International Journal of Environmental Engineering and Management
ISSN 2231-1319 Volume 8, Number 1 (2017), pp. 27-38

http://www.ripublication.com

Multivariate Statistical Analysis of Seasonal Variations of Water Quality In River Godavari at Polavaram, A.P.

Rajesh Kumar M, Olive Eunice P, Sreejani Tp*, Sravya Pvr, Srinivasa Rao Gvr*

JRF, Civil Engg. Department, Andhra University, Visakhapatnam, India - 530003
**Corresponding author*

Abstract

An attempt has been made in this present work, to study the seasonal variation of water quality of River Godavari at Polavaram in Andhra Pradesh using Multivariate Statistical techniques viz., Cluster analysis (CA), Principal Component Analysis (PCA) and Factor Analysis (FA) applying SPSS 20.0 software. The Water Quality Indices (WQI) are computed using WAIWQI and NSFWQI methods. The data obtained from the Central Water Commission, Hyderabad, Andhra Pradesh, for a period of decade, i.e. 2002-2012 is taken into consideration for conducting the study. The data comprised of 20 Physico-chemical parameters and 7 biological parameters. From the Principal Component Analysis (PCA) and the Factor Analysis (FA) of the water quality data of the study area, the principal parameters affecting the quality of water during the study period are obtained. The Cluster Analysis (CA) conducted on temporal basis displayed clusters of the years based on their similarity. When this Cluster Analysis is compared with the Water Quality Indices determined using WAIWQI and NSFWQI methods, it is found that the clusters vis-à-vis the WQI are in good correlation. The study has shown that the water quality of river Godavari at Polavaram is not biologically contaminated and the surface runoffs are causing a moderate change in the water quality w.r.t physical and chemical characteristics.

Keywords: Polavaram, Godavari, Cluster Analysis (CA), Principal Component Analysis (PCA), Factor Analysis (FA), SPSS.

