October 2012 1(1)

*The Macrotheme Review*

***A multidisciplinary journal of global macro trends***

# Development of Water Quality Index for assessment of quality of water in the coastal water of Bay of Bengal at Visakhapatnam zone, India

Sangeeta Pati*, M. K. Dash, C. K. Mukherjeec

Indian Institute of Technology, patisangeeta@gmail.com*

**Abstract**

*This paper proposes a technique to describe the water quality quantitatively for the coastal water from multiple measured water quality parameters through an index called Water Quality Index (WQI) using different multivariate statistical techniques. Numerical classification and discriminant analysis were used to derive the WQI. Cluster analysis (CA) technique is carried out in the Euclidean space to classify the groups. Three groups of water quality i.e. good, average and poor have been identified based on the nutrient load. Discriminant analysis was used to generate a discriminant function for developing WQI. A suitable discriminant function has been generated to describe the maximum variance in the data set. Group centroids are calculated using the discriminant function for each group. The Water Quality Index (WQI) has been generated using these centroid values and found within the range -3.2 to +2.1. The negative value representing the centroid for potentially good and positive value representing the poor water quality. The interval of this range establishes the average water quality of coastal water. The application of the WQI is suggested as a very helpful tool that enables the public and decision makers to evaluate the coastal water quality.*

**Keywords:** *quality of water, India*

## 1. Introduction

In order to understand easily about the status of the water quality of different time and places, the Water Quality Index (WQI) must reduce the vast quantity of water quality information into its simplest form. Being a comprehensive indicator, it turns out to be multifaceted information of water quality parameters into a single number with scientific basis. But many times it is difficult to interpret, rendering meaningful conclusion from a complex data matrix comprised of a large number of physical, chemical and biological parameters, which differ in time and space (Almeida, Quintar, González and Mallea, 2007; Dixon and Chiswell 1996). To analyze the complex data matrix, different multivariate statistical techniques like cluster analysis (CA), discriminate analysis (DA), principal component analysis (PCA) and factor analysis (FA) have been used earlier.

The coastal ecosystem of east coast of India is vulnerable. Therefore, regular monitoring of WQ is very much required to know the status of the coastal ecosystem.

The purpose of the present study is to develop WQI for coastal water of east coast of India (Visakhapatnam zone) using different multicriteria analysis and to analyze the seasonal variation of water quality for the study region.

## 2. Materials and methods

### *2.1 Study area*

The study area is located at the coastal area of Visakhapatnam, India (Fig. 1) stretching from latitude 17°5' N to 17°7' N and longitude 83°18' E to 83°3'E. This area has significant importance in terms of fishery and aquaculture, particularly for cage-culture which has been started very recently having high socio-economic value to fishermen community of the area.

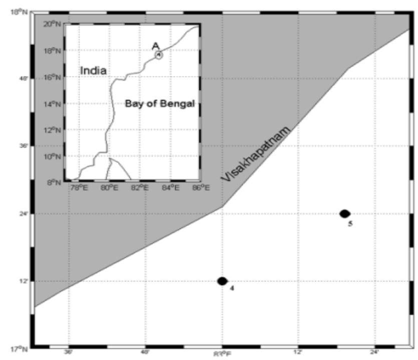


**Fig. 1**. Map showing the study area with location of water quality monitoring stations (4 and 5)

### *2.2. Data used*

Water samples were collected monthly at two depths; 5 – 10 m and 25 – 30 m; at two stations located in the continental shelf from 2007 to 2010 (Fig. 1). Different water quality parameters like ammonia nitrogen ($NH_3$-N), Nitrite nitrogen ($NO_3$-N), phosphate phosphorous $PO_4$-P), Biochemical oxygen demand (BOD), TSS (Total suspended solid) are considered for developing the WQI. Analysis of water samples were done according to standard methods described in (APHA-AWWA-WPCE, 1995). The pH and Dissolved oxygen (DO) of each water sample were measured at the sampling points by a digital pH and DO meter, respectively.

