

Assessment of Water Quality in Imo River Estuary Using Multivariate Statistical Method

Akpan, I. I.

Department Of Science Laboratory Technology Akwa Ibom State College Of Arts And Science

Abstract: The water quality of Imo River Estuary, the Niger Delta region was studied for a duration of 12 months. This study was aimed at the assessment of water quality parameter of the water body. In order to have an indepth knowledge to the physical and chemical processes as well as their associated spatial distribution, the study analyses some parameters recorded at the three sampling sites through multivariate statistical methods. The principal component analysis (PCA) and factor analysis (FA) was employed to extract and recognize the major underlying factors contributing to the variations among the water quality measured. Results indicate that three principal components, that is nutrients, organic and meteorological factor account for 99.91% of the total variance among the water quality parameters. The spatial distribution of principal components further confirms that nutrient sources constitute the main pollutant contribution.

Keywords: Assessment, Principal Component Analysis, Factor Analysis, Estuary, Source

I. Introduction

Water quality is the major factor which control healthy and diseased states in both humans and animals. Surface water quality is an important component of the natural environment and a subject of serious concern nowadays. The variation of water quality are essentially the combination of both anthropogenic and natural combinations. In general, the anthropogenic outputs constitute a constant source of pollution, whereas surface runoff is a seasonal phenomenon which is affected by climate within the water catchment basin (Liu et al., 2011). Among them, due to the much human duties, the anthropogenic inputs from a variety of sources are primarily the main factors which affect the water quality of most water bodies, mostly those close to a highly urbanized areas (Akpan, 2012).

Various studies have been carried out on anthropogenic pollutants of the ecological systems (Heikka, 2007; Akpan, 2012; Palma *et al.*, 2010). Because of the spatial and temporal variations in water health quality, a monitoring procedure which provides a reliable estimation of the quality of water bodies is essential. The monitoring results produce a large and complicated data matrix that is cumbersome to explain and interpret so as to draw conclusion that is meaningful. Multivariate statistical methods are tools for analyzing a large numbers of samples assessing anthropogenic effects on water health quality and the biodiversity.

The usage of different multivariate statistical techniques, such as principal component analysis (PCA), factor analysis (FA), cluster analysis (CA) aid in the interpretation of complex data matrixes for a better knowledge of water quality and ecological characteristics of the sampled area. These methods give the identification of possible sources that affect water environmental systems and offer a valuable tool for reliable management of water resources as well as rapid solution for pollution issues (Singh *et al.*, 2004). Multivariate statistical methods have been widely used to analyse and evaluate water quality and are useful to study variations (temporal and spatial) occasioned by natural and anthropogenic factors linked to seasonality (Akpan, 2012, Kim et al., 2009).

The objective of this study was to analyse water quality parameters in water samples collected on monthly basis from January 2010 – December, 2010 in Imo River Estuary, Niger Delta Area, Nigeria. The data matrix was treated to some multivariate statistical methods.

II. Materials and Methods

Study Site and Sample Collection

The Imo River Estuary (4°.33' - 7°.36', 4°.36' - 7°.41' (Fig.1) originates in the vicinity of Okigwe (Imo State, Nigeria) and takes a southerly course until it empties into Atlantic Ocean in the Bight of Bonny. It drains an area of 9,100km² with a relatively small flood plain area of 160km² (Udo, 1995). The Imo River basin is situated in the tropical rain forest belt with an equatorial climate which comprises two major seasons. The annual mean rainfall is 350mm with July as the middle of the wet season. The relative humidity is high (over 80%) because of the warm wet air masses that prevail during the period. The River is an important site for studying water quality parameters. Recently, the river has been subjected to pollution sources, therefore the investigation of the water quality is very essential.

The main channel and creeks of the estuarine zone are fringed by mangrove swamp forests on tidal mudflats. The substrate of the saline swamp consists of peaty fibrous mud. The macrophytes are dominated by the native red mangroves and the exotic nipa palm.

The sampling network including three sites were designed to cover a wide range of key locations accounting for the upstream – downstream: site I (Esene Creek), which is close to the housing estate, site II (ALSCON Harbour), which the harbor area of the ship and site III (Jaja creek), which is close to the effluent discharge point of Aluminium Smelting Company.(Fig.1)

Water temperatures, pH was measured insitu with a thermistor chain (Templine, Apprise Technologies, Inc. Duluth, MN, USA) and Graffin pH meter (model 80), after standardization with buffer solution. Conductivity was measured insitu using a temperature – compensated battery power electronic switch-gear meter. Dissolved oxygen concentration was measured with a dissolved oxygen metre (yellow springs instruments company USA, model 550A). The water samples were collected using an open water sampler and were analysed and measured in the laboratory to obtain Total Alkanity (TA), total suspended solids (TSS), Total Hardness (TH), Biochemical Oxygen Demand (BOD) and nutrients.