1. Introduction

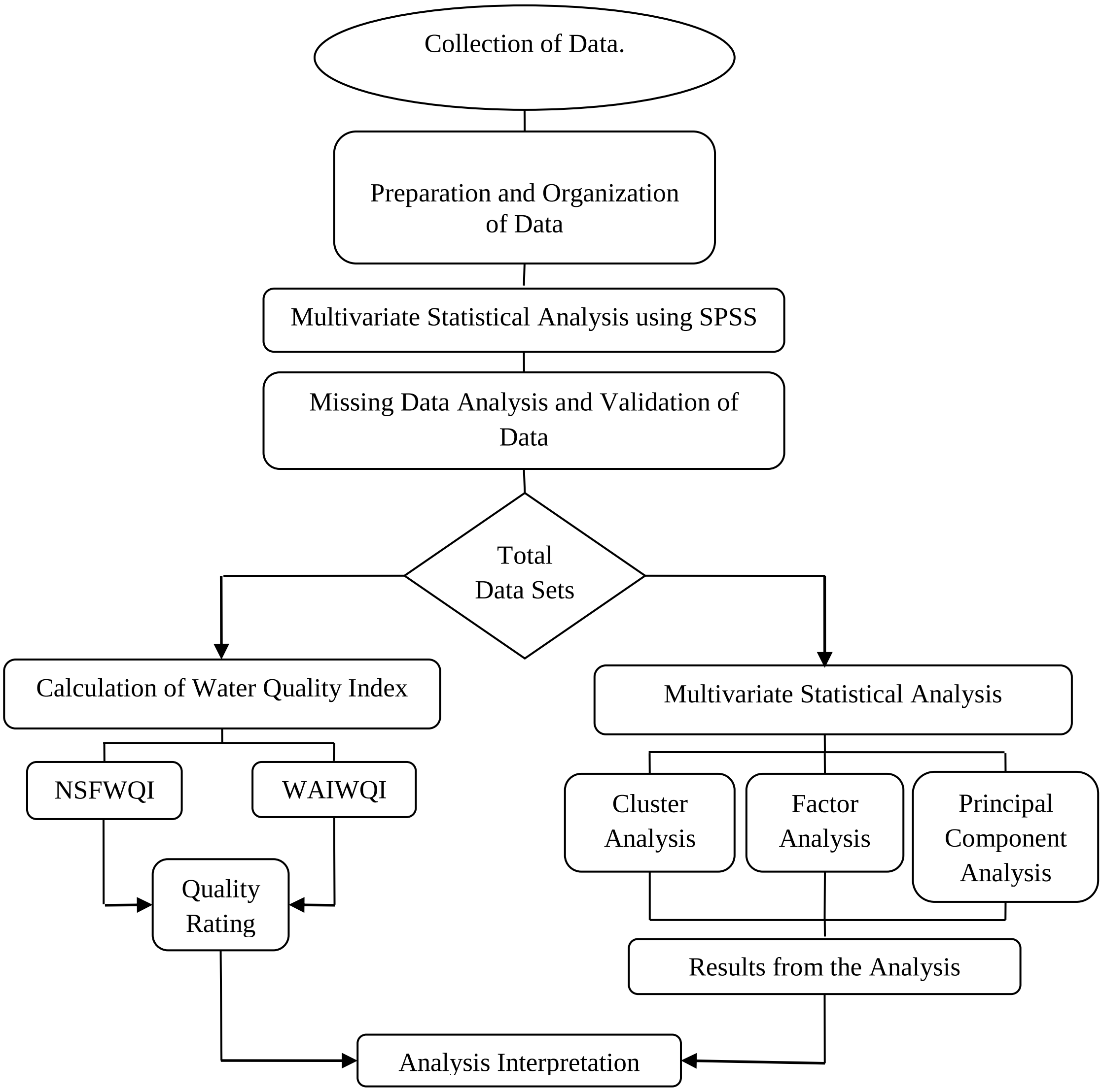
Water is one of the vital resources required to sustain life and is a key resource in all economic activities ranging from agriculture to industry. With the increasing pressure of human population, there developed severe stress on water resources. Both the natural processes, such as precipitation inputs, erosion, weathering of crustal materials, as well as the anthropogenic influences like urban, industrial and agricultural activities, increasing exploitation of water resources, together determine the quality of surface water in a region. At the same time surface waters are most vulnerable to pollution due to their easy accessibility for disposal of wastewaters. Rivers play a major role in assimilation or carrying off the municipal and industrial wastewater and run-off. This is a seasonal phenomenon and is largely affected by climate in the basin. Water quality refers to the physical, chemical, and biological characteristics of water. The quality of surface water is an important parameter in the determination of water usage for different types of activities. Therefore, it is necessary to monitor and evaluate water quality on regular basis. Determination of a large number of physico-chemical and biological parameters is usually conducted resulting in a large data matrix, which needs a complex data interpretation. The data sets contain rich information about the behaviour of the water resources. The overall process of evaluation of the physical, chemical and biological nature of the water ultimately results in the overall water quality assessment. Assessment of surface water quality can be a complex process undertaking multiple parameters capable of causing serious effects on overall water quality. The impact of anthropogenic activities on Rivers as a result of the increasing rate of urbanization is of a great concern due to the fact that water from these rivers is the source of water distributed for public use. These rivers are of particular importance in the study of surface water pollution because effluents from industries, municipal waste, agricultural and urban run-off are discharged into it thereby deteriorating the quality (Amadi . A.N[1] et.al 2010, Himangshu Shekhar Mandal[4] 2011). The multivariate statistical techniques served as an excellent exploratory tool in analysis and interpretation of complex data set on water quality and in understanding their temporal and spatial variation. There is a need for integrated approach where spatial analysis is one of the most important aspects. Hence, the study by Mohd Saiful Samsudin[8] et.al 2011 illustrates the environmetric techniques for analysis and interpretation of complex data, water quality assessment, identification of pollution sources, and investigating spatial variations of water quality as an effort toward a more effective river basin management. Various multivariate statistical techniques were adopted to develop WQI using different water quality parameters. Hierarchical cluster analysis used to classify the total data set into major groups based on similarity of water quality. Suitable discriminant function was generated using DA to develop the WQI. The information obtained through this work may be used to improve the management practices of the environment considering the selected water quality constituents

(Sangeeta Pati [9]et.al 2012). The population growth, economic development, with the consequent anthropogenic activities and global climate change pose to reduce the quality trends of surface water resources (Asama A. Agrama[2] 2012, Manjusha Bhor[7] et.al 2013). Continuous monitoring is useful for the sustainable development through planning and for the implementation of remediation methods in the future, in order to mitigate the adverse effects of the poor quality of water on human health, as well as on plant growth. The correlation and multiple regression analysis applied to the datasets indicated their interrelationships, for evaluating water quality during the pre monsoon, monsoon, and post monsoon seasons.(Hemant Pathak[3] 2013). Multivariate statistical techniques, such as cluster analysis (CA), Principal component analysis (PCA), Factor analysis (FA) and Discriminant analysis (DA), use for the evaluation of temporal/spatial variations and the interpretation of a large complex variable surface water quality data set in water quality for effective River water quality management. (Jinal Patel[5] 2015).