### *2.3. Data treatment and multivariate statistical analysis*

The water quality data sets, described above were subjected to different multivariate statistical analysis like cluster analysis (CA) and discriminant function analysis (DA) to obtain a water quality index. Concentration orders of all physico-chemical parameters differed greatly and the statistical results were highly biased by the parameters with high concentration. Therefore, standardization (z-scale) was made on each parameter prior to the statistical analysis (Singh, Malik, Mohan and Sinha, 2004; Shrestha and Kazama, 2007). It minimizes the variance in the data and eliminates the influence of different units of measurements. In this study, data used in cluster analysis were standardized using Eq. 1 to provide equal weightage to each variable.

$$(x_{ij} - x_i)/\sigma_i = z_{ij} \quad (1)$$

where $x_{ij}$ is the matrix element to be standardized, $x_i$ is mean value of the data of the $i^{th}$ row, $\sigma_i$ is standard deviation of the $i^{th}$ row data and $z_{ij}$ *is* the standardized value.

All mathematical and statistical computations were made using MATLAB and SPSS Statistic 17.0.

### *2.3.1. Cluster analysis (CA)*

Cluster analysis is used to group the data into different clusters based on the similarity features and characteristics they possess (Vega, Pardo, Barrado and Deban, 1998). It was performed based on Euclidean distance. In this study hierarchical agglomerative clustering method, using squared Euclidean distance as proximity measure and Ward's linkage algorithm were used to group the total nutrient data into different groups. The resulting groups from cluster analysis are generally illustrated by a dendrogram (tree diagram) (McKenna, 2003) and provides a visual summary of the clustering process.

The squared Euclidean distance between two objects *i* and *i'* is (Huang, Qiu and Guo, 2009)

$$D^2_{ii'} = \sum_{j=1}^{m} (x_{ij} - x_{i'j})^2 \quad (2)$$

where *m* is the number of variables.

### *2.3.2. Discriminant analysis (DA)*

Discriminant analysis (Jennrich and Sampson, 1983) was used to find one or more functions of quantitative measurements that make it possible to discriminate among groups previously established. The DA technique builds up a discriminant function for each group. It operates on each raw data matrix and constructs a discriminant function for each group (Singh, Malik, Mohan and Sinha, 2004; Shrestha and Kazama, 2007) as in Eq. 3.

$$f(G) = k_i + \sum_{j=i}^{n} w_{ij} p_{ij} \quad (3)$$

where *f(G)* discriminant function, *i* is the number of groups *(G)*, $k_i$ is the constant inherent to each group, *n* is the number of parameters used to classify a set of data into a given group, $w_{ij}$ *is* the weight coefficient, assigned by DA to a given selected parameters $(p_{ij})$.

## 3. Results and Discussion

### *3.1. Development of WQI*

The results of cluster analysis yielded a dendrogram (Fig. 2) which grouped all data set into three groups of clusters based on the similar characteristics. To validate the separation of clusters within each group, significance test is done using F-statistic under 1% level of significance and the clusters were found significantly different from each other. To analyze the water quality of each group, different water quality parameters were revisited. It was found that group 1 is characterized by high concentration of DO and low concentration of nutrient loads ($NH_3$-N and $PO_4$-P) and was represented as good water quality. Group 2 had comparatively less DO and moderate nutrient load. So it was considered as average water quality. Group 3 possessed lesser DO and high concentration of nutrient loads and was characterized as poor water quality.

DA was conducted with a series of previously established three groups. Two sets of discriminant functions were generated taking two groups at a time. The statistical summary of the DA is shown in Table 1.

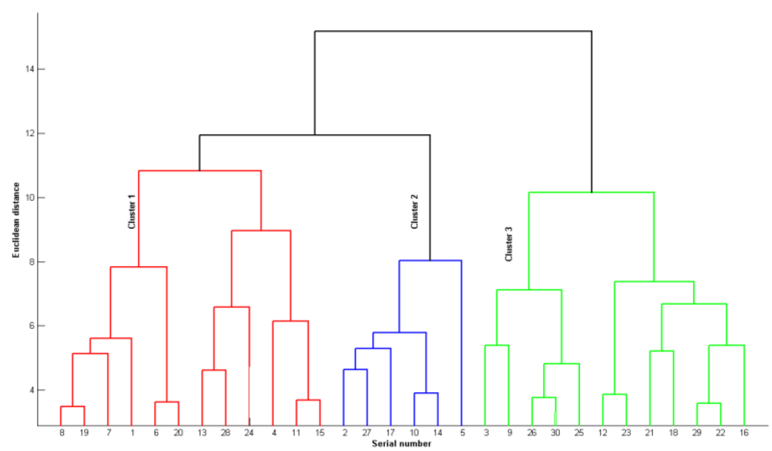


**Fig. 2.** Dendrogram for group average clustering

Table 1
Statistical summary for first and second discriminant functions

| Discriminant function | Eigen value | % of variance | Canonical correlation |
|---|---|---|---|
| 1 | 1.527[a] | 57 | 0.782 |
| 2 | 1.185[a] | 43 | 0.736 |

