River water quality rapid judgment through water quality index and multivariate statistics

G. Sri Bala[1], G.V.R. Srinivasa Rao[2] and P.A.R.K. Raju[3]

[1]Research Scholar, Assistant Professor, Department of Civil Engineering , SRKR Engineering College, Bhimavaram, Andhra Pradesh, India.
[2]Professor& Head, Department of Civil Engineering, College of Engineering (A), Andhra University, Visakhapatnam, Andhra Pradesh, India.
[3]Professor, Department of Civil Engineering, SRKR Engineering college, Bhimavaram, Andhra Pradesh, India.

Abstract

In this study, multivariate statistical investigation and calculation of water quality index was made to analyze the water quality at seventeen sampling stations along the man made irrigation channel called Eluru canal originating at tail end of River Godavari. Eleven parameters were analyzed as per the standard procedures of Drinking water specification (Indian standards) during pre-monsoon, monsoon and post-monsoon seasons and were used to calculate Weighted Arithmetic Water Quality Index. The status of the index was found really poor in all the seasons. The data matrix (17x11) was very difficult to reveal the internal patterns. In the present study, Cluster Analysis was found to be useful for sampling strategy to locate spatial variations in water along the stream. It has formed two big clusters grouping seventeen stations. Similarly, Principal Component Analysis revealed that the underlying factors responsible for pollution were especially salts. Discriminant Function Analysis was very helpful in identification of the contribution of pH and Bio-chemical oxygen demand to index as a whole. Thus these techniques were found useful to reduce the voluminous data and identify the optimal rule for selection of sampling stations and variables. It helps for effective water quality supervision.

Keywords Cluster analysis; Discriminant function analysis; Principal component analysis; Weighted arithmetic mathematical index.

Introduction

Next to air, water is most central for the endurance of living beings on earth, available in two forms as surface and ground water. Surface water quality is very crucial as rivers are catch basins for waste and pollution carriers (Singh et al., 2005). And it is very much essential to understand the severity of pollution and the protection of surface waters from these pollutants.

Human proceedings and land use patterns are important factors causing pollution of surface waters (Zhang et al., 2007; Hussain et al., 2008). In order to predict pollution levels in water from time to time, it is necessary to go for frequent monitoring programs. Monitoring of water quality incorporates the study of various parameters and it becomes essential to understand the relationship between chemistry and biology involved among the parameters of water (Ouyang, Y., 2005, Kazi et al., 2009). In this regard, the mathematical equations called water quality indices were developed by Horton and Brown and also many more researchers used those indices to define water quality data.

Several indices were developed based on assortment of different parameters for diverse water bodies. The developed indices have various constraints. Weighted arithmetic water quality index (WQI_WA) is one of the specific consumption indices that is specifically adopted to represent the quality of water for drinking purpose. Many researchers have used this WQI to define water purity (Chauhan, 2010; Chowdhury et al., 2012; Rao et al., 2010; Balan et al., 2012).

Unless mentioned, the water data is termed as multivariate in nature. Multivariate statistics popularly known as pattern recognition techniques (environmetric methods) such as Principal Component Analysis (PCA), Cluster Analysis (CA) and Discriminant Function Analysis (DFA) are widely used to resolve such tedious problems involved in dealing multiple data. Multivariate methods grant speedy solution and steadfast management measures in water monitoring programs (Zhou et al., 2007).

PCA is a part of Factor Analysis (FA) which is a dimension reduction technique that is very useful to trim down the data (Panda et al., 2006). CA is used to detect the spatial similarity and dissimilarity among different sampling sites (Massart and Kaufmann, 1983). DFA was used by many researchers to detect the spatial and temporal changes in water quality (Shrestha and Kazama, 2007).

The basic objective of the present research was to reveal the internal classification of data by the following: (i) to analyse the data for the required parameters as per the standard procedures; (ii) to calculate water quality index that simplifies the data during three seasons; (iii) To analyse the data to the maximum possible extent statistically by CA, PCA and DFA.

In this study, the water samples were collected from seventeen sampling stations of Eluru stream, Godavari River during three seasons in the year 2015 to examine the structure of water body. The findings were helpful to investiage the suitability of water for drinking purpose and estimate the sampling and experimental strategies in assessment of big water bodies spatially and temporally in an optimistic manner.

Materials and methods

Study sites

River Godavari is the holy perennial river in the district of West Godavari, Andhra Pradesh, India. Agriculture in West Godavari is carried on through six streams viz., Eluru canal, Narsapur canal, Venkayya Vayyeru canal, Gostani canal and Kakaraparru canal and Attili canal originating from river Godavari.

The study was carried out at seventeen sampling stations along Eluru stream of River Godavari. The sampling stations were SS1-Nidadavole, SS2 – Nandamuru, SS3 – Arrula, SS4 – Navabpalem, SS5 – Krishnayapalem, SS6 –Pathipadu, SS7 – Tadepalligudem, SS8 - Pentapadu, SS9 – Unguturu, SS10 – Narayanapuram, SS11 – Chebrolu, SS12 – Kaikaram, SS13 – Pulla, SS14 – Bhimadolu, SS15 – Gundugolanu, SS16 - Pothunuru, SS17 – Kovali.

Water sampling and analysis

Water samples were collected along the seventeen sampling stations in three seasons i.e., pre - monsoon, monsoon and post monsoon and were analysed for various parameters viz., pH, Electrical conductivity (EC), Total dissolved solids (TDS), Alkalinity, Hardness, Calcium (Ca^{2+}), Magnesium (Mg^{2+}), Dissolved oxygen(DO), Bio-chemical oxygen demand(BOD), Nitrates (NO^{3-}) and Chlorides (CL^{-}). All the parameters were analyzed in the laboratory as per the standard APHA procedures.

