American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-6, Issue-8, pp-112-120
www.ajer.org
Research Paper Open Access

Statistical Modelling of Water Quality of River Godavari at Rajahmundry and Dhawaleswaram, AP.

T.P.Sreejani[1]*, M. Rajesh Kumar[2], P.V.R.Sravya[3], G.V.R. Srinivasa Rao[4]

[1]JRF, Civil Engineering Dept., Andhra University, Visakhapatnam, India
[2] Research Scholar, Civil Engineering Dept., Andhra University, Visakhapatnam, India
[3]Asst. Professor, Civil Engineering Dept., ANITS, Visakhapatnam, India
[4]Professor of Civil Engineering Dept., Andhra University, Visakhapatnam, India
Corresponding Author: T.P.Sreejani

ABSTRACT**:** *An attempt has been made in this present work, to study the seasonal variations of water quality of River Godavari for the period of 2004-2012 at Rajahmundry and Dhawaleswaram located on the banks of River Godavari in Andhra Pradesh. This water quality modelling is done with the help of Water Quality Indices (WQI) computed using NSFWQI and WAIWQI methods. Multivariate statistical techniques namely Factor Analysis (FA), Cluster analysis (CA), and Principal Component Analysis (PCA) are used to conduct the statistical modelling of river water quality in the study area using SPSS 20.0 software. The study using the Water Quality Indices (WQI) indicated that the quality of water is found to be moderate w.r.t physico-chemical parameters and poor w.r.t biological parameters. The WQI values projected for the future period i.e. 2014, 2015, 2016 indicated no perceptible change in the quality of water. The Multivariate Analysis resulted in the study has shown the highest pollution loadings for the Monsoon and Post-monsoon seasons. The variation in DO_Sat% is found to vary the degree of pollution as per the variances obtained from the Multivariate Analysis. Therefore, it is concluded that the quality of River Godavari at Rajahmundry and Dhawaleswaram is moderate at present and its quality in the future might not have a perceptible change if the present conditions prevail and unchanged.*

Keywords: *Godavari River, Rajahmundry, Dhawaleswaram, Water Quality Index (WQI), Factor Analysis (FA), Cluster Analysis (CA), Principal Component Analysis (PCA).*

Date of Submission: 02-07-2017 Date of acceptance: 14-08-2017

I. INTRODUCTION

Earth is the only known planet in this universe where life is possible because of the availability of water and oxygen. Water is most important necessity of life for all the living beings on the earth. The water quality and human health are closely related. The quality of water is now being recognized as a major concern as it is getting deteriorated by human activities, day by day. The water quality of a river can be predicted by using the water quality modeling (WQM). WQM can be conducted in so many ways such as analytical, numerical, and statistical. The statistical method is the most popular one which utilizes a good number of multivariate techniques like Factor Analysis (FA), Principal Component Analysis (PCA), Cluster analysis (CA) etc. The water quality in both pre- and post-monsoon periods throughout the year can be characterized. The evaluation of concentrations of major ions as a part of Hydro chemical analysis of water will enable the researcher for WQI based assessment. (Tripathy and Sahu[9] (2005)). Multivariate statistical techniques emphasizes the significance of monitoring networks in detecting water pollution and provides a useful tool that help the decision makers in determining the extent of pollution via practical pollution indicators. (Hulya Boyacioglu[3] et.al, (2005)). Also these methods illustrates the usefulness of multivariate statistical techniques for analysis and interpretation of complex data sets, and in water quality assessment, identification of pollution sources/factors and understanding temporal/spatial variations in water quality for effective river water quality management. (Shrestha and Khazama[7] (2006)). In order to get better information about the quality of surface water of large and complex databases, there is need in usage of multivariate statistical techniques.