2. Study Area

Godavari is the largest river of all peninsular rivers in India. It has a total drainage area of 312,812 sq. km of which 48.6% lies in Maharashtra, Telangana 18.8%, 4.5% in Andhra Pradesh, 10.9% in Chattisgarh, 5.7% in Orissa, 10.0% in Madhya Pradesh (MP) and 1.4 % in Karnataka. Principal tributaries of Godavari in AP are Pranhita, Indravati and Sabari. These tributaries contribute 80% flow of the total river in AP. The average yearly water flow in Godavari is nearly 110 billion cubic meters. Polavaram is a mandal in West Godavari district of Andhra Pradesh. It is about 35 km away from the banks of Godavari River at Papi Hills of Eastern Ghats. It consists of 23 villages, 4 in plain area and 19 in scheduled area polavaram village also fall in the mandal. The Geddapalli panchayat and its villages are in deep forest, which is 20 km away from plains.

3. Methodology



4. Results and Discussions

4.1 Water Quality Indices

WQI values are evaluated taking into consideration the monthly averages of 13 Physico-chemical parameters and 7 biological parameters during the period 2002-2012, using WAIWQI and NSFWQI method. To observe the variation in WQI, the WQI w.r.t the three seasons viz., pre-monsoon, monsoon and post-monsoons are evaluated. The results are tabulated in the following table 1.

Table 1. Seasonal variations of WAIWQI w.r.t Physico-chemical and Biological parameters for 2002-12

Season	**WAIWQI**				**NSFWQI**			
	Physico-chemical	Quality Rating	Biological	Quality Rating	Physico-chemical	Quality Rating	Biological	Quality Rating
Pre-monsoon	54.04	Poor	29.61	Good	53.09	Medium	61.62	Medium
Monsoon	51.29	Poor	23.94	Excellent	55.07	Medium	52.16	Medium
Post-monsoon	60.68	Poor	23.71	Excellent	58.47	Medium	58.97	Medium

4.2 Cluster Analysis

Cluster analysis (CA) is used to classify variables based on their similarity level. Hierarchal CA was performed using Ward's method and Squared Euclidean Distance to form a combination of clusters. The result is illustrated by dendrogram, presenting the clusters and their proximity. The cluster analysis is conducted taking into consideration 20 physico-chemical and 7 biological parameters. The dendrograms for the physico-chemical and biological parameters for a period of ten years w.r.t the seasonal variations (monsoon, pre-monsoon and post-monsoon) are shown in the following fig 1 to 6. The dendrograms are constructed using ward linkage method and are drawn for rescaled distance cluster combine. The water quality parameters for a particular year are considered as one case and in total 10 cases are used representing ten years of study period, for constructing dendrograms.

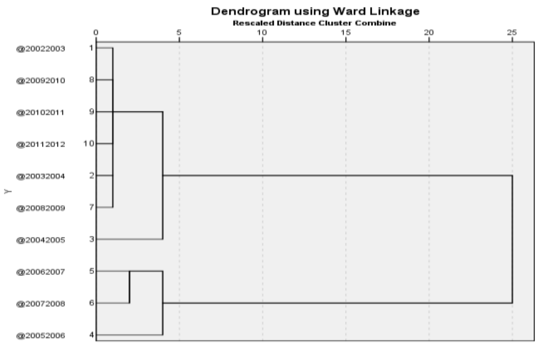


Figure 1: CA for Physico-chemical parameters (Pre-monsoon seasons) (2002-2012)

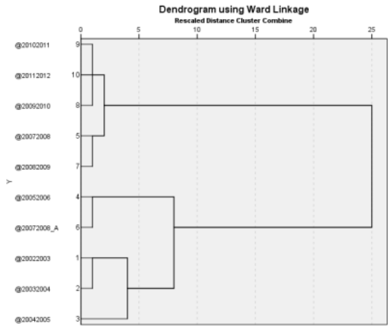


Figure 2: Dendrogram: Biological parameters (Pre-monsoon seasons) (2002-2012)