PRINCIPAL COMPONENT ANALYSIS

Principal component analysis type of multivariate analysis, which is concerned with data analytical method focused on a particular collection of variables. Consider the form of the first principal component. The score for individual on component, c_{i1} , uses weight w_{i1}, \dots, w_{ip1} in the linear combination:

$$c_{i1} = y_{i1}w_{i1} + y_{i2}w_{i2} + \dots + y_{ip}w_{ip1} \dots \dots \dots (1)$$

The linear combination was chosen so that the sum of squares of C_1 is large as possible subject to the condition that $W_{i1}^2 + \dots + W_{ip1}^2 = 1$. The second principal component is another linear combination of y_i

$$c_{i2} = y_{i1}w_{i12} + y_{i2}w_{i22} + \dots + y_{ip}w_{ip2} \dots \dots \dots (2)$$

where the variance C_2 is the maximal, subject to the conditions that Correlation (C_1, C_2) = 0 and that $w_{i12}^2 + \dots + w_{ip2}^2 = 1$. The criterion of summarizing the information in P variables by a few components is a desirable as a way of reducing the number of variables needed in an analysis (Helena *et al.*, 2000).

Factor analysis (FA) follows Principal Component Analysis (PCA), FA focuses on reducing the contribution of less significant variables to simplify more of the data structure from PCA. This was achieved by rotating the axis defined by PCA based on well established rules, PCA of the normalized variables was done to extract significant Principal Components (PCs) and to further reduce the contribution of variables with minor significance (Brumelis *et al.*, 2000) Abdul-Wahah, 2005.

The FA can be written as

$$y_{ji} = f_{j1}z_{ji} + f_{j2}z_{i2} + \dots + f_{jm}z_{im} + e_{ij} \dots \dots \dots 3$$

where Y is the measured variable, f is the factor loading, z is the factor score, e is the residual term accounting for errors, i is the sample number and m is the total number of factors.

III. Results And Discussion

The measured results of the water quality parameters of Imo River Estuary at the three sampling sites from January, 2010 – December, 2010 are shown in Table 1.

Principal Component Analysis and Pollution Establishment

Pattern recognition of correlation among the water quality parameters was best summarized by Principal Component Analysis/Factor Analysis (PCA/FA). The Bartlett test was used on the data set to determine the suitability of these data for PCA/FA. The covariance matrix which coincided with the correlation matrix is presented in Table 2, because FA/PCA was applied to normalized data. There are some positive correlation some variables and also negative correlation between some (Table 2).

The scree plot was applied to a determine the member of PCs to be retained to understand the underlying data structure. Based on the scree plot and eigenvalues >1 criterion, three factors were chosen as principal factors, explaining 99.91% of the total variance in the data set. The corresponding VFs, variable loadings, eigenieclues and explained variance are presented in Table 3.

