located to the northeast of Mysore city at a distance of 15 Kms on the Bangalore – Mysore highway. The town has developed in two areas consisting of Patna area, which is like an urban area, and Ganjam, which resembles a typical village.

Though it is a town of medium population, the temples and historically significant monuments of this town attracts a large number of tourist people resulting in a very high floating population. Because of this reason the river Cauvery along Srirangapatna town stretch is prone to anthropogenic activities such as bathing, washing and disposal of wastes.

River Cauvery in this town divides into two major branches – north branch and south branch. There exists another small stream branch called Paschima Vahini river, almost parallel to south branch. These branches unite at a place called Sangama. The ground level in the town slopes from south branch towards north branch so that most of the storm and sewerage drains discharge into branch of river Cauvery. There are four stream monitoring stations and two drains located in this town stretch. Three of these stations are on the north Branch of the river and one station after the point of confluence of these branches. Two bathing ghats exist in this stretch. The stations are shown in Figure 1.

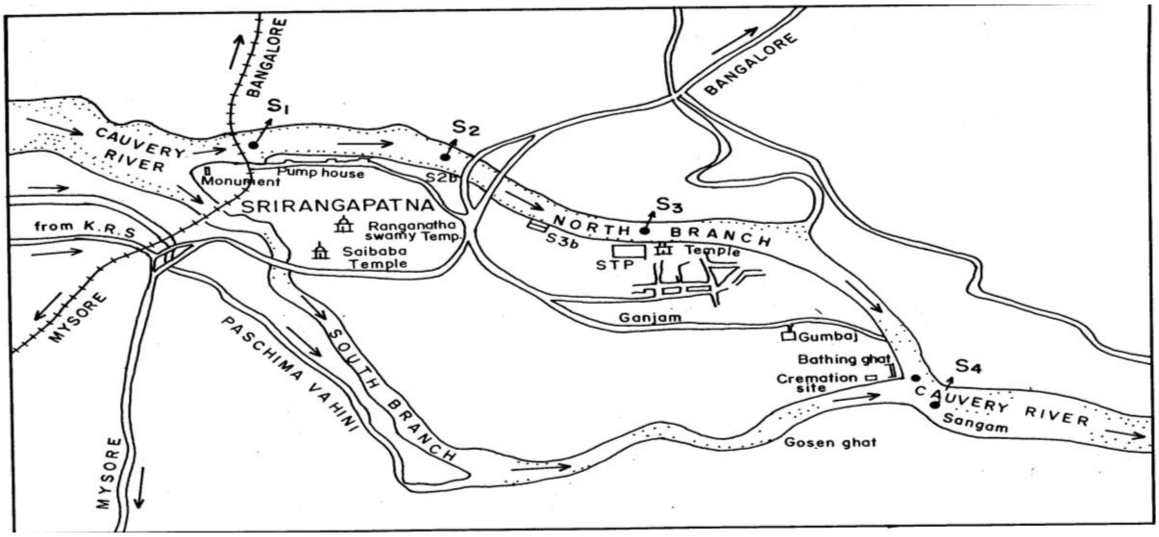


**Figure 1. Map of water quality monitoring stations at Srirangapatna town**

## 2.2 Monitoring Stations

**STATION - S1**

Station S1 is located on the north branch of the river, near the Bangalore – Mysore railway bridge. It is an upstream station and near this station water is being drawn for supply to the town.

**STATIONS - S2a and S2b**

The station S2a is located at a distance of about 150 m upstream of the Wellesly road bridge on the north branch of the river. This station is about 300 m downstream of station S1.

The station S2b is located on a drain that enters the river from the right bank just downstream of S2a. The flow in the drain is mainly comprised of sullage from Srirangapatna town.

**STATIONS - S3a and S3b**

The station S3a is an impact station and is positioned near the Nimishamba temple. It is downstream of the sewage disposal point, approximately 500 m from the station S2a. A bathing ghat exists near this Station. The station S3b is located on a relatively small drain that enters the river downstream of station S3a. The flow in the drain comprises mainly of wastewater from Ganjam village area of the town.