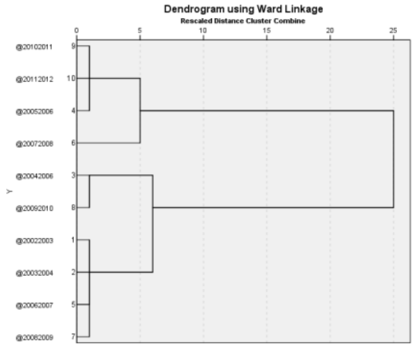


Figure 3: Dendrogram: Physico-chemical parameters (Monsoon seasons) (2002-2012)

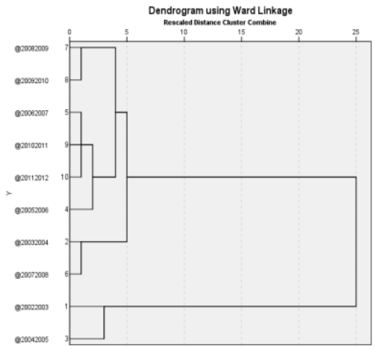


Figure 4: Dendrogram: Biological parameters (Monsoon seasons) (2002-2012)

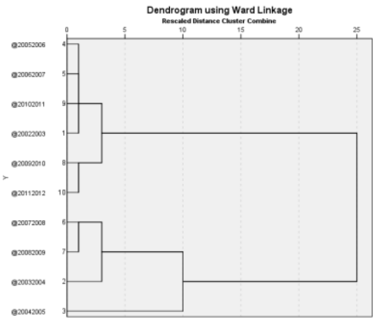


Figure 5: Dendrogram: Physico-chemical parameters (Post-monsoon seasons) (2002-2012)

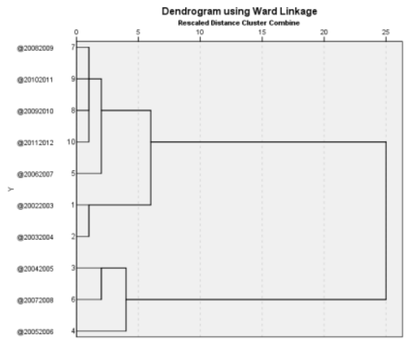


Figure 6: Dendrogram: Biological parameters (Post-monsoon seasons) (2002-2012)

3 Principal Component Analysis

PCA is conducted for an accurate evaluation of the clustering behaviour. Principal components (PC) with Eigen values greater than one are only extracted as components. PCs having Eigen values more than unit value are represented as factor loadings. Factor loading > 0.75, 0.75 – 0.5 and < 0.5 can be classified as strong, moderate and weak respectively. (Liu[6] et al. (2003)). PCA is conducted taking into consideration 20 physico-chemical and 7 biological parameters and the results are tabulated in tables 2 and 3. The results presented in Table 2 shows that 6 principal components are generated explaining 69.20% of the total variance in the water data sets. The principal components having high factor loadings are as follows: PC1: EC, Alk-Tot, HCO_3, PC2: Alk-Phenols, CO_3, PC3: NO_2+NO_3, NO_3-N, PC5: NO_2-N. These parameters have high factor loadings which indicate that these parameters are in high concentrations and are the major pollutants among other parameters. PC4 and PC6 displayed from moderate to weak loadings.

Table 2. PCA of Physico-chemical parameters (2002-2012): Rotated component matrix

Parameters	**Component**					
	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
EC_GEN (µmho/cm)	.767	.365	.183	.348	.007	-.060
pH_GEN (pH units)	.283	.729	-.024	-.061	.024	.008
Temp (deg C)	-.414	-.182	.102	.122	.076	.473
ALK-TOT (mgCaCO3/L)	.895	.311	-.033	.165	-.031	.032
Alk-Phen (mgCaCO3/L)	.162	.937	.019	.016	-.082	-.066
Ca (mg/L)	.750	.224	.014	-.363	.007	-.223
Cl (mg/L)	.265	.192	.611	.364	.024	-.041
CO3 (mg/L)	.130	.939	.019	.044	-.131	-.032
F (mg/L)	.101	.018	.092	.016	.669	-.437
Fe (mg/L)	-.315	-.078	-.142	.620	-.165	-.039
HCO_3 (mg/L)	.946	.025	-.045	.168	.012	.047
K (mg/L)	.065	-.029	.193	.529	-.374	.118
Mg (mg/L)	.309	.176	-.087	.705	.093	.239
Na (mg/L)	.461	.193	.549	.303	.050	-.275
NO2+NO3 (mg N/L)	-.076	-.095	.898	-.015	-.067	.102
NO2-N (mgN/L)	-.071	-.042	.020	-.054	.812	.277
NO3-N (mgN/L)	-.123	-.097	.836	-.169	.096	-.045
o-PO4-P (mg P/L)	.033	-.043	-.032	-.027	.002	.697
SiO2 (mg/L)	.049	.459	-.025	.043	.092	-.063
SO4 (mg/L)	.244	.018	.132	.679	.148	-.268
Eigenvalues	3.722	2.949	2.322	2.211	1.361	1.276
% of Variance	18.612	14.745	11.611	11.056	6.805	6.378
Cumulative %	18.612	33.357	44.968	56.024	62.829	69.207