(Arzu altin[1] et.al, (2008)). Application of these type of analysis allows the classification of data and investigates the patterns of water quality and pollution sources in different seasons and the influence of water quality parameters. (Prakash raj Kennel[6] et.al, (2010)), 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, one can obtain functions that will categorize future new samples according to the pollution status or water quality condition which makes the study significant to design easy and effective monitoring programs of river water quality, using parameters that indicate organic pollution. (Soledad[9] Olive Gonzalez et.al, (2011)). Along with the assessment tools, different National and International Agencies involved in water quality assessment and pollution control defines water quality criteria for different uses of water considering different indicator parameters, so there are numerous WQI specific to any region or area. Different water quality indices developed worldwide, however, it is difficult to recognize a unique water quality index for assessing surface water quality of any nation or area with a definitive solution. (Bharti and Katyal[2] (2011)). The application of chemometrics allows data reduction in the dimensionality of large data set, delineating a few indicator parameters responsible for variations in the water quality. (Indrani Gupta[4] et.al. (2013)). Multivariate statistical techniques were effective analytical techniques for classification and processing of large datasets of water quality and the identification of major sources of water pollution in tropical pastures. It is possible to plan for future sampling events by optimizing the number of sampling stations and selecting appropriate water quality parameters, and reduce affiliated recurring costs. However, it should be mentioned that extra sampling stations must be included in water quality monitoring program to assess the spatial variations of water quality of the catchment in further studies. (Majid Ajorlo[5] et.al, (2013)).

II. STUDY AREA

River Godavari is a major waterway in central India, originating in the Western Ghats at Trimbakeshwar, in the Nasik District of Maharastra, India and flowing eastward by 1465 km across Deccan plateau through the state of Maharastra, Telangana, and Andhra Pradesh. It enters Telangana at Basar in Adilabad district. Rajahmundry is the Second largest city on the banks of River Godavari in Andhra Pradesh. At Rajahmundry, the Godavari is in its widest form having a width of approximately 5 km from Rajahmundry to the other bank at Kovvur. The study was conducted at the areas Rajahmundry, Dhawaleswaram and Polavaram, located on the banks of River Godavari. Rajahmundry is located at 16.98° N and 81.78° E with an average elevation of 14 m MSL (45 feet). Dhawaleswaram located at 16.57°N and 81.48°E is a suburb situated near Rajahmundry in the East Godavari district of Andhra Pradesh. It is approximately four kilometers away from Rajahmundry. Sir Arthur Cotton built a barrage across river Godavari at Dhawaleswaram. The barrage provides water for agriculture purpose for both the East and West Godavari districts.

III. METHODOLOGY

3.1 Collection of data:

The data was collected from Irrigation & CAD Department, Hydrology Project Circle, Hyderabad, Govt. of Andhra Pradesh for Rajahmundry and Dhawaleswaram areas for a period of ten years i.e., from 2002 to 2012. Then it is divided as per the seasons considered viz. pre-monsoon (March - June), Monsoon (July - October), Post-monsoon (November - February). The data include physico-chemical and biological parameters.

3.2 Calculation of Water quality indices:

Water Quality Indices (WQI) are calculated in both Weighted Arithmetic Water Quality Index (WAIWQI) and National Sanitation Foundation Water Quality Index (NSFWQI) methods as given below.

a. $WAIWQI = \sum Q_i W_i / \sum W_i$ (1)

The quality rating scale (Q_i) for each parameter is calculated by using the following equation

$Q_i = 100*\{(V_i - V_o) / (S_i - V_o)\}$ (2)

Where, V_i =estimated concentration of i^{th} parameter in the analysed water

V_o =The ideal value of pure water

S_i =Standard value of i^{th} parameter

The unit weight (W_i) for each water quality parameter is calculated by using the formula

$W_i = K / S_i$ (3)

Where, K = proportionality constant $=1/\sum(1/S_i)$

b. $NSFWQI= \sum SI_i$ (4)
Where $SI_i= w_i \times q_i$ (5)
SI_i = sub-index of i^{th} parameter
w_i= relative weight and is calculated as
$w_i=1/s_i$
si= standard values of parameters
q_i = quality rating of i^{th} parameter
c_i = experimental value

3.3 Multivariate Statistical Analysis

SPSS 20.0 is used to conduct Factor Analysis (FA), Cluster Analysis (CA) and Principal Component Analysis (PCA) are considered in this study for the water quality modelling. The methodology of Multivariate Statistical Analysis is as follows.

3.3.1. Missing value incorporation and validation of data

Missing data occur when no data value is stored for the variable in an observation. The collected data is given as input to the software to obtain the missing value patterns. Then the variables are classified into categorical and quantitative variables. From this the missing value patterns are analyzed. T-test performed to give the final dataset as output.

Validation of data aims at identification of suspicious and invalid cases and data variables in a data set. It produces - lists of variables, cases, and data values that fail various checks, counts of violations of variables, checks the cross variable rules, gives descriptive summaries of analysis of variables and ignores the weight variable specification. The data obtained from the missing value analysis is given as input for data validation. Duplicate cases are identified in the process and descriptive statistics are analyzed to conduct few statistical tests to validate the data.