Data treatment

WQI_WA development

WQI_WA is calculated taking into consideration ICMR, BIS and WHO standards using the following expression

$WQI\_WA = \sum QR.Wi/\sum Wi$

The quality rating QR for each attribute is calculated as given below:

$$QR = 100 * (Vni - Vio)/(Sni - Vio)$$

Where

Vni = the analyzed value of attribute found in laboratory

Vio = the ideal value as per BIS and ICMR for each parameter. The value is zero for all the parameters except for pH and DO. The value for pH is 7.0 and for DO is 14.6 respectively.

Sni = the standard value for each attribute

Wi = recommended unit weight for each attribute by standard agencies considered from the following Table 1.

The overall water quality index was computed by aggregating the quality rating with the unit weight linearly (Tripathy and Sahu, 2005).

Table 1. Criterions for water (All attributes are expressed in mg/L except pH and EC)

SNO	Parameters	Standards	Recommended Agency	Unit Weight
1	pH	6.5 - 8.5	ICMR/BIS	0.219
2	Electrical Conductivity	300	ICMR	0.371
3	Total Dissolved Solids	500	ICMR/BIS	0.0037
4	Total Alkalinity	120	ICMR	0.0155
5	Total Hardness	300	ICMR/BIS	0.0062
6	Calcium	75	ICMR/BIS	0.025
7	Magnesium	30	ICMR/BIS	0.061
8	Chlorides	250	ICMR	0.0074
9	Nitrate	45	ICMR/BIS	0.0412
10	Dissolved Oxygen	5	ICMR/BIS	0.3723
11	Biological Oxygen Demand	5	ICMR	0.3723

The calculated WQI values are interpreted using the following classification given in Table 2.

Table 2. Status and index level (Source: Chaterjee and Raziuddin, 2002)

WQI_WA Level	Status
0-25	Excellent
26-50	Good
51-75	Poor
76-100	Very Poor
>100	Unsuitable

Multivariate Statistical Analysis

Different types of analysis viz., Pearson's correlation analysis, principle component analysis, cluster analysis and discriminant function analysis were made using Excel 2007 and SPSS 22.

Karl Pearson's correlation analysis

Correlation analysis was performed using Excel 2007. This analysis evaluates the correlation coefficients between variables and will help to identify the significant relationships among the variables. This is a basic step for any statistical method to process with the multiple data.

Principle component analysis (PCA)

PCA is designed to renovate the original variables into new uncorrelated variables called principle components (Helena et al., 2000). Multiple parameters were transformed into few components. The first principle component explains the maximum amount of variance and the second one explains the maximum residual variance (Vikas, 2012). The axis is rotated to the maximum extent to reduce the effect of less significant variables in the data. This method of reduction provides a few number of factors that usually consider the importance of original variables as a whole (Vega et al., 1998). PCA was applied using varimax rotation with Kaiser normalization.

Cluster analysis (CA)

It is a method that organizes variables or cases into clusters. The data is not classified on statistical strategy but is based on similarities and dissimilarities (Shrestha and Kazama, 2007). In this procedure, the similar attributes fall into identical cases. Hierarchical clustering analysis is most widely used to group the clusters. The similarity and dissimilarity fashion decides the pattern to form an icicle plot called dendogram. Agglomerative hierarchical cluster analysis using Ward's method and squared Euclidean distance was used to grade variables into clusters (Andrade et al., 2000).

Discriminant function analysis (DFA)

It is a technique used to differentiate the attributes between two or more groups. It is performed on the data matrix in various methods (Varol, 2009). This method is similar to logistic regression. The analysis forms various groups called discriminant functions. The discriminant function group is of the form:

$$D = C + D1X1 + D2X2 + \cdots \dots \dots \dots \dots . DnXn$$

Where D = discriminant score (Z score)

C = Y-intercept

D1, D2,....Dn = discriminant function coefficients

X = discriminant raw attribute score

n = number of discriminant variables.

Results and discussion

The physico-chemical attributes and WQI_WA values for three seasons were presented in Tables 3, 4 and 5.

Table 3. WQI_WA and characteristics of water samples (Pre Monsoon)
(All values are in mg/L, except pH and EC, no units for WQI_WA)

SNO	pH	EC	TDS	Alkalinity	Hardness	Ca^{2+}	Mg^{2+}	CL^{-}	NO^{3-}	DO	BOD	WQI_WA	Status
S-1	9.36	200.00	140.00	93.00	95.00	22.00	9.72	21.27	4.15	4.00	8.50	**112.57**	unsuitable
S-2	8.50	190.00	130.00	100.00	65.00	16.00	6.08	17.72	3.49	4.00	8.50	102.68	unsuitable
S-3	8.30	230.00	150.00	110.00	95.00	18.00	12.15	17.72	2.33	4.00	8.50	104.97	unsuitable
S-4	9.17	220.00	140.00	105.00	70.00	16.00	7.29	21.27	2.38	4.40	6.00	98.37	very poor
S-5	9.33	190.00	130.00	45.00	125.00	22.00	17.01	17.72	0.00	4.00	8.00	109.31	unsuitable
S-6	9.28	220.00	150.00	95.00	125.00	26.00	14.58	28.36	2.50	4.00	8.20	112.67	unsuitable
S-7	8.48	200.00	140.00	105.00	85.00	16.00	10.93	53.17	4.81	4.40	8.00	100.67	unsuitable
S-8	6.74	430.00	280.00	110.00	165.00	60.00	3.68	70.90	6.86	4.00	8.00	104.12	unsuitable
S-9	8.09	370.00	240.00	105.00	100.00	20.00	12.15	21.27	13.30	5.60	6.00	98.63	very poor
S-10	8.49	250.00	170.00	135.00	105.00	42.00	0.03	21.27	2.50	4.20	6.80	98.63	very poor
S-11	7.62	380.00	260.00	165.00	120.00	22.00	15.80	53.17	2.33	4.00	7.20	105.48	Unsuitable
S-12	7.97	310.00	220.00	150.00	150.00	30.00	18.23	28.36	1.72	4.80	6.80	99.35	very poor
S-13	9.95	250.00	170.00	125.00	90.00	16.00	12.15	28.36	3.80	4.80	6.80	112.40	Unsuitable
S-14	8.29	400.00	270.00	180.00	140.00	22.00	20.65	46.08	5.37	3.60	4.80	103.76	Unsuitable
S-15	8.15	430.00	280.00	125.00	85.00	16.00	10.93	35.45	3.80	4.80	6.50	108.11	Unsuitable
S-16	8.50	375.00	240.00	110.00	165.00	60.00	3.68	53.17	0.00	4.00	8.00	116.28	Unsuitable
S-17	8.25	315.00	220.00	135.00	70.00	16.00	7.29	28.36	3.60	4.80	6.80	100.58	Unsuitable