**STATION - S4**

This is a downstream station, located after the confluence of the north and south branches of the river Cauvery. A bathing ghat exists upstream of this station.

## 2.3 Data Preparation

The data sets of 6 water quality monitoring stations which comprised of 10 water quality parameters monitored monthly over 12 years (2000-2011) are used for this study. The data is obtained from the water Quality Monitoring work of Cauvery River Basin in Mysore District, Karnataka State assigned to Sri Jayachamarajendra College of Engineering, Mysore under a nationwide River Water Quality Monitoring Project of the National River Conservation Directorate (NRCD), Ministry of Environment and Forests, Government of India, under its National River Conservation Project (NRCP). Although there are more water quality parameters in these stations, only 10 most important parameters are chosen because of their continuity in measurement through the 12 years. The 10 selected water quality parameters include Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Chlorides (Cl), Total Dissolved Solids (TDS), Conductivity, Total Coliform (TC), Faecal Coliform (FC), Temperature and pH.

### 2.4 Principal Component Analysis

Principal Component Analysis (PCA) is a statistical method used to determine components that are linear combinations of the original variables (Garizi *et al,* 2011). It is a powerful pattern recognition technique that attempts to explain the variance of a large dataset of intercorrelated variables with a smaller set of independent variables or principal components (Simeonov *et al,* 2003). In PCA, a set of p correlated variables is transformed to a smaller set of uncorrelated hypothetical constructs called principal components. For this purpose, a covariance matrix or correlation matrix is used (Garizi *et al,* 2011). Mazlum *et al,* (1996) suggested that for water quality data analysis, a correlation matrix is used because variables are different in scale and equal in importance. An important aspect of PCA is the generation of eigenvalues which give a measure of the significance of the components: the components with the highest eigenvalues are the most significant. Also eigenvalues of 1.0 or greater are considered significant (Pejman *et al,* 2009; Garizi *et al,* 2011).

## III Results and Discussions

In the present study, PCA was applied to the six different sampling stations at Srirangapatna, each station consisting of 120 data (10 parameters for 12 years). The data was analyzed using MINITAB 15 Software Package.

### 3.1 PCA of Baseline Stations S1 and S2a

As seen in table 1, the first four PCs of station S1 show eigenvalues greater than 1, representing 88% of the total variance. Thus these four PCs are retained. The PC1 represents 41.8% of the total variance and is concerned mainly with BOD, TC and FC. PC2 has 19.8 % of the total variance and is concerned with conductivity and chlorides. PC3 having 13.3% of the total variance is again associated with TC. Also PC4 with 13.1% of the total variance is associated with DO.

The first four PCs of station S2a show eigenvalues greater than 1, representing 91.4% of the total variance. Thus these four PCs are retained. The PC1 represents 41.6% of the total variance and is concerned mainly with BOD, TC and FC. PC2 has 19.2 % of the total variance and is concerned with DO. PC3 having 18.2% of the total variance is associated with TDS, conductivity and chlorides. Further PC4 with 12.4% of the total variance is associated with COD and Temperature.

The scree plots of stations S1 and S2a are shown in figures 2 and 3 respectively. From the figures, it is clearly established that the first four eigenvalues are greater than 1 and hence significant.

From the four PCs in stations S1 and S2a, the variables that are significant include BOD, TC, FC, Chlorides and Conductivity. This shows that these stations are polluted from human activities. The baseline stations are polluted and are a serious concern as these stations are water collecting points for drinking water and hence treatment has to be undertaken for the water supply. This indicates the volume of pollution of water bodies like rivers.