3.3.2. Factor Analysis (FA)

Factor analysis is a method for investigating whether a number of variables of interest are linearly related to a smaller number of unobservable factors. FA is a process in which the values of observed data are expressed as functions of a number of possible factors in order to reduce the co-relational data to a smaller number of dimensions or factors. A correlation matrix (rectangular array of the correlation coefficients) is generated for all the variables. Factors are extracted from correlation matrix based on the correlation coefficients of the variables. The factors are rotated in order to maximize the relationship between the variables and some of the factors.

3.3.3. Cluster Analysis (CA)

This is a statistical technique used to divide the data into group cases and further into homogenous sub groups based on responses to variables. Hierarchical clustering is adopted in the present study since it is the most commonly used method and computes the proximity matrix of the difference or similarity of every case with every other case.

3.3.4. Principal Component Analysis (PCA)

PCA is a process in which a new variable is formed from the set of existing variables. PCA is Eigen vector based multivariate analysis. It transforms the data in such a way that the greatest variance by any projection of the data comes to lie on the first component and the second greatest variance on the second component and so on.

IV. RESULTS AND DISCUSSIONS

The seasonal WQI values are evaluated using NSFWQI and WAIWQI methods considering 18 physico-chemical 3 and biological parameters at both Rajahmundry and Dhawaleswaram stations.

Table 1: Seasonal Variation of WQI values at Rajahmundry and Dhawaleswaram

WQI	NSFWQI				WAIWQI			
Study Area	Rajahmundry		Dhawaleswaram		Rajahmundry		Dhawaleswaram	
Parameters	Phyisco-Che	Biological	Phyisco-Che	Biological	Phyisco-Che	Biological	Phyisco-Che	Biological

Seasonal Variations								
Pre-Monsoon	48.55	66.51	46.60	69.14	42.70	68.32	40.54	65.34
Quality Rating	Fair	Med.	Fair	Med.	Good	Poor	Good	Poor
Monsoon	51.05	66.61	53.50	71.86	47.77	68.10	47.42	70.37
Quality Rating	Med.	Med.	Med.	Good	Good	Poor	Good	Poor
Post-Monsoon	52.15	70.76	67.24	69.19	49.63	72.30	62.76	68.58
Quality Rating	Med.	Good	Med.	Med.	Good	Poor	Poor	Poor

The following tables vide no. 2,3,4,5,6,7,8,9 depict the results of descriptive statistics from the FA and PCA, w.r.t Physico-chemical and Biological parameters respectively.

Table 2 Descriptive Statistics (FA) w.r.t physico-chemical parameters considering months as cases.

Cases	Rajahmundry			Dhawaleswaram		
	Mean	Std. deviation	Analysis (N)	Mean	Std. deviation	Analysis (N)
Mar	32.33	53.59	18	29.00	47.68	18
Apr	30.27	49.63	18	28.50	46.23	18
May	26.05	42.37	18	27.94	46.85	18
Jun	28.61	47.00	18	26.77	45.42	18
Jul	26.94	44.53	18	26.77	46.22	18
Aug	22.72	36.77	18	24.33	40.40	18
Sep	27.66	45.81	18	27.16	44.75	18
Oct	**37.55**	**63.56**	18	33.27	57.70	18
Nov	37.11	62.99	18	**39.22**	**66.57**	18
Dec	28.44	48.04	18	31.05	52.83	18
Jan	31.44	52.50	18	36.77	61.76	18
Feb	32.00	52.72	18	33.11	56.31	18

Table 3 Descriptive Statistics (FA) w.r.t biological parameters considering parameters as cases.

Cases	Rajahmundry			Dhawaleswaram		
	Mean	Std. deviation	Analysis (N)	Mean	Std. deviation	Analysis (N)
BOD_{3-27} (mg/L)	2.00	0.60	12	2.47	0.27	12
DO (mg/L)	7.00	0.60	12	7.79	0.21	12
DO_SAT (%)	**92.00**	**6.09**	12	**102.46**	**3.39**	12

From the above tables vide table 2 and 3, it was observed that the highest mean and standard deviation values are obtained for the month of October (monsoon period) and November (post-monsoon) and the parameter DO_Sat% influences the quality of water.

Table 4 Descriptive Statistics (PCA) w.r.t physico-chemical parameters considering months as cases.