Table 4. WQI_WA and characteristics of water samples (Monsoon)
(All values are in mg/L, except pH and EC, no units for WQI_WA)

SNO	pH	EC	TDS	Alkalinity	Hardness	Ca^{2+}	Mg^{2+}	CL^{-}	NO^{3-}	DO	BOD	WQI_WA	Status
S-1	8.38	180.00	130.00	80.00	110.00	24.00	12.15	35.45	15.80	5.20	2.40	68.86	Poor
S-2	8.34	180.00	130.00	90.00	90.00	28.00	4.87	28.36	15.20	5.60	2.40	66.53	Poor
S-3	8.14	210.00	150.00	70.00	100.00	12.00	17.00	35.45	16.00	6.00	2.40	67.23	Poor
S-4	8.30	200.00	140.00	90.00	100.00	24.00	9.73	28.36	14.26	4.40	4.80	83.40	very poor
S-5	7.86	200.00	140.00	80.00	90.00	20.00	9.72	35.45	14.16	4.30	2.40	67.22	Poor
S-6	8.05	210.00	150.00	80.00	100.00	36.00	2.45	35.45	12.90	6.40	2.30	63.27	Poor
S-7	8.48	220.00	160.00	110.00	90.00	28.00	4.87	28.36	15.10	4.80	4.80	85.43	very poor
S-8	8.23	220.00	160.00	90.00	100.00	32.00	4.88	14.18	11.68	5.60	1.60	64.66	Poor
S-9	8.23	220.00	160.00	80.00	100.00	12.00	17.00	35.45	11.00	6.20	1.60	64.22	Poor
S-10	8.08	250.00	170.00	90.00	110.00	32.00	7.30	35.45	11.53	5.20	7.00	93.99	very poor
S-11	8.15	260.00	170.00	100.00	130.00	28.00	14.59	42.54	12.30	4.40	4.80	87.70	very poor
S-12	7.93	280.00	190.00	110.00	100.00	28.00	7.30	42.54	9.05	4.40	4.80	86.07	very poor
S-13	7.98	270.00	190.00	90.00	150.00	52.00	4.89	49.63	10.49	5.20	5.00	84.86	very poor
S-14	7.90	330.00	220.00	90.00	120.00	32.00	9.73	42.54	10.57	4.00	4.80	91.34	very poor
S-15	7.95	320.00	220.00	90.00	130.00	32.00	12.16	49.63	11.30	5.20	4.80	88.29	very poor
S-16	8.00	330.00	230.00	120.00	130.00	40.00	7.31	49.63	8.47	4.40	2.40	79.33	very poor
S-17	7.80	350.00	240.00	120.00	130.00	52.00	0.04	42.54	6.37	4.50	3.00	80.90	very poor

Table 5. WQI_WA and characteristics of water samples (Post Monsoon)
(All values are in mg/L, except pH and EC, no units for WQI_WA)

SNO	pH	EC	TDS	Alkalinity	Hardness	Ca^{2+}	Mg^{2+}	CL^-	NO^{3-}	DO	BOD	WQI_WA	Status
S-1	8.96	200.00	140.00	100.00	90.00	20.00	9.70	44.50	5.00	7.40	3.60	**75.53**	very poor
S-2	8.96	200.00	140.00	100.00	90.00	14.00	13.36	42.50	5.00	6.40	3.60	78.48	very poor
S-3	8.98	200.00	140.00	95.00	75.00	18.00	7.29	21.27	8.20	6.60	3.40	76.52	very poor
S-4	8.94	200.00	140.00	100.00	65.00	20.00	3.65	28.36	3.00	6.50	3.50	76.16	very poor
S-5	8.83	190.00	130.00	95.00	80.00	18.00	8.51	42.54	8.30	6.50	3.40	74.70	Poor
S-6	8.76	190.00	130.00	85.00	75.00	14.00	9.72	28.36	6.80	6.20	2.80	71.67	Poor
S-7	8.60	210.00	140.00	90.00	70.00	18.00	6.08	28.36	9.80	5.00	5.50	88.14	very poor
S-8	8.34	200.00	140.00	110.00	80.00	14.00	10.93	35.45	0.02	5.60	5.20	81.90	very poor
S-9	8.54	210.00	140.00	95.00	75.00	20.00	6.08	28.36	2.80	6.80	3.60	73.09	Poor
S-10	8.64	210.00	140.00	100.00	65.00	18.00	4.86	35.45	9.00	7.60	1.40	61.24	Poor
S-11	8.96	210.00	140.00	85.00	75.00	16.00	8.50	21.07	12.80	7.30	2.40	70.67	Poor
S-12	8.55	210.00	140.00	90.00	90.00	20.00	9.70	21.27	10.20	5.60	4.20	80.20	very poor
S-13	8.24	230.00	150.00	100.00	80.00	18.00	8.51	28.36	11.50	4.40	3.60	78.91	very poor
S-14	7.96	210.00	140.00	100.00	75.00	18.00	7.29	28.36	13.70	5.20	4.60	77.39	very poor
S-15	8.41	250.00	170.00	115.00	90.00	22.00	8.51	49.63	13.80	5.00	2.40	75.12	very poor
S-16	8.05	260.00	170.00	105.00	100.00	22.00	10.94	35.45	12.30	4.60	3.50	79.09	very poor
S-17	8.05	230.00	160.00	95.00	90.00	20.00	9.72	28.36	10.80	4.50	4.00	78.93	very poor