**Table 1: PCA of Baseline Stations**

| **Variables** | **Baseline Stations** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **S1** | | | | **S2a** | | | |
| | PC1 | PC2 | PC3 | PC4 | PC1 | PC2 | PC3 | PC4 |
| DO | -0.210 | -0.562 | -0.237 | 0.187 | -0.195 | 0.551 | 0.047 | 0.380 |
| BOD | 0.426 | -0.287 | 0.193 | -0.039 | 0.438 | 0.070 | -0.308 | 0.097 |
| COD | 0.242 | -0.307 | -0.135 | 0.564 | 0.360 | 0.197 | 0.125 | 0.461 |
| CL | 0.218 | 0.273 | -0.509 | -0.024 | 0.244 | 0.147 | 0.456 | -0.057 |
| TDS | 0.394 | 0.030 | -0.348 | -0.156 | 0.370 | 0.136 | 0.367 | -0.194 |
| Conductivity | 0.217 | 0.581 | 0.005 | 0.141 | 0.098 | -0.548 | 0.351 | -0.203 |
| Temperature | 0.137 | 0.184 | 0.005 | 0.728 | 0.115 | -0.320 | 0.388 | 0.596 |
| pH | 0.298 | -0.200 | -0.503 | -0.229 | 0.248 | 0.426 | 0.220 | -0.443 |
| TC | 0.413 | -0.128 | 0.420 | -0.057 | 0.414 | -0.143 | -0.362 | 0.023 |
| FC | 0.434 | -0.074 | 0.282 | -0.122 | 0.434 | -0.084 | -0.303 | 0.007 |
| Eigenvalues | 4.1790 | 1.9782 | 1.3296 | 1.3051 | 4.1584 | 1.9158 | 1.8182 | 1.2361 |
| Proportion of Variance | 0.418 | 0.198 | 0.133 | 0.131 | 0.416 | 0.192 | 0.182 | 0.124 |

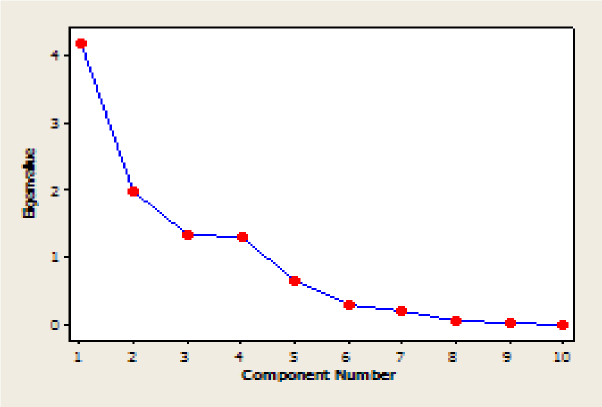

Fig 2: Scree Plot of eigenvalues in Station S1

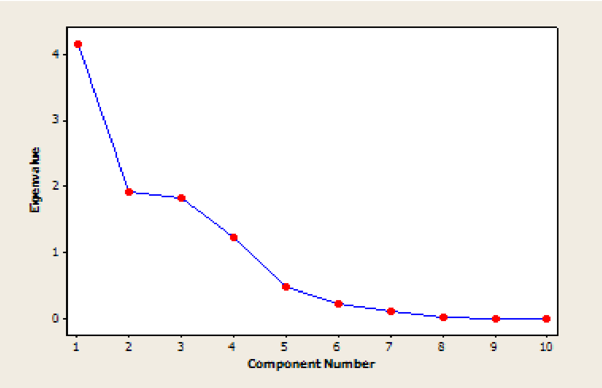

Fig 3: Scree Plot of eigenvalues in station S2a

### 3.2 PCA of Impact Stations S2b and S3b

As seen in table 2, the first three PCs of station S2b show eigenvalues greater than 1, representing 86.8% of the total variance. Thus these three PCs are retained. The PC1 represents 60.8% of the total variance and is concerned mainly with BOD, COD, Cl, Conductivity and TDS. PC2 has 15.8 % of the total variance and is concerned with TC and FC. PC3 having 10.2% of the total variance is associated with Temperature.

The first three PCs of station S3b show eigenvalues greater than 1, representing 91.3% of the total variance. Thus these three PCs are retained. The PC1 represents 54.5% of the total variance and is concerned mainly with BOD, COD, Cl, Conductivity and TDS. PC2 has 22.5 % of the total variance and is concerned with Temperature and pH. PC3 having 14.3% of the total variance is associated with TDS.