Table 3: PCA of Biological parameters (2002-2012) Rotated component matrix			
Parameters	**Component**		
	PC1	**PC2**	**PC3**
BOD3-27 (mg/L)	.052	.001	.910
COD (mg/L)	.000	.741	-.324
DO (mg/L)	.979	.032	.047
DO_SAT% (%)	.981	-.002	.031
NH3-N (mg N/L)	-.027	-.749	-.316
Eigenvalues	1.924	1.112	1.036
% of Variance	38.482	22.243	20.720
Cumulative %	38.482	60.725	81.445

Principal component analysis w.r.t biological parameters vide Table 3 shows that three principal components are generated explaining 81.44% of the total variance in the water data sets. The principal components having high factor loadings are as follows: PC1: DO (mg/l) and DO_SAT%, PC3: BOD_3-27. These parameters have high factor loadings which indicate that these parameters are in high concentrations and are the major pollutants among other parameters.

4.3 Factor Analysis

FA follows principal component analysis. The main purpose of factor analysis is to reduce the contribution of less significant variables and to simplify even more the data structure coming from the principal component analysis. This purpose can be achieved by rotating the axis defined by principal component analysis according to well established rules, and constructing new variables, also called varifactors. As a result, a small number of factors will usually account for approximately the same amount of information as do the much larger set of original observations (Shrestha and Kazama[10], 2007). FA is conducted taking into consideration 20 physico-chemical and 7 biological parameters and the results are tabulated in tables 4 and 5.

Table 4 FA w.r.t Physico-chemical parameters (2002-2012) : Rotated factor matrix				
Parameters	**Factor**			
	1	**2**	**3**	**4**
EC_GEN (µmho/cm)	.613	.365	.607	.114
pH_GEN (pH units)	.298	.589	.052	-.027
Temp (deg C)	-.389	-.214	-.013	.042
ALK-TOT (mgCaCO3/L)	.764	.319	.454	-.122
Alk-Phen (mgCaCO3/L)	.146	.977	.061	.024
Ca (mg/L)	.855	.241	-.156	.019
Cl (mg/L)	.162	.163	.452	.488
CO3 (mg/L)	.089	.982	.092	.014
F (mg/L)	.232	-.013	-.043	.148
Fe (mg/L)	-.354	-.074	.340	-.115
HCO3 (mg/L)	.807	.031	.469	-.140
K (mg/L)	-.112	.020	.451	.081
Mg (mg/L)	.039	.142	.713	-.146
Na (mg/L)	.391	.194	.437	.480
NO2+NO3 (mg N/L)	-.077	-.104	.097	.814
NO2-N (mgN/L)	.012	-.130	-.107	.048
NO3-N (mgN/L)	-.015	-.122	-.101	.793
o-PO4-P (mg P/L)	-.094	-.061	.019	-.079
SiO2 (mg/L)	.126	.289	.039	-.007
SO4 (mg/L)	.128	.030	.554	.112
Eigen values	3.030	2.825	2.397	1.895
% of Variance	15.151	14.126	11.983	9.474
Cumulative %	15.151	29.276	41.259	50.733

The factor matrix vide table 4 generated four significant factor loadings which explained 50.73% of the variance in data sets. The following parameters are indicated as high factor loadings: Factor 1: Alk-Tot, Ca, HCO_3, Factor 2: Alk- phenols, CO_3, Factor 4: NO_2+NO_3, NO_3-N. These parameters have high factor loadings which indicate that these parameters are in high concentrations and are the major pollutants among other parameters.