In the monsoon and post monsoon seasons, the WQI_WA value was found to vary from 60 to 100. The water was relatively poor and very poor in all the sampling stations. There was no large variation in terms of quality from upstream to mid and downstream. In pre- monsoon, the values of WQI_WA were found to be higher crossing 100 at most of the sampling stations and the water was found to be unsuitable and very poor for drinking purpose. The seasonal variations were significant as some parameters were found to be high in summer such as pH and BOD compared to other two seasons. DO was minimum in summer season.
WQI_WA status was readily understood whereas the data matrix (17x11) was difficult to read and interpret the relations in three seasons.

Pearson correlation coefficient analysis
Pearson correlation coefficient analysis is made on the crowd of variables to spot the literal rapport between attributes. The association might be strongly positive or negative between attributes and this is a deep-seated measure required for any statistical course of action. The correlation matrix of eleven parameters was presented in Tables 6. for the three seasons respectively.

Table 6. Correlation matrix of the water quality parameters considered (three seasons)

Pre Monsoon											
	pH	**EC**	**TDS**	**Alkalinity**	**Hardness**	**Ca^{2+}**	**Mg^{2+}**	**CL^-**	**NO^{3-}**	**DO**	**BOD**
pH	1										
EC	-0.698	1									
TDS	-0.705	**0.994***	1								
Alkalinity	-0.416	0.565	0.616	1							
Hardness	-0.388	0.488	0.492	0.156	1						
Ca^{2+}	-0.43	0.408	0.38	0.012	0.778	1					
Mg^{2+}	0.133	0.042	0.09	0.199	0.175	-0.482	1				
CL^-	-0.602	0.63	0.632	0.352	0.573	0.555	-0.07	1			
NO^{3-}	-0.319	0.344	0.329	0.093	-0.121	-0.11	0.004	0.088	1		
DO	-0.008	0.168	0.168	0.008	-0.332	-0.288	-0.013	-0.258	0.537	1	
BOD	0.081	-0.478	-0.489	-0.609	0.026	0.204	-0.284	-0.015	-0.346	-0.358	1
Monsoon											
	pH	**EC**	**TDS**	**Alkalinity**	**Hardness**	**Ca^{2+}**	**Mg^{2+}**	**CL^-**	**NO^{3-}**	**DO**	**BOD**
pH	1										
EC	-0.707	1									
TDS	-0.681	**0.993***	1								
Alkalinity	-0.21	0.639	0.647	1							
Hardness	-0.483	0.717	0.707	0.332	1						
Ca^{2+}	-0.446	0.612	0.629	0.583	0.69	1					
Mg^{2+}	0.184	-0.206	-0.238	-0.509	-0.06	-0.764	1				
CL^-	-0.631	0.693	0.671	0.293	0.748	0.398	0.118	1			
NO^{3-}	0.696	-0.853	-0.861	-0.653	-0.596	-0.646	0.359	-0.515	1		
DO	0.346	-0.49	-0.433	-0.584	-0.272	-0.256	0.11	-0.323	0.326	1	
BOD	-0.13	0.323	0.26	0.239	0.323	0.253	-0.06	0.324	-0.142	-0.445	1
Post Monsoon											
	pH	**EC**	**TDS**	**Alkalinity**	**Hardness**	**Ca^{2+}**	**Mg^{2+}**	**CL^-**	**NO^{3-}**	**DO**	**BOD**
pH	1										
EC	-0.655	1									
TDS	-0.589	**0.962***	1								
Alkalinity	-0.371	0.482	0.58	1							
Hardness	-0.328	0.545	0.605	0.327	1						
Ca^{2+}	-0.318	0.644	0.636	0.309	0.32	1					
Mg^{2+}	-0.125	0.134	0.2	0.131	0.795	-0.32	1				
CL^-	0.028	0.174	0.273	0.669	0.393	0.132	0.309	1			
NO^{3-}	-0.417	0.595	0.485	-0.075	0.234	0.369	-0.001	-0.109	1		
DO	0.79	-0.638	-0.617	-0.288	-0.425	-0.263	-0.258	0.034	-0.429	1	
BOD	-0.326	-0.117	-0.1	0.01	0.111	-0.1	0.175	-0.186	-0.227	-0.497	1

***Correlation is significant at the 0.01 level**

It was found from Table 6 that the value of correlation coefficient for the attributes TDS and EC was very high. The value was greater than 0.9 nearing 1.0. Next to EC and TDS, hardness and calcium were moderately correlated. These relation coefficients strongly tell the liaison between parameters that will help in further statistical modelling. Further, the relation

between hardness and Ca^{2+}, Mg^{2+} was clear that there is a clear impact of these divalent cations on hardness property as a whole in three seasons.
Principal Component Analysis (PCA)

PCA was executed on eleven parameters on the normalized data matrix for three seasons separately. The factors with high Eigen value were considered highly significant. The principal components were extracted considering the value of eigen value greater than 1. From the results of PCA as shown in Table 7, the data generated three components showing a cumulative variance of 79.39% and 75.92% in monsoon and pre monsoon seasons.
Similarly, the first four components were derived out of eleven parameters showing maximum variance of 86.35% in post monsoon. The four components derived have a subsequent eigen value greater than 1 satisfying Kaiser criteria.