**Table 2: PCA of Impact Stations**

| **Variables** | **Impact Stations** | | | | | |
|---|---|---|---|---|---|---|
| | **S2b** | | | **S3b** | | |
| | PC1 | PC2 | PC3 | PC1 | PC2 | PC3 |
| DO | -0.372 | -0.038 | -0.102 | -0.117 | 0.625 | 0.119 |
| BOD | 0.360 | -0.314 | -0.111 | 0.373 | -0.265 | -0.130 |
| COD | 0.281 | -0.330 | -0.226 | 0.411 | 0.032 | 0.004 |
| CL | 0.380 | -0.124 | 0.114 | 0.400 | -0.142 | 0.021 |
| TDS | 0.342 | -0.258 | -0.325 | 0.405 | -0.069 | 0.194 |
| CONDUCTIVITY | 0.384 | 0.133 | 0.061 | 0.417 | -0.086 | 0.052 |
| TEMP | 0.277 | -0.116 | 0.170 | 0.309 | 0.441 | 0.073 |
| pH | -0.177 | -0.026 | -0.848 | 0.256 | 0.524 | 0.085 |
| TC | 0.267 | 0.568 | -0.178 | 0.137 | 0.065 | -0.676 |
| FC | 0.254 | 0.595 | -0.157 | 0.001 | 0.179 | -0.678 |
| Eigenvalues | 6.0775 | 1.5829 | 1.0225 | 5.4453 | 2.2461 | 1.4326 |
| Proportion of Variance | 0.608 | 0.158 | 0.102 | 0.545 | 0.225 | 0.143 |

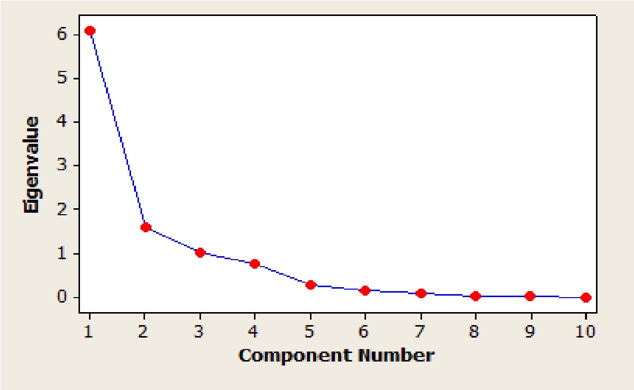


Fig 4: Scree Plot of eigenvalues in station S2b

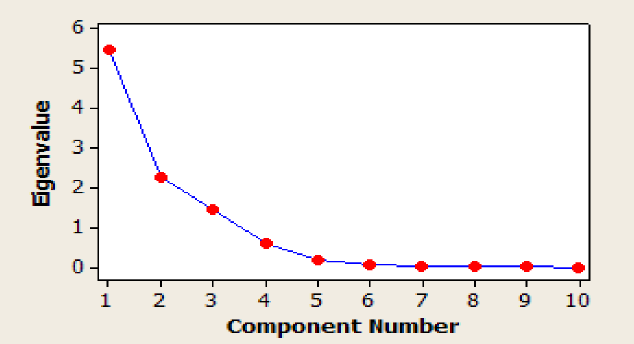


Fig 5: Scree Plot of eigenvalues in station S3b

The scree plots of stations S2b and S3b are shown in figures 4 and 5 respectively. From the figures, it is clearly established that the first three eigenvalues are greater than 1 and hence significant.

The PCs in impact stations include BOD, COD, TC, FC, TDS, Chlorides, Conductivity. These are the drains that enter the river and pollute the water downstream. This is a major concern as the quality gets deteriorated over time and huge amounts are invested for drainage works and treatment units to protect the quality of river water. The pollution in river water quality raises questions on effectiveness of treatment units.