Cases	Rajahmundry			Dhawaleswaram		
	Mean	Std. Deviation	Analysis (N)	Mean	Std. Deviation	Analysis (N)
Mar	32.33	53.59	18	29.00	47.68	18
Apr	30.28	49.63	18	28.50	46.23	18
May	26.05	42.37	18	27.94	46.85	18
Jun	28.61	47.00	18	26.77	45.42	18
Jul	26.94	44.53	18	26.77	46.22	18
Aug	22.72	36.77	18	24.33	40.40	18
Sep	27.66	45.81	18	27.16	44.75	18
Oct	**37.55**	**63.56**	18	33.27	57.70	18
Nov	37.11	62.99	18	**39.22**	**66.57**	18
Dec	28.44	48.04	18	31.05	52.83	18
Jan	31.44	52.50	18	36.77	61.76	18
Feb	32.00	52.72	18	33.11	56.31	18

Table 5 Descriptive Statistics (PCA) w.r.t biological parameters considering parameters as cases

Cases	Rajahmundry			Dhawaleswaram		
	Mean	Std. deviation	Analysis (N)	Mean	Std. deviation	Analysis (N)
$BOD_{3\text{–}27}$ (mg/L)	2.00	0.60	12	2.47	0.27	12
DO (mg/L)	7.00	0.60	12	7.79	0.21	12
DO_ SAT% (%)	92.00	6.09	12	102.46	3.39	12

From the tables 4 and 5, it was observed that the highest mean and standard deviation values are obtained for the month of October (monsoon period) and November (post-monsoon) and the parameter DO_Sat% influences the quality of water.

From the analysis, the percent of variance attributable to each factor and the cumulative variance of the factor and the previous factors are shown. The higher the absolute value of the loading, the more the factor contributes to the pollution. The highest positive loadings obtained contribute the maximum pollution. Whereas the component matrix obtained from the PCA gives the highest pollution components.

Table 6 Factor Matrix (Factor Analysis of physico-chemical parameters at Rajahmundry and Dhawaleswaram)

Cases	Rajahmundry			Dhawaleswaram		
	Factors			Factors		
	1	2	3	1	2	3
Mar	0.998	0.030	-0.027	0.999	-0.026	0.009
Apr	1.000	-0.020	-0.001	0.999	-0.026	0.026
May	0.998	0.048	0.031	0.999	0.033	0.025
Jun	0.998	0.043	-0.004	0.999	0.018	-0.009
Jul	0.997	-0.073	0.012	0.999	0.004	0.004
Aug	0.994	-0.105	0.035	0.991	-0.125	0.040
Sep	**0.999**	**0.025**	**0.029**	0.996	0.005	-0.078
Oct	0.997	-0.028	-0.057	0.989	0.145	0.033
Nov	0.998	0.038	-0.032	0.998	0.037	-0.003
Dec	0.999	-0.038	-0.007	**1.000**	**0.009**	**0.013**
Jan	0.990	0.131	0.019	0.998	-0.015	-0.053
Feb	0.998	-0.050	0.000	0.999	-0.041	-0.006

The higher the absolute value of the loading, the more the factor contributes to the pollution. The highest positive loadings obtained for the month of September (Monsoon period) is 0.999 in table 6. Hence, the degree of pollution is maximum in the Monsoon period with highest loadings. The highest positive loadings obtained for the month of December (Post-Monsoon period) is 1.000 in table 6. Hence, the degree of pollution is maximum in the Post-Monsoon period with highest loadings.

Table 7 Factor Matrix (Factor Analysis of biological parameters at Dhawaleswaram)

Parameters	Factors	
	1	2
BOD3-27 (mg/L)	0.224	0.536
DO (mg/L)	0.931	-0.278
DO_SAT% (%)	**0.987**	**0.141**

The highest positive loadings obtained for DO_Sat% is 0.987 in table 7. Hence, the degree of pollution is maximum because of the parameter DO_Sat% with highest loadings. The factor matrix was not extracted since the software attempted to extract 2 factors, in iteration 25, the communality of a variable exceeded 1.0. Extraction was terminated. In the similar way the analysis was also carried out in the PCA.