a. first two discriminant functions were used in the analysis

The eigenvalues, total variance, and canonical correlations obtained in this process indicate that the contribution of the second discriminant function is less in comparison to the first. Thus, the first discriminant function has been considered in this study for further analysis. The DF generated is given in Eq. 4.

$$D = -0.653[DO]-0.587[pH] +2.080[NH_3\text{-}N]-3.036[NO_3\text{-}N] +2.007 [PO_4\text{-}P] +1.283 [BOD]-0.003[TSS] +5.256 \quad (4)$$

Where, *D* is the discriminant score and *DO, pH,* $NH_3$*-N,* $NO_3$*-N,* $PO_4$*-P, BOD* and *TSS* are the independent variables. The standardized coefficients generated from DA for each variable are: DO (-0.497), pH (-0.080), $NH_3$-N (0.424), $NO_3$-N (-0.477), $PO_4$-P (-0.311), BOD (0.808), TSS (-0.026). The discriminant function coefficients of this equation highlight the relative importance of the variables. It indicates that the main contributing parameters to the equation are $NH_3$-N, $PO_4$-P and BOD. These inorganic nutrients are considered very important in the differencing the groups. The efficacy of the first function that was derived from the discrimination process can be assessed by the proportion of correctly classified observations. In our case, the efficacy obtained was high, with 90.06%. The classification matrix obtained from DA which shows the actual group to which a sample belongs versus the predicted one is shown in Table 2. Scatter plot for all the observations points in the space of 2 DF is shown in Fig. 3.

Table 2
Original group versus predicted group generated by first discriminant function

| Original group | Total observations (N) | Predicted group | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| 1 | 56 | 51(91.1%) | 5 (8.9%) | 0 (0%) |
| 2 | 50 | 3 (6%) | 46 (92%) | 1(2%) |
| 3 | 22 | 1(4.5%) | 2 (9.1%) | 19(86.4%) |

The discriminant procedure generates centroids or central values for every group. It was found that the D-values were uniformly distributed about zero. Therefore, WQI value was computed by simple division of discriminant score D by the standard deviation to get an adequate scale for water quality index which is expressed in Eq. 5. Negative value close to -3.2 indicates good water quality as it representative of low nutrient concentration and positive value close to +2.1 indicates poor water quality possessing high nutrient concentration.

$$WQI= D/1.591 \quad (5)$$

where, D is the discriminant score generated using first discriminant function and 1.591 is standard deviation of the total data set used in this study.