### 3.3 PCA of Trend Stations S3a and S4

As seen in table 3, the first three PCs of station S3a show eigenvalues greater than 1, representing 80.3% of the total variance. Thus these three PCs are retained. The PC1 represents 41.8% of the total variance and is concerned mainly with BOD, COD, TC and FC. PC2 has 21.4 % of the total variance and is concerned with DO. PC3 having 17.1% of the total variance is associated with DO, TDS and pH.

**Table 3: PCA of Trend Stations**

| **Variables** | **Trend Stations** | | | | | |
|---|---|---|---|---|---|---|
| | **S3a** | | | **S4** | | |
| | PC1 | PC2 | PC3 | PC1 | PC2 | PC3 |
| DO | 0.052 | 0.472 | 0.414 | 0.092 | 0.511 | -0.289 |
| BOD | 0.440 | 0.095 | -0.294 | 0.412 | -0.255 | -0.267 |
| COD | 0.461 | -0.026 | 0.103 | 0.337 | 0.336 | 0.091 |
| CL | 0.134 | -0.218 | 0.227 | 0.190 | 0.245 | 0.388 |
| TDS | 0.320 | -0.101 | 0.484 | 0.433 | 0.173 | 0.113 |
| CONDUCTIVITY | -0.022 | -0.621 | 0.059 | -0.016 | -0.232 | 0.572 |
| TEMP | 0.094 | -0.545 | 0.183 | 0.251 | -0.047 | 0.542 |
| pH | 0.336 | 0.159 | 0.415 | 0.372 | 0.324 | -0.075 |
| TC | 0.408 | 0.013 | -0.409 | 0.343 | -0.442 | -0.129 |
| FC | 0.428 | -0.034 | -0.263 | 0.407 | -0.338 | -0.174 |
| Eigenvalues | 4.1765 | 2.1392 | 1.7100 | 3.8007 | 2.4474 | 2.1824 |
| Proportion of Variance | 0.418 | 0.214 | 0.171 | 0.380 | 0.245 | 0.218 |

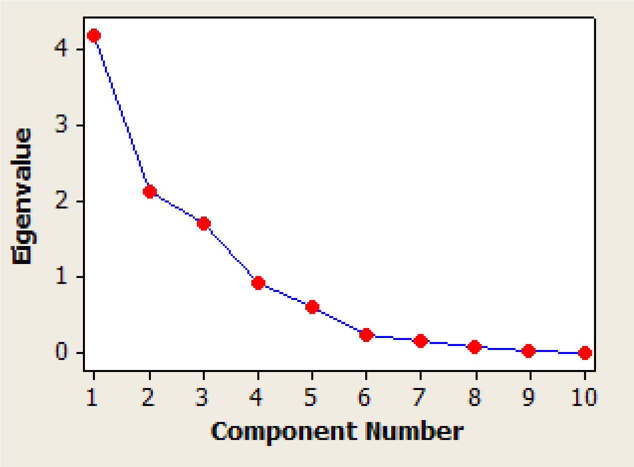


Fig 6: Scree Plot of eigenvalues in station S3a

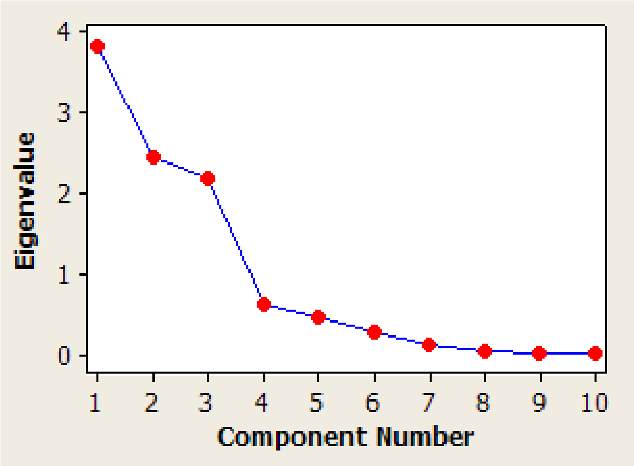


Fig 7: Scree Plot of eigenvalues in station S4

The first three PCs of station S4 show eigenvalues greater than 1, representing 84.3% of the total variance. Thus these three PCs are retained. The PC1 represents 38% of the total variance and is concerned mainly with BOD, COD, TDS, pH, TC and FC. PC2 has 24.5 % of the total variance and is concerned with DO, COD and pH. PC3 having 21.8% of the total variance is associated with Conductivity and Temperature.