Table 5 FA w.r.t Biological parameters (2002-2012) : Rotated factor matrix			
	Factor		
Parameters	**1**	**2**	**3**
BOD3-27 (mg/L)	.054	.000	.326
COD (mg/L)	.002	.362	-.161
DO (mg/L)	.953	.064	.126
DO_SAT% (%)	.958	-.001	.086
NH3-N (mg N/L)	-.026	-.371	-.147
Eigen values	1.946	1.110	1.016
% of Variance	36.609	5.460	3.533
Cumulative %	36.609	42.069	45.602

The factor analysis shown in table 5, generated four significant factor loadings which explained 45.60% of the variance in data sets. The following parameters are indicated as high factor loadings: Factor 1: DO, DO_Sat%. These parameters have high factor loadings which indicate that these parameters are in high concentrations and are the major pollutants among other parameters.

5. Conclusions

1. From the Principal Component Analysis (PCA) and the Factor Analysis (FA) of the water quality data of the study area vide table 2, 3, 4 and 5 it is observed that, the carbonates, bi-carbonates, alkaline total, alkaline phenols and nitrite nitrogen are the principal parameters affecting the quality of water during the study period.
2. The Cluster Analysis (CA) conducted on temporal basis w.r.t physico-chemical and biological parameters fig 1 to 6 is compared with the water quality indices determined using WAIWQI and NSFWQI methods, it is found that the clusters formed vis-à-vis the WQI are in good correlation.
3. From the study related to WQI and the Cluster Analysis it is observed that the water quality is found to be medium w.r.t physico-chemical parameters during different seasons and w.r.t biological parmeters, it is found that the water is good in all the seasons.
4. Therefore, it can be concluded that, the water quality of the river Godavari at Polavaram is not biologically contaminated and the surface runoffs are causing a moderate change in the water quality w.r.t physical and chemical characteristics.

References

[1] Amadi A.N., Olasehinde. P.I, Okosun. E.A, Yisa.J (2010)., Assessment of the

Water Quality Index of Otamiri and Oramiriukwa Rivers, *Science Publications Physics International* 1 (2): 116-123.

[2] Asmaa A. Agrama (2012), Assessment of Surface Water Quality in Egypt, *Journal of Applied Sciences Research*, 8(4), 1944-1951.

[3] Hemant Pathak (2013), Water Quality Studies of Two Rivers At Bundelkhand Region, Mp, India: A Case Study**,** *U.P.B. Sci. Bull., Series B*, Vol. 75, Iss. 2, 81-90.

[4] Himangshu Shekhar Mandal and Amrita Das (2011), Assessment of seasonal variation in physico-chemical characteristics and quality of Torsha River water for irrigation used in Cooch Behar and Jalpaiguri districts of West Bengal, India, *J. Chem. Pharm. Res*., 2011, 3(6):265-270.

[5] Jinal Patel and Minakshi Vaghani, Assessment Of Surface Water Quality Using Multivariate Statistical Techniques – A Review, *International Journal of Advanced Research in Engineering, Science, Management*, ISSN: 2394-1766.

[6] Chen-Wuing Liu, Kao-Hung Lin, Yi-Ming Kuo, (2003), Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan, *The Science of the Total Environment* 313, 77–89.

[7] Manjusha Bhor, Prakash Kadave,Abhijit Bhor, Sheetal Bhor, Manisha Bhosale, Bholay A.D, (2013), Water Quality Assessment of the River Godavari, At Ramkund, Nashik, (Maharashtra), India, Research Inventy: *International Journal Of Engineering And Science*, Issn: 2278-4721, Vol. 2, Issue 2, Pp 64-68.

[8] Mohd Saiful Samsudin, Hafizan Juahir, Sharifuddin M. Zain, Nur Hazirah Adnan, (2011) Surface River Water Quality Interpretation Using Environmetric Techniques: Case Study at Perlis, River Basin, Malaysia, *IJEP* Vol.1 No.5, PP.1-8.

[9] Sangeeta Pati. M.K.Dash, C.K.Mukherjee, (2012), Development of Water Quality Index for assessment of quality of water in the coastal water of Bay of Bengal at Visakhapatnam zone, India, The Macrotheme Review, *A Multidisciplinary Journal of Global Macro Trends*, Volume 1, Issue 1, 85-91.

[10] Srestha.S and Kazama.F, (2007), Assessment of surface Water Quality Using Multivariate Statistical Techniques: A Case Study of the Fuji River Basin, Japan, *Environmental Modelling & Software* 22, 464-475.