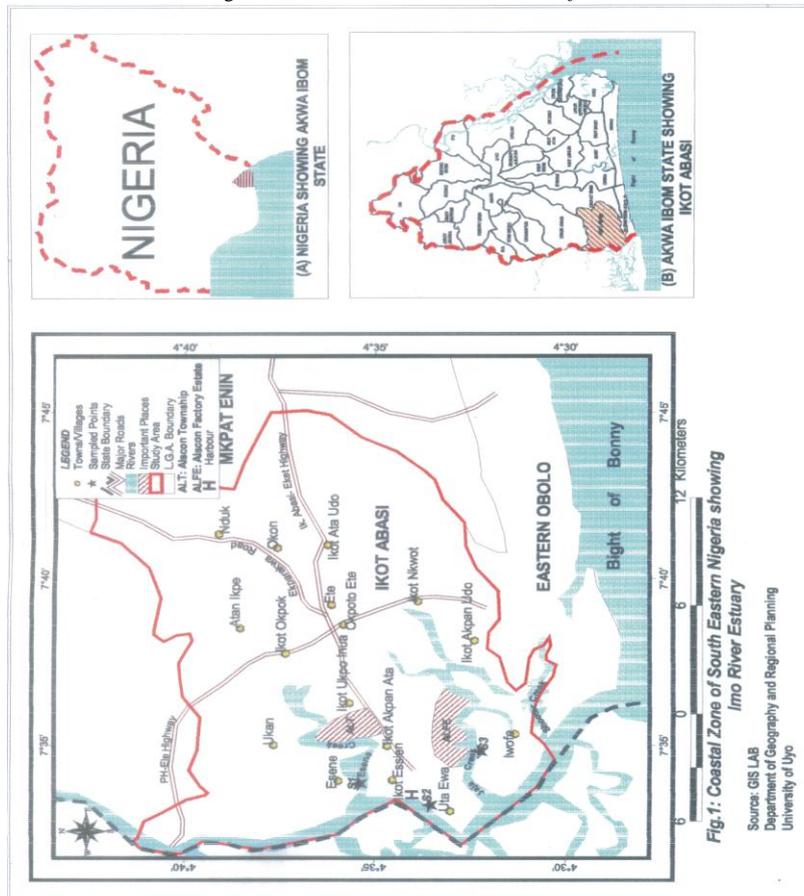
Liu *et al.*, (2003), segmented the factor loadings as “strong”, moderate’ and Weak, corresponding to absolute loading values of > 0.75, 0.75 – 0.50, and 0.50 – 0.30, respectively. The first factor (Vf_1), explaining 45.86% of the total variance, had a weak positive loading on Na (sodium), K (potassium), SO_4 (sulphate) $PO_4 - P$ (phosphate-phosphorus), TA (Total Alkalinity), Cl (chloride), DO (dissolve oxygen) and Ca (calcium). This could be attributed as nutrient source pollution indicator. Vf_2 , which explained 38.95% of total variance, had weak positive loadings on Na (sodium), BOD (Boochemical Oxygen Demand) and TSS (Total Suspend Solids) acid represent organic source. VF_3 explaining a moderate positive loading on Temp (Temperature) and A (Acidity) and represented the metrological factor. The analysed results showed that FA/PCA is a reliable means to determine the main factors which affect water quality in Imo River Estuary.

IV. Conclusion

Water quality data collected from three monitoring stations located around Imo River estuary, Niger Delta Area has been examined by display methods, that is principal component analysis and factor analysis to yield correlations between variables and water quality similarity in the estuary. The PCA/FA assisted to extract and recognize the factors responsible for water quality variations. PCA/FA established two latent factors that explained 99.91% of total variance, namely: nutrient –pollution source and organic pollution source. This explains that nutrient sources and organic pollutions source are the most important factors that contribute to affect the water quality of Imo River Estuary. Therefore, this work underscore the important of multivariate statistical methods for the analysis and explanation of complex array of data, water quality monitoring and determination of pollution sources to the water bodies.

References

- [1]. Akpan, I. I. (2012). Physico-chemical and Biological Studies of Imo River Estuary, Nigeria, Ph.D Thesis, Department of Fishery, Michael Okpara University of Agriculture, umudike, 184.
- [2]. Heikka, R. A. (2005). Multivariate Monitoring of Water Quality: A Case Study of Lake Simple, Finland. *J. Chmonet*, 22, 747-751.
- [3]. Palma, P., Albarenga, P., Palma, V. L., Fernandes, R. M., Soares, A. M. V. M., Barbosa, I. R. (2010). Assessment of Anthropogenic Sources of Water Pollution Using Multivariate Statistical Techniques. A Case Study of Alqueva’s Reservoir, Portugal, *Environ. Monit. Assess* 165, 539-552.
- [4]. Helena B., Pardo, B., Vega, M., Barrado, E., Fernandez, J. M., Fernandez, L. (2000). Temporal Evolution of Groundwater Composition in an Alluvial Aquifer (Pisuerga River, Spain) by Principal Component Analysis. *Water Res.* 34, 807-816.
- [5]. Brumelis, G., Lapina, L., Nikodemus, O., Tabors, G. (2000). Use of an Artificial Model of Monitoring Data to Aid Interpretation of Principal Component Analysis. *Environ. Modell. Softw.* 2000, 15, 755-763.
- [6]. Singh, K. P., Malik, A., Mohan, D., Sinha, S. (2004). Multivariate Statistical Techniques for the Evaluation of Spatial and Temporal Variations in Water Quality of Gomti River (India): A Case Study: *Water Res.* 38, 3980 – 3992.
- [7]. Abdul-Wahab, S. A., Bakheit, C. S., Al-Alawi, S. M. (2005). Principal Component and Multiple Regression Analysis in Modelling of Ground-level Ozone and Factors Affecting its Concentration. *Environ. Modell. Softw.* 20, 1263-1271.