The PCs in trend stations include BOD, COD, TDS, pH, TC and FC. The wastes from bathing ghat enter the river in this stretch and pollute the river. Hence all the monitoring stations like baseline, impact and trend are polluted from human activities. The scree plots of stations S3a and S4 are shown in figures 6 and 7 respectively. From the figures, it is clearly established that the first four eigenvalues are greater than 1 and hence significant.

Similar work of evaluation of river water quality using PCA has been carried out by several researchers like Mazlum *et al*, (1999); Praus *et al*, (2007); Li *et al*, (2009); Kumar *et al*, (2010) and Thareja *et al*, (2010).

## IV. Conclusions

Multivariate statistical techniques like Principal Component Analysis are used to interpret large and complex water quality data in Cauvery River at Srirangapatna. PCA helped to identify the important water quality parameters responsible for the deterioration of water quality in river Cauvery. The main parameters included BOD, COD, TC, FC, Conductivity, TDS and Chlorides across baseline, impact and trend stations. PCA rendered a significant reduction in water quality data from 120 in each station to three or four principal components. Thus this study illustrates the significance of multivariate statistics in simplifying complex data sets of monitoring stations in river Cauvery at Srirangapatna and also helps in understanding the monitoring strategy for effective water quality management. Huge amounts are invested for drainage works and treatment units to protect the quality of river water. The pollution in river water quality raises questions on effectiveness of these treatment units.

## V. Acknowledgement

The authors thank National River Conservation Directorate (NRCD), Ministry of Environment and Forests MoEF, Government of India for the financial support under National River Conservation Directorate Project.

## References

[1]. Garizi A. Z., Sheikh V. and Sadoddin A., (2011), "Assessment of seasonal variations of chemical characteristics in surface water using multivariate statistical methods", International Journal of Environmental Science and Technology, Vol. 8, No. 3, pp. 581-592.
[2]. Kumar M. and Singh Y., (2010), "Interpretation of Water Quality Parameters for Villages of Sanganer Tehsil, by Using Multivariate Statistical Analysis", Journal of Water Resources and Protection, Vol. 2, pp. 860-863.
[3]. Li Y, XU L and Li S., (2009), "Water Quality Analysis of the Songhua River Basin Using Multivariate Techniques", Journal of Water Resource and Protection, Vol. 2, pp 110-121.
[4]. Mazlum N., Ozer A. and Mazlum S., (1999), "Interpretation of Water Quality Data by Principal Components Analysis",, Turkish Journal of Engineering and Environmnetal Science, Vol. 23, pp. 19-26.
[5]. Pejman A.H., Bidhendi G. R. N., Karbassi A. R., Mehrdadi N. and Bidhendi M.E., (2009), "Evaluation of spatial and seasonal variations in surface water quality using multivariate statistical techniques", International Journal of Environmental Science and Technology, Vol. 6, No. 3, pp. 467-476.
[6]. Praus P., (2007), "Urban Water Quality Evaluation using Multivariate Analysis", Acta Montanistica Slovaca, Vol. 12, No. 2, pp. 150-158.
[7]. Simeonov V., Stratis J.A., Samara C., Zachariadis G., Voutsa D., Anthemidis A., Sofoniou M. and Kouimtzis Th., (2003), "Assessment of the surface water quality in Northern Greece", Water Research, Vol. 37, pp. 4119-4124.
[8]. Thareja S., Trivedi P., (2010), "Assessment of Water Quality of bennithora River in Karnataka through Multivariate Analysis", Nature and Science, Vol. 8, No. 6, pp 51-56.