Table 7. Extracted values (three seasons)

Pre Monsoon						
Component	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**
1	4.369	39.719	39.719	4.199	38.177	38.177
2	2.528	22.986	62.705	2.121	19.285	57.461
3	1.453	13.212	75.917	2.03	18.455	75.917
Monsoon						
Component	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**
1	5.905	53.679	53.679	4.468	40.62	40.62
2	1.615	14.677	68.357	2.452	22.289	62.909
3	1.213	11.03	79.386	1.813	16.478	79.386
Post Monsoon						
Component	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**
1	4.595	41.776	41.776	3.719	33.813	33.813
2	1.928	17.528	59.304	1.941	17.642	51.455
3	1.779	16.172	75.475	1.939	17.628	69.083
4	1.196	10.874	86.349	1.899	17.267	86.349

Scree plot is one more criteria to decide the number of components from the data. The scree plots for the data matrix in three seasons obtained were shown in Fig 1, Fig 2 and Fig 3 respectively.

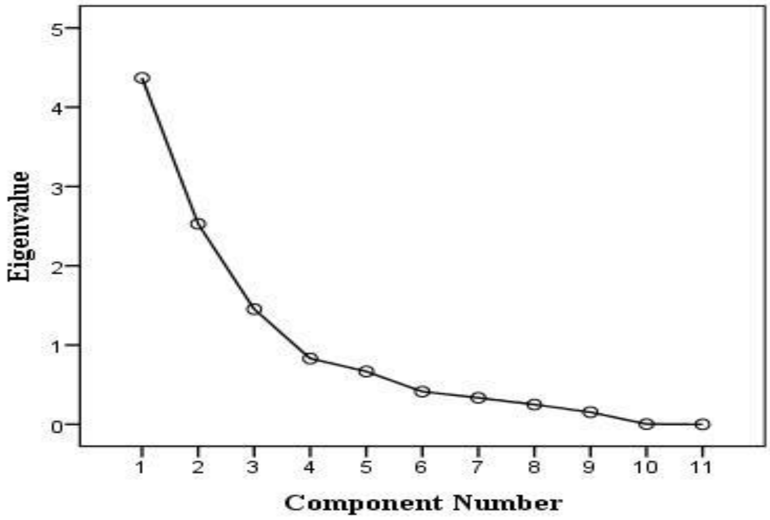

Fig 1. Scree plot of the Eigen values (Pre Monsoon)

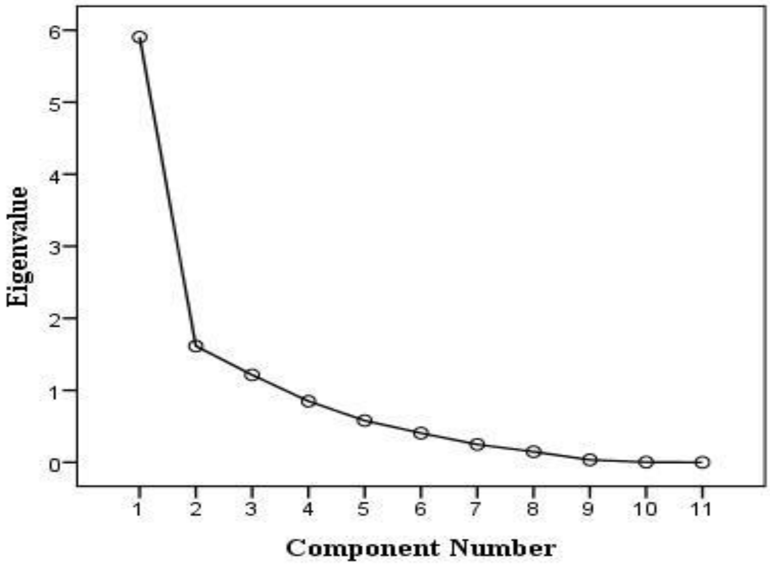

Fig 2. Scree plot of the Eigen values (Monsoon)

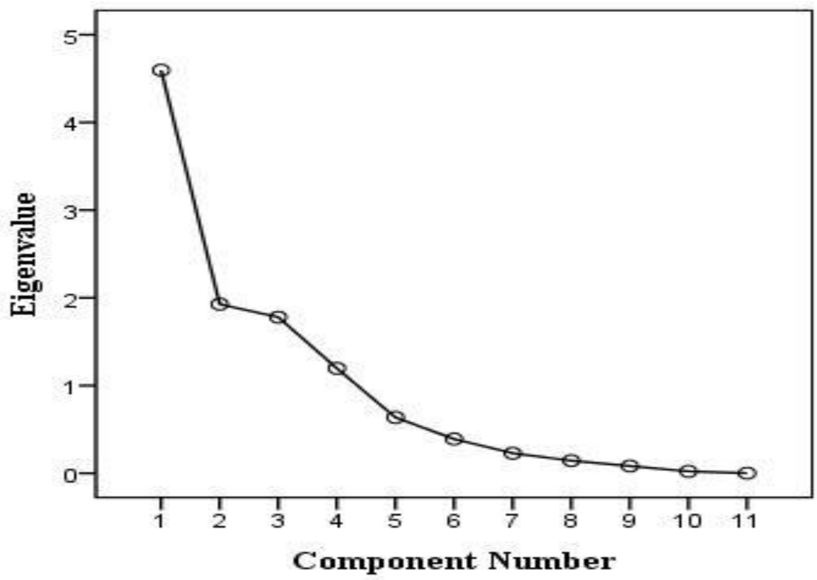

Fig 3. Scree plot of the Eigen values (Post Monsoon)

The scree plots were clear that three components were descending with more variance compared to other eight parameters in both pre monsoon and monsoon periods. It was also evident that the four factors gradually vary among eleven parameters and form major components for extraction in post monsoon period. Factor loading matrix reveals the level of pollution with the important constituents affecting it.
Liu et al (2003) has classified the factor loadings such as strong, moderate and weak classes. If the range of loading is from 0.3 to 0.5, then it is termed as weak. If the loading range varies from 0.5 to 0.75, it is termed as moderate and if the loading is greater than 0.75, it is termed as strong.