Table 8 Component Matrix (PCA of physico-chemical parameters at Rajahmundry and Dhawaleswaram)

Cases	Rajahmundry	Dhawaleswaram
	Component	Component
	1	1
Mar	0.998	0.999
Apr	**1.000**	0.999
May	0.998	0.999

Jun	0.998	0.999
Jul	0.997	0.999
Aug	0.994	0.991
Sep	0.999	0.996
Oct	0.997	0.989
Nov	0.998	0.998
Dec	0.999	**1.000**
Jan	0.990	0.998
Feb	0.998	0.999

Since only one component was extracted from PCA, the higher the absolute value of the loading, the more the factor contributes to the pollution. The highest positive loadings obtained for the month of April (Pre-Monsoon period) is 1.000 in table 8. Hence, the degree of pollution is maximum in the Monsoon period with highest loadings. The highest positive loadings obtained for the month of December (Post-Monsoon period) is 1.000 in table 8. Hence, the degree of pollution is maximum in the Post-Monsoon period with highest loadings.

Table 9 Component Matrix (PCA of biological parameters at Rajahmundry and Dhawaleswaram)

Parameters	Component	Component
	1	1
BOD3-27 (mg/L)	-0.756	0.364
DO (mg/L)	0.934	0.929
DO_SAT% (%)	**0.965**	**0.976**

The highest positive loadings obtained for DO_Sat% is 0.965 and 0.976 in table 9. Hence, the degree of pollution is maximum because of the parameter DO_Sat% with highest loadings. In the Cluster Analysis, the first outcome of the analysis is proximity matrix. The proximity matrix table shows the squared Euclidean distance of each pair of cases. The values on the diagonal are 0, since a case does not differ from itself. The agglomeration schedule table shows the change in the distance measure is as additional cases are merged into clusters. The column labeled coefficients has the values of the distance statistic used to form the cluster; and a good cluster sees a sudden jump in the distance coefficient. The solution before the gap indicates the good solution. The dendrogram plot shows how far (or close) cases were when they were combined. The length of the branch (or link) shows how far apart each case is from the other cases in its cluster. The plot rows represent each case on the y-axis and the x-axis is a rescaled distance coefficients. Cases with low distance/ high similarity are close together.

Table 10 Agglomeration Schedule (Cluster Analysis of physico-chemical parameters at Rajahmundry)

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	2	10	81.50	0	0	7
2	4	7	197.00	0	0	4
3	8	9	443.00	0	0	11
4	3	4	699.50	0	2	6
5	1	12	979.50	0	0	7
6	3	5	1347.25	4	0	9
7	1	2	1833.50	5	1	8
8	1	11	2826.55	7	0	10
9	3	6	4333.20	6	0	10
10	1	3	8078.90	8	9	11
11	1	8	17928.41	10	3	0

The agglomeration schedule in table 10 shows the change in the distance measure is as additional cases are merged into clusters. The column labeled coefficients has the values of the distance statistic used to form the cluster; and a good cluster sees a sudden jump in the distance coefficient. The solution before the gap indicates the good solution. For example, we can see that there is a large jump between cases 7 and 8, corresponding to combining cluster 1(cases 2, 10) and cluster 2 (cases 4, 7, 3, 5, 6) and cluster 3(cases 1, 11, 12) and cluster 4(cases 8 and 9). Another possible cluster solution is a two- cluster solution because of the large jump between stages 10 and 11, forming cluster 1 (cases 2, 10, 1, 12, 11) and cluster 2 (cases 4, 7, 3, 5, 6, 8, 9). The agglomeration schedule shows that cases 2 and 10 are combined in a cluster first at stage 1(the cluster is labeled 2 since the cluster number is always the lowest of the case numbers in the cluster). At stage 2, cases 4 and 7 form to become cluster 2. At stage 3, cases 8 and 9 to form cluster 4. At stage 4, the case 3 is combined with case 4 which is already clustered with case 7 at stage 2. At stage 5, the case 1 is combined with case 12 to form cluster 3. At stage 6, case 5 which are combined with case 3 to form cluster. At stage 8, case 11 is combined with 1 to form 3rd cluster. At stage 9, cases 3 and 6 are combined to form a cluster. At stages 10 and 11, cases 3 and 8 are combined to form the final cluster. Fig. 1 shows the formed dendrogram.