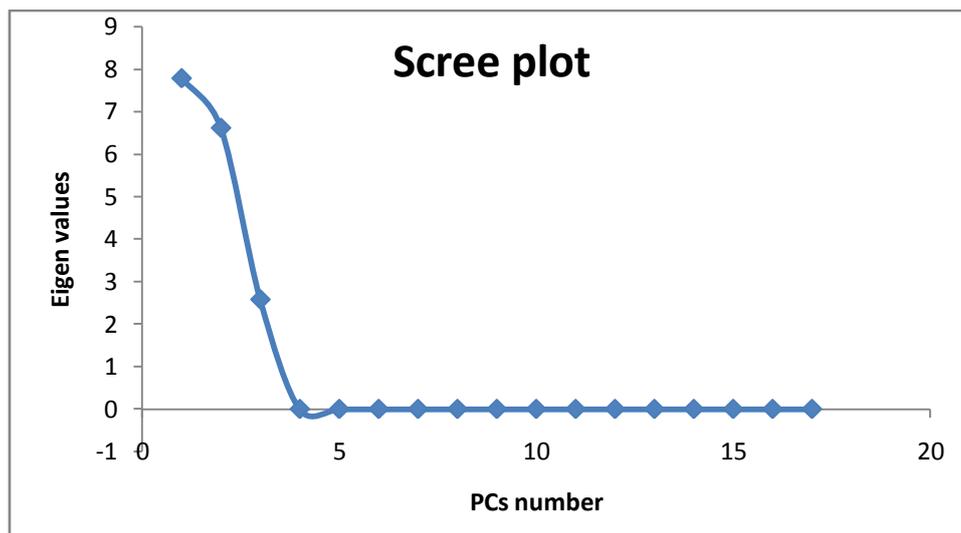


Fig. 2: Scree plot of the characteristics roots (Eigen values) of principal component analysis.

Table 1: The Range, Mean and Standard Deviation of Water Chemistry Parameters at 3 Sites of the Imo River Estuary

Parameters	S1		S2		S3	
	Range	X ± SD	Range	X ± SD	Range	X ± SD
Temp (°C)	22.90 – 28.10	26.79 ± 1.07	28.10 – 27.3	26.14 ± 0.54	24.20- 26.70	25.39 ± 0.65
pH	4.70 -6.50	5.60 ± 0.45	5.20 – 6.80	6.01 ± 0.39	6.00 – 8.30	7.02 ± 0.73
Ec (µ/cm)	10.10 – 15.1	13.2 ± 3.55	12.13 – 19.23	17.80 ± 4.94	18.1 – 24.10	22.41 ± 843.12
Na (mg/l)	0.0001 – 0.400	0.15 ± 0.11	0.07 – 0.90	0.39 ± 0.22	4.11 – 8.00	6.33 ± 1.24
K (mg/l)	0.01 – 1.00	0.59 ± 0.30	0.16 -14.10	1.48 ± 2.74	4.81 – 6.20	39.59 ± 15.28
So4 (mg/l)	0.001 – 0.300	0.04 ± 0.58	0.01 – 0.80	0.15 ± 0.20	0.06 – 6.01	2.85 ± 1.45
Po4 (mg/l)	0.00 – 0.20	0.02 ± 0.04	0.00 – 0.70	0.08 ± 0.18	0.01 – 0.80	0.13 ± 0.18
Ta (mg/l)	11.20 – 30.10	18.81 ± 5.07	13.20 – 41.20	25.99 ± 8.06	20.10 -88.20	55.31 ± 20.49
Cl (mg/l0	17.18 – 70.20	29.78 ± 13.08	20.10 – 90.18	38.99 ± 17.60	550.40 – 1984.0	1412.0 ± 515.04
TH (mg/l)	11.38-17.18	14.24 ± 1.49	14.17 – 21.50	16.47 ± 1.85	16.15 – 42.17	23.81 ± 7.00
Mg (mg/l)	5.01 -22.30	13.26 ± 4.64	12.00 – 37.00	23.73 ± 6.07	25.38 – 58.30	37.31 ± 9.34
Do (mg/l)	2.90 – 5.31	3.94 ± 0.42	3.70 – 5.81	4.39 ± 0.48	5.10 – 7.10	5.78 ± 0.53
A (mg/l)	9.80 -18.90	15.87 ± 2.52	10.20 – 21.50	18.60 ± 2.42	18.54 – 30.10	23.32 ± 2.71
TDs (mg/l)	2.00 – 100.20	21.59 ± 28.62	3.00 -702.10	48 ± 1509.78	15.68-1821.0	1114.00 ± 467.48
BOD (mg/l)	1.58-2.18	1.95 ± 0.18	2.00 – 3.15	2.27 ± 0.32	2.15 – 4.31	3.19 ± 0.47
TSS (mg/l)	0.002-0.370	0.08 ± 0.13	0.003 – 0.420	0.10 ± 0.15	0.22 -0.62	0.44 ± 0.11
Ca (mg/l)	2.30-50.20	26.05 ± 10.41	27.30 – 65.10	48.70 ± 11.14	50.30-90.10	69.35 ± 12.39