Table 8. Rotated Component Matrix (three seasons)

Parameter	**Pre Monsoon**			**Monsoon**			**Post Monsoon**			
	1	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**4**
EC	-0.798	-0.223	0.023	-0.805	-0.129	-0.016	**0.918**	0.222	0.17	-0.13
TDS	**0.856**	0.298	0.309	**0.855**	0.3	0.314	**0.855**	0.352	0.221	-0.125
NO3-	**0.851**	0.283	0.361	**0.85**	0.34	0.247	**0.789**	-0.378	0.132	0.104
Ca2+	0.46	0.115	0.657	0.288	0.678	0.477	0.742	0.269	-0.268	
Alkalinity	0.732	-0.453	0.044	**0.811**	0.104	0.188	0.275	**0.886**		-0.174
CL-	**0.754**	-0.27	-0.471	0.498	**0.756**	0.101		**0.829**	0.302	0.22
Mg2+	-0.162	-0.21	**0.8**	0.036	-0.951	0.029			**0.981**	-0.127
Hardness	**0.828**	-0.149	0.004	**0.851**	-0.148	0.252	0.39	0.259	**0.809**	-0.103
BOD	0.18	0.799	0.036	-0.767	-0.482	-0.086	-0.239			-0.898
DO	-0.133	**0.844**	0.063	-0.219	-0.173	-0.811	-0.559		-0.23	**0.728**
pH	-0.163	-0.442	-0.709	0.126	-0.014	**0.806**	-0.608			**0.643**

In post monsoon period, from Table 8, the first factor derived showed maximum loading in EC, TDS and nitrates with loading greater than 0.75. It was also evident from the table that the pollution was high with salts that clearly depict the intrusion of sea water into the stream in respective sampling stations under the concerned study. The factor loadings were high for EC and TDS in monsoon and pre monsoon periods. Even the findings of correlation analysis confirmed the factor loadings of PCA in three seasons showing high relations between EC and TDS.
Cluster Analysis (CA)
The seventeen sampling stations with eleven parameters each were assembled into two big clusters in all the three seasons respectively. Dendograms of the pattern were presented in figures 4, 5 and 6 for three seasons.

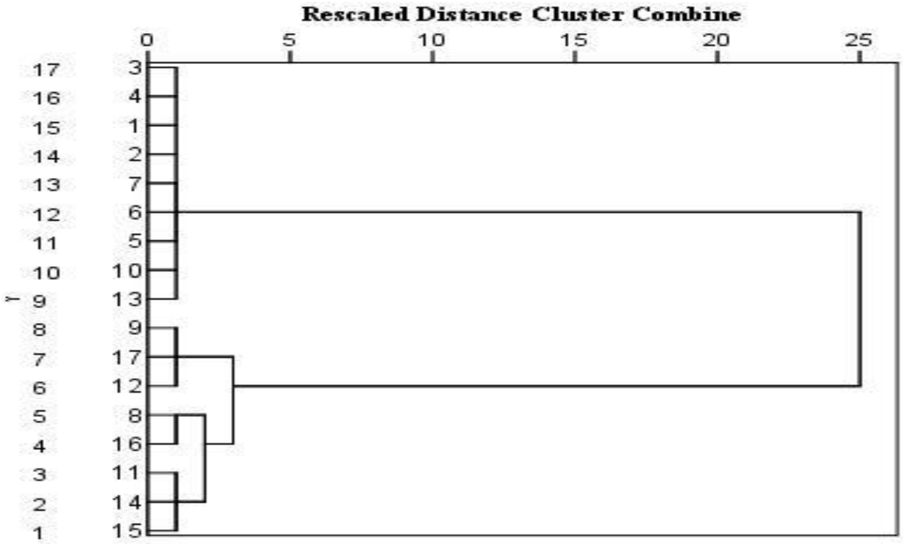

Fig 4. Dendogram of CA for sampling stations (Pre Monsoon)

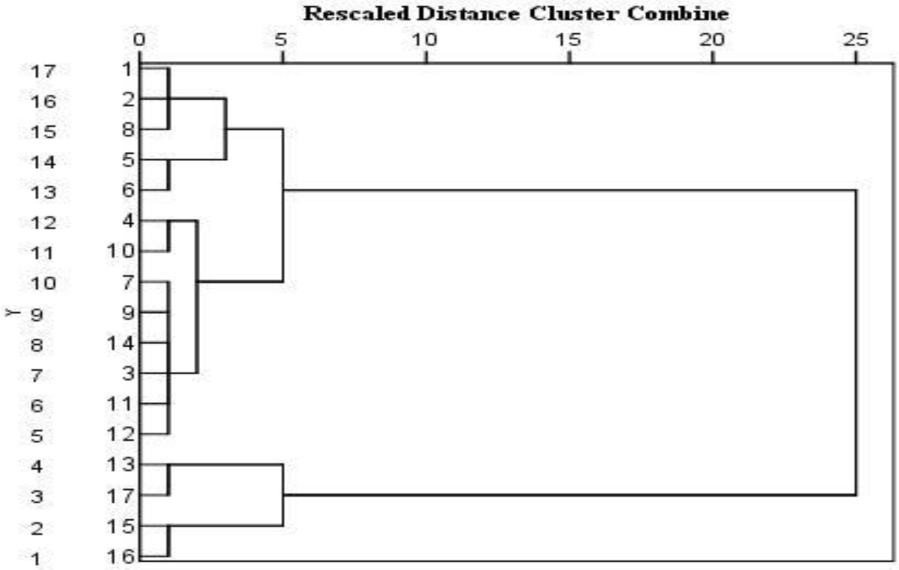

Fig 5. Dendogram of CA for sampling stations (Post Monsoon)

From Fig 4, it was clear that two big clusters were generated using agglomerative cluster analysis. One cluster i.e., cluster A grouping sampling stations 3, 4, 1, 2, 7, 6, 5, 10, 13.The other cluster i.e., cluster B covers two sub clusters. First sub cluster covers sampling stations 9, 17 and 12. The other sub cluster groups the sampling stations 11, 14 and 15.