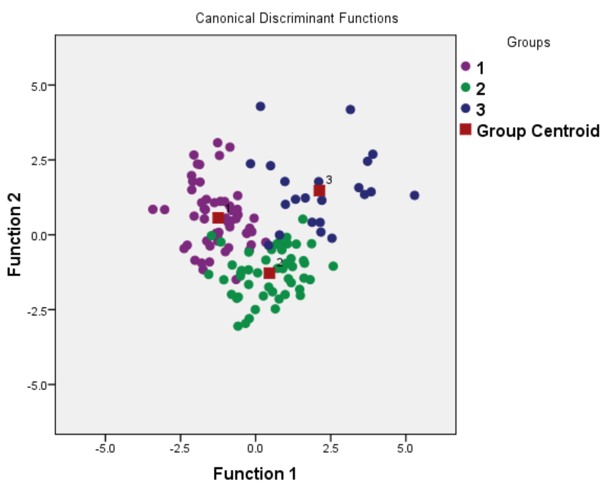


**Fig. 3.** Scatter plot of all samples in space of the two discriminant functions.

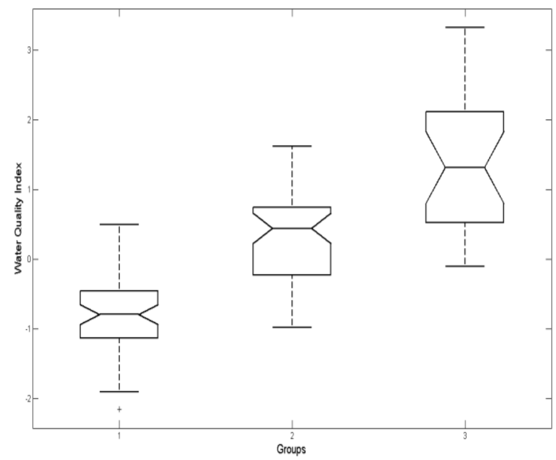


**Fig. 4.** Boxplot of water quality Index for three clustering groups (the solid line connecting the notch represent medians; the lower and upper whisker represent 10$^{th}$ and 90$^{th}$ percentile respectively. The lower and upper boundaries of boxes represent 25$^{th}$ and 75$^{th}$ percentiles respectively.

The water quality index obtained for three groups were tested with a box plot coupled with Kruskal-Wallis rank sum test, which compared the medians of the groups. Significant differences among three groups were highlighted with rejection of the null hypothesis at the 5% significance level ($\alpha$=5%). A very small probability ($p<0.001$) of overlapping among the groups was found in this test. The comparisons of medians among the three clustering groups are shown in Fig. 4. Group 1 exhibits clear separation with a median of -0.7935 compared to other two groups. The distributions of 10% and 90% population are shown by lower and upper whiskers and the distribution of 25% and 75% of population are shown by lower and upper boundaries of the box (Fig. 4). Figure shows that WQI for group 1 is uniformly distributed about the meridian, though a few outlier points exist. Group 3 represents relatively poor water quality having high nutrient content with a median of 1.31. The group 2 represents average water quality, which stands in between other two groups.

## 4. Conclusions

In this study various multivariate statistical techniques were adopted to develop WQI for the coastal water of Visakhapatnam, India using different water quality parameters. Hierarchical cluster analysis was used to classify the total data set into 3 major groups based on similarity of water quality. Statistically significant 3 groups were ranked as good, average and poor according to the presence of DO and $NH_3$-N and $PO_4$-P as they are the potential indicators of water quality. Suitable discriminant function was generated using DA to develop the WQI. WQI value close to -3.2 indicated good water quality as it was having low nutrient load and that close to +2.1

indicates poor water having high nutrient load. WQI showed significant differences among the three groups. Group 1 and group 3 possessed good and poor water quality respectively based on the nutrient loads ($NH_3$-N and $PO_4$-P) and DO concentration. The information obtained through this work may be used to improve the management practices of the coastal environment considering the selected water quality constituents. This study also illustrates the usefulness of WQI for analyzing the complex nutrient data for assessing the coastal water and identifying different pollution sources for variation in water quality

## Acknowledgements

We are thankful to Central Marine Fisheries Research Institute, Visakhapatnam, India for supporting this research. We also like to acknowledge the help and support provided by Agricultural and Food Engineering Department and Oceans, Rivers, Atmosphere and Land Sciences Centre, Indian Institute of Technology, India for availing the facility to conduct the research work.

### References

Almeida, C.A., Quintar, S., González, P., Mallea, M.A., 2007. Influence of urbanization and tourist activities on the water quality of the Potrero de los Funes River (San Luis – Argentina). Environmental Monitoring Assessment, 133, 459–465.

APHA-AWWA-WPCF, 1995. Standard methods for examination of water and wastewater. 19th ed. Washington, District of Columbia, American Public Health Association, American Water Works Association, Water Pollution Control Federation.

Dixon, W., Chiswell B., 1996.Review of aquatic monitoring programme design. Water Research 30, 1935-1948.

Huang, J.Y., Qiu, Y.B., Guo, X. P., 2009.Cluster and discriminant analysis of electrochemical noise statistical parameters. Electrochemica Acta, 54, 2218-2223.

McKenna Jr., J.E., 2003. An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. Environmental Modelling & Software, 3, 205-220.

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.

Singh, K.P., Malik, A., D. Mohan, Sinha, S., 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomatiriver (India)—A case study. Water Research, 38, 3980–3992.

Vega, M., Pardo, R., Barrado, E., Deban, L., 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Research, 32, 3581-3592.