Table 2: Pearson Correlation Coefficient of Water Chemistry Parameters in Imo River Estuary

Tem	pH	EC	Na	K	SO ₄	PO ₄	TA	Cl	TH	Mg	DO	A	TDS	BO D	TSS	Ca
pH	-0.47**															
EC	-0.55**	0.78**														
Na	-0.54**	0.77**	0.95**													
K	-0.48**	0.77**	0.92**	0.94*												
SO ₄	-0.47**	0.79**	0.89**	0.91**	0.86**											
PO ₄	-0.18	0.29*	0.19	0.23	0.22	0.28*										
TA	-0.52**	0.73**	0.85**	0.85*	0.81**	0.85**	0.19**									
Cl	-0.52**	0.75**	0.95**	0.95*	0.92**	0.92**	0.22**	0.89**								
TH	-0.39*	0.63**	0.71**	0.77*	0.67**	0.78**	0.16	0.85**	0.75**							
Mg	-0.52**	0.76**	0.79**	0.18*	0.75**	0.81**	0.33*	0.84**	0.77**	0.77**						
DO	-0.58**	0.75**	0.79**	0.85*	0.78**	0.82**	0.35*	0.75**	0.79**	0.70**	0.81**					
A	-0.55**	0.63**	0.71**	0.75*	0.74**	0.69**	0.09	0.77**	0.76**	0.61**	0.69**	0.71*				
TDS	-0.48**	0.70**	0.86**	0.89*	0.89**	0.85**	0.25*	0.78**	0.88*	0.70**	0.74**	0.75*	0.71**			
BO D	-0.45**	0.75**	0.79**	0.84*	0.82**	0.75**	0.21	0.74**	0.83**	0.69**	0.68**	0.75*	0.77*	0.76**		
TSS	-0.44**	0.04**	0.77**	0.77*	0.72**	0.67**	0.05	0.64**	0.75**	0.59**	0.56**	0.71*	0.60**	0.69**	0.77**	
Ca	-0.47**	0.66**	0.72**	0.77*	0.71**	0.78**	0.31**	0.77**	0.76**	0.72**	0.82**	0.75*	0.69**	0.71**	0.73**	0.61**

** Values are statistically significant at P < 0.01.

* Values are statistically significant at P < 0.05.

Table 3: Loading of Parameters on Significant Vfs for Water Quality Data Set

Parameters	Three Significant PCs		
	VF ₁	VF ₂	VF ₃
Temp (°C)	0.0023	0.1876	0.5450
pH	-.2704	0.2254	-0.1903
Ec (µ/cm)	-0.2317	-0.2873	0.1163
Na (mg/l)	-0.4379.	-0.3740	0.1509
K (mg/l)	0.3044	-0.1949	0.1001
So4 (mg/l)	0.3363	0.1268	0.0675
Po4 (mg/l)	0.3044	-0.1949	0.1001
Ta (mg/l)	0.3410	0.1186	-0.0065
Cl (mg/10)	0.3109	0.1927	0.0099
TH (mg/l)	0.2211	-0.1672	-0.4099
Mg (mg/l)	0.2509	0.2552	0.1740
Do (mg/l)	-0.3100	0.1894	-0.0709
A (mg/l)	-0.1760	0.2859	-0.3001
TDs (mg/l)	-0.2828	0.2376	-0.0305
BOD (mg/l)	-0.0862	0.3466	0.2381
TSS (mg/l)	-0.0125	0.3883.	-0.0102
Ca (mg/l)	0.3084	0.1975	0.0126
Eigenvalue	7.79	6.62	2.58
Percentage of total variance	45.86	38.95	15.10
Cumulative percentage of variance	45.86	84.81	99.91