Similarly, from Fig 5 & Fig 6, it was clear that two big clusters were formed and these clusters group most of the stations based on similarities and dissimilarities among the data. Based on the clusters formed, in future the number of sampling points can be minimised in water quality monitoring and management for a particular stream. Similarly Simeonov et al. 2003 had classified the surface water course using hierarchical cluster analysis to optimise the cost of monitoring network.

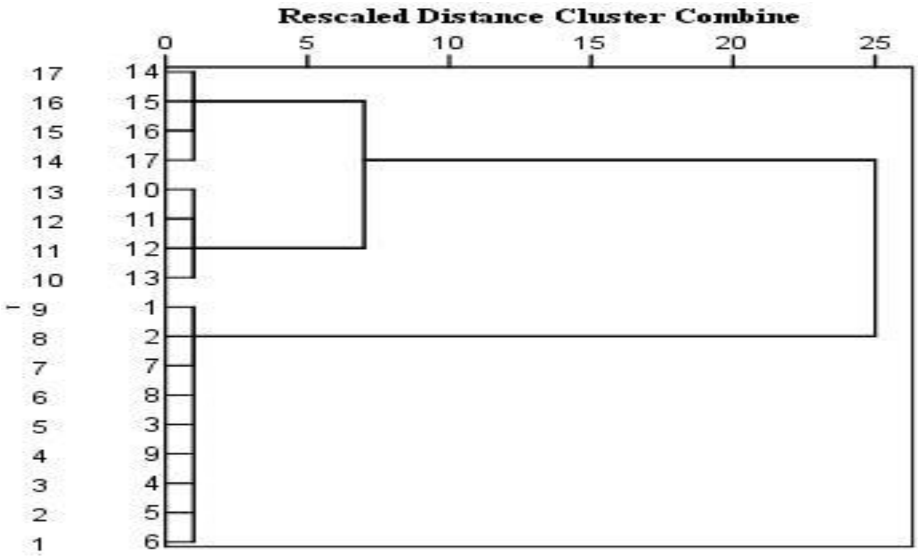

Fig 6. Dendogram of CA for sampling stations (Monsoon)

Discriminant Function Analysis (DFA)

The data matrix with eleven attributes was implemented with DFA in SPSS 22 version and the results were found to discriminate between two groups such as Group A (WQI_WA = 50) and Group B (WQI_WA = 100). Two groups were selected in such a way that if the value of the index is in between 50 to 75, the parameter discriminates itself into Group A and If the index is greater than 75, the parameter distinguishes into Group B. Wilk's Lambda test was performed and the test confirmed that the discriminant function was statistically significant as seen in Table 9. Further it was evident that 100% variance was there among the data in matrix. The value of canonical correlation coefficient was very high nearing 1.0. Hence, the data has significant correlations in three seasons and can be well classified into distinct groups.

Table 9. Wilk's Lambda and Eigen values (All seasons)

Pre Monsoon				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	7.752	100.0	100.0	.941
Test of Function	Wilks' Lambda	Chi-square	df	Sig.
1	.114	21.693	10	.017
Monsoon				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	30.357	100.0	100.0	.984
Test of Function	Wilks' Lambda	Chi-square	df	Sig.
1	.032	34.454	10	.000
Post Monsoon				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	5.22	100.00	100.00	0.92
Test of Function	Wilks' Lambda	Chi-square	df	Sig.
1	.161	18.278	10.000	.050

Further from Table 10, it was confirmed that in the data, pH and BOD were more significant with the higher coefficients compared to the others. All the data is 100% originally classified as per statistical inference.

Table 10. Classification function coefficients (Three seasons)

Parameter	Pre Monsoon		Monsoon		Post Monsoon	
	WQI_WA		WQI_WA		WQI_WA	
	50.00	100.00	50.00	100.00	50.00	100.00
pH	117.20	133.61	1516.05	1543.02	458.36	477.10
EC	0.04	-0.23	9.18	8.40	-0.55	-1.13
TDS	1.72	2.52	-8.50	-6.94	5.62	6.48
Alkalinity	0.66	0.50	-9.37	-8.19	18.81	20.03
Hardness	-0.27	-0.46	-8.25	-7.65	3.23	3.51
Ca2+	-2.26	-2.85	11.58	11.05	-1.48	-1.01
CL-	1.53	1.97	14.80	14.93	-10.24	-10.92
NO3-	5.32	5.87	-44.22	-44.63	16.90	17.94
DO	-41.65	-64.65	-121.23	-135.32	-16.55	-20.18
BOD	90.06	103.50	39.58	62.03	120.97	127.50
(Constant)	-882.63	-1051.65	-5486.79	-5942.33	-3319.28	-3615.38

Conclusions

1. The Weighted arithmetic water quality index (WQI_WA) used in this study has shown that the water quality is very poor in the study area and the status was easy to understand.
2. Results of correlation analysis has shown that EC, TDS were strongly related to each other with the coefficient crossing 0.9.
3. The data for three seasons was subjected to different multivariate statistics such as CA, PCA and DFA. CA divided seventeen stations spatially into two big clusters in all three seasons under study. PCA generated three components in monsoon and pre monsoon periods. Similarly, PCA extracted four principal components in post monsoon period out of eleven attributes experimented in the laboratory showing the major pollutant causes were salt intrusions in the region.
4. Discriminant Function Analysis was exclusively used to find the participation of each attribute in complete water purity as a whole. In all the seasons, it was clear that the data was correctly classified and the variables pH and BOD were influencing the water purity to the maximum extent for the complete water body Eluru stream with its seventeen sampling stations.

5. The results also tell that the degree of pollution was high in the study area and the water needs treatment before it is used for drinking purpose.

Thus, this study illustrates the application of multivariate statistics for the researchers to (i) simplify large data set; (ii) identify pollution sources (iii) understand spatial and temporal changes in water (iv) identify major responsible parameters affecting quality as a whole.