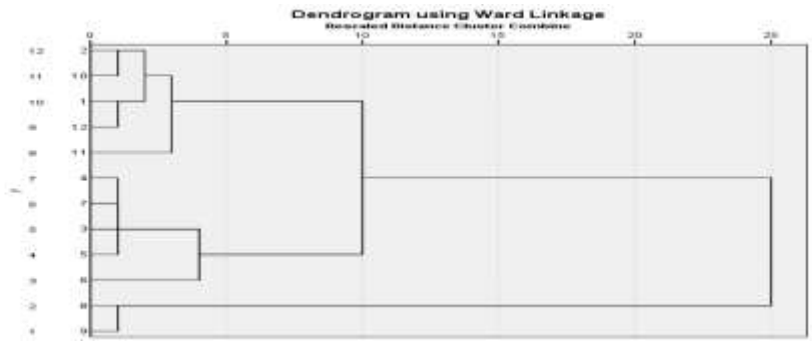

Fig.1 Dendrogram of physico-chemical parameters at Rajahmundry

From the CA of biological parameters showed at Rajahmundry vide table 11, there is a large jump between cases 1 and 2, corresponding to combining cluster 1 (cases 1 and 2) and cluster 2 (cases 1, 3). Agglomeration Schedule shows that cases 1 and 2 are combined in a cluster at first stage (the cluster is labeled 1 since the cluster number is always the lowest of the case numbers in the cluster). At stage 2, cases 1 and 3 form cluster 2(final cluster) as represented in the Dendrogram vide fig. 2.

Table 11 Agglomeration Schedule (Cluster Analysis of biological parameters at Rajahmundry)

Stage	Cluster Combined		Coefficient	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	1	2	156.00	0	0	2
2	1	3	61669.33	1	0	0

Fig. 2 Dendrogram (Cluster Analysis of biological parameters at Rajahmundry)

The agglomeration schedule tables vide table 12, shows the change in the distance measure is as additional cases are merged into clusters. The column labeled coefficients has the values of the distance statistic used to form the cluster; and a good cluster sees a sudden jump in the distance coefficient. The solution before the gap indicates the good solution. For example, we can see that there is a large jump between cases 6 and 7, corresponding to combining cluster 1(cases 1, 2, 4, 5, 3, 7) and cluster 2 (cases 10, 12, 8) and cluster 3(cases 9, 11) and cluster 4(cases 1 and 6). Another possible cluster solution is a two- cluster solution because of the large jump between stages 10 and 11, forming cluster 1 (cases 1, 2, 4, 5, 3, 7, and 6) and cluster 2(cases 9, 10, 11, 12, 8).

Table 12: Agglomeration Schedule (Cluster Analysis of physico-chemical parameters at Dhawaleswaram)

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	1	2	34.50	0	0	5
2	4	5	70.50	0	0	3
3	3	4	147.16	0	2	5
4	10	12	330.66	0	0	9
5	1	3	577.50	1	3	6
6	1	7	902.66	5	0	8
7	9	11	1373.66	0	0	10
8	1	6	2535.64	6	0	11
9	8	10	3763.47	0	4	10
10	8	9	6160.34	9	7	11
11	1	8	18703.91	8	10	0

The agglomeration table vide table 12 shows that cases 1 and 2 are combined in a cluster first at stage 1 (the cluster labeled 1 since the cluster number is always the lowest of the case numbers in the cluster). At stage 2 cases 4 and 5 to form a second cluster. At stage 3, cases 3 and 4 are combined. Cases 10 and 12 are combined at stage 4. At stage 5, case 3 and at stage 6, case 7 are combined with case 1 which is already combined with 2 at stage 1. At stage 7 the cases 9 and 11 forms a new cluster. Again at stage 8, the case 1 is combined with case 6 to form a cluster (1, 2, 4, 5, 3, 7, and 6). At stage 9, a case 8 is combined with 10 which are already combined with case 12 to form a cluster (cases 8, 10, 12). At stages 10, case 9 are combined with case 8 to form a final cluster (8, 9, 10, 11, and 12). At final stage the case 1 is combined with case 8 to form a final cluster represented in figure 3.

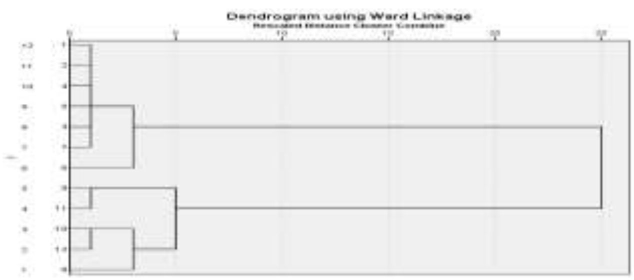

Fig.3 Dendrogram (Cluster Analysis of physico-chemical parameters at Dhawaleswaram)

Agglomeration Schedule vide table 13, shows that cases 1 and 2 are combined in a cluster at first stage (the cluster is labeled 1 since the cluster number is always the lowest of the case numbers in the cluster). At stage 2, cases 1 and 3 form cluster 2 (final cluster is represented in fig.4).