Acknowledgement

The authors appreciate the financial support provided by Technology Mission Cell: Water and Solar Energy, Department of Science and Technology, New Delhi. through Grant no. DST/TM/WTI/2K14/183 (G).

References

Andrade, E.M. (2000). Regionalization of small watersheds in Arid and Semi-arid regions: Cluster and Andrews Curve approaches. *Journal of Brazillian Association of Agricultural Engineering*, 18(4), 39-52.

APHA (2005). *Standard Methods for the examination of Water and Wastewater*, 20th edition, Washington D.C: American Public Health Association.

Balan, I.N., Shivakumar, M. and Kumar, P.D.M. (2012). An assessment of ground water quality index in Chennai, Tamil Nadu, India. *Chronicles Young Scient*., 3(2). 146 – 150.

BIS 2012. Bureau of Indian Standards, IS: 10500, Manak Bhavan, New Delhi, India.

Boyacioglu, H., & Boyacioglu, H. (2007). Water pollution sources assessment by multivariate statistical methods in the Tahtali Basin, Turkey. *Environmental Geology*, 54, 275 – 282.

Brown R.M., Mclelland N.I., Deininger R.A., & Tozer R.G. (1970). A water quality index – DO we dare? *Water Sew Works,* 117 (10): 339 -343.

Chaterjee, C., & Raziuddin, M. (2002). Determination of water quality index (WQI) of a degraded river in Asanol Industrial area, Raniganj, Burdwan, West Bengal, Nature. *Environment and pollution Technology*, 1(2), 181 – 189.

Chauhan, A., & Singh, S. (2010). Evaluation of Ganga water for drinking purpose by water quality index at Rishikesh, Uttarakhand, India, Raport opinion, 2(9). 53 – 61.

Chowdhury, R.M., Muntasir, S.Y., & Hossain, M.M. (2012). Water quality index of water bodies along Faridpur – Barisal road in Bangladesh", *Glob. Eng. Tech. Rev.*, 2(3). 1- 8.

Helena, B., Pardo,R., Vega, M., Barrado, E., Fernandez, J.M., & Fernandez, L. (2000). Temporal evolution of groundwater composition in an alluvial (Pisuerga river, Spain) by principal component analysis. *Water Research*, 34, 807-816.

Horton RK (1965). An index number system for rating water quality. *J Water Pollution Control Fed* 37(3): 300 – 306.

Hussain, M., Ahmed, S.M., Abderrahman, W. (2008). Cluster analysis and quality assessment of logged water at an irrigation project, eastern Saudi Arabia. *Journal of Environmental Management*, 86:297 – 307.

Kazi T.G., Arain M.B., Jamali M.K., Jalbani, N., Afridi, H.I., Sarfraz, R.A., Baig, J.A., & Shah, A.Q. (2009). Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. *Ecotox.Environ.Safe*, 72(20):301-309.

Liu, C.W., Lin, K.H., & Kuo, Y. M. (2003). Application of factor analysis in the assessment of ground water quality in a balckfoot disease area in Taiwan. *The Science of the Total Environment*, 313, 77 – 89. Doi:10.1016/S0048-9697(02)00683 - 6.

Massart, D.L., & Kaufman, L.(1983), the interpretation of chemical data by the use of cluster analysis. New York: Wiley.

Ouyang, Y. (2005). Evaluation of river water quality monitoring stations by principal component analysis. *Water Res*, 39, 2621 -2635 [Cross Ref] [Pub Med].

Panda, U.C., Sundaray, S.K., Rath, P., Nayak, B.B., & Bhatta, D. (2006). Application of factor and cluster analysis for characterization of river and estuarine water systems – a case study: Mahanadi River (India). *J. Hydrol* 331: 434 – 445.

Rao, C.S., Rao, B.S., Hariharan, A.V.L.N.S.H. and Bharathi, N.M. (2010). Determination of water quality index of some areas in Guntur district Andhra Pradesh. *Int. J., Appl. Bio. Pharm. Tech*, I (1). 79 – 86.

Simeonov, V., Stratis, J.A., Samara, C., Zachariadis, G., Voutsa, D., & Anthemidis, A., et al. (2003). Assessment of the surface water quality in Northern Greece. *WaterResearch,37*, 4119 – 4124. doi: 10.1016/S0043-1354(03)00398 – 1.

Singh, K.P., Malik, A., & Sinha, S. (2005). Water quality assessment and apportionment of pollution sources of Gomti River (India) using multivariate statistical techniques – A case study. *AnalticaChimicaActa*, 538, 355 – 374, doi:

10.1016/j.aca.2005.02.006.

Shrestha, S., & 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.

Tripathy, J.K. & Sahu K. C. (2005). Seasonal hydrochemistry of groundwater in the barrier- spit system of Chilika lagoon. *Journal of Environmental Hydrology*, *12* (7), 1-9.

Varol, M., & Sen, B. (2009). Assessment of surface water quality using multivariate statistical techniques: a case study of Behrimaz Stream, Turkey. *Environmental Monitoring and Assessment*, *159*, 543-553.

Vikas, V., & Lokendra Singh., T. (2012). Multivariate Statistical Approach for the assessment of Groundwater Quality in Ujjain City, India. *Environ. Science &Engg*. *54(4),* 33-543.

Zhang Q.L., Shi, X.Z., Huang, B., Yu, D.S., Obornc, I., Blombacke, K., Wang, H.J., Pagella TF, Sindai FL (2007). Surface water quality of factory based and vegetable – based peri – Urban areas in the Yangtze River delta region, *China Catena*, 69(1): 5– 64.

Zhou, F., Guo, H.C., & Liu, Y. (2007). Quantitative identification and source apportionment of anthropogenic heavy metals in marine sediment of Hong Kong. *Environmental Geology*, Vol . 53, pp. 295 – 305.