Table 13 Agglomeration Schedule (Cluster Analysis of biological parameters at Dhawaleswaram)

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	1	2	170.22	0	0	2
2	1	3	76038.45	1	0	0

Fig.4 Dendrogram (Cluster Analysis of biological parameters at Dhawaleswaram)

V. CONCLUSIONS

1. From the Average seasonal WQI analysis it was found that the water quality is found to be fair and medium at Rajahmundry and Dhawaleswaram respectively.
2. From the Factor Analysis (FA) conducted for the Physico-chemical at Rajahmundry and Dhawaleswaram the highest loadings obtained for the months of Monsoon and Post-monsoon period with variances of 99.437% and 99.422% and DO_Sat% is influencing the quality of the river at both Rajahmundry and Dhawaleswaram with the highest loadings with a variance of 63.017%.
3. The Cluster Analysis (CA) conducted at Rajahmundry and Dhawaleswaram shows that Monsoon season bears the higher pollution loadings and DO_Sat% affects the degree of pollution.
4. Principal Component Analysis (PCA) conducted for the Physico-chemical at Rajahmundry and Dhawaleswaram yielded the highest loadings for Pre and Post-Monsoon periods with variances of 99.444% and 99.429%. From the Principal Component Analysis (PCA) of the biological parameters, DO_Sat% is found to be influencing factor of the quality of the river at both Rajahmundry and Dhawaleswaram with the highest loadings and with a variance of 79.205% and 64.935%.

REFERENCES

[1]. Arzu Altin, Zeynip Filiz, Canzu Filik Iscen, Assessment of Seasonal Variations of Surface Water Quality Characteristics of Porsuk Stream, *Springer- Environ Monit Assess* (2009) 158:51–65.

[2]. Bharti N, Katyal.D, Water Quality Indices Used for Surface Water Vulnerability Assessment, *International Journal of Environmental Sciences,* Volume 2, No 1, 2011.

[3]. Hulya Boyaciogla, Hayal Bwater oyaciogla, Orhan Gundaz, Application of Factor Analysis in the Assessment of Surface Quality in Buyuk Menderes River Basin, *European water 9/10*: 43-49, 2005.

[4]. Indrani Gupta, Abahsinh Saalunkhe, Nanda Rohra, Rakesh Kumar, Chemometrics Data Analysis of Marine Water Quality in Maharastra, West Coast of India, *Indian journal of Geo Marine Sciences*, Vol 42(1), Feb 2013, 97-105.

[5]. Majid Ajorlo & Ramdzani B. Abdullah & Mohd Kamil Yusoff & Ridzwan Abd. Halim & Ahmad Husni Mohd. Hanif &Walter D. Willms & Mahboubeh Ebrahimia, Multivariate Statistical Techniques for the Assessment of Seasonal Variations in Surface Water Quality of Pasture Ecosystems, *Springer- Environment Monit Assess* (2013) 185: 8649-8658.

[6]. Prakash Raj Kannel , Sushil R. Kanel, Seockheon Lee, Y.S. Lee, Chemometrics in Assessment of Seasonal Variation of fresh Water Systems, *Springer - Environ Monit Assess* (2011) 174:529–545.

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

[8]. Soledad Oliva González, César A. Almeida, Sylvia Quintar, Miguel A. Mallea, Patricia S. González, Application of Multivariate Statistical Techniques to Evaluate Organic Pollution on a River in Argentina, *An Interdisciplinary Journal of Applied Science*: v. 6, n. 3, 2011

[9]. Tripathy. J.K., and Sahu K.C., Seasonal Hydrochemistry of the Groundwater in the Barrier Spit System of the Chilika Lagoon, India, *Journal of Environmental Hydrology*, Volume 12, 2015.

*T.P.Sreejani " Statistical Modelling of Water Quality of River Godavari at Rajahmundry and Dhawaleswaram, AP." American Journal of Engineering Research (AJER) 6.8 (2017): 112-120.