

Modelling and Prediction of Water Quality Index of Selected Pond Water InAbohMbaise Local Government Area, Imo State, Nigeria.

Comment [H1]: Area of Imo State, Nigeria

Comment [H2]: There should not be a full stop after a title

Comment [H3]:

Abstract

The aim of this study was the development of water quality models against water quality parameters from 5 selected ponds in Aboh-Mbaise local government area (LGA) of Imo state.

Comment [H4]: The aim of this study is to develop water quality models against water quality parameters

Water quality index (WQI) as dependent variable computed based on water quality parameters which were taken as independent variables and modelled as multiple linear regression. Given that there are over 25 water quality parameters (physiochemical, heavy metals and microbials), it was necessary to adopt factor reduction technique using principal component analysis. In this approach, 3 principal component factors were generated having corresponding factors (independent variables of 5, 6 and 5 respectively). The resulting multiple regression for the 3 principal component factors yielded goodness of fit of 92.3, 99.6 and 96.6% as well as root mean square error (RMSE) of 66.673, 0.672 and 51.968, respectively. The model verification was accomplished by plotting the computed WQI against predicted values from the developed models and the best option was the one with 99.0% R^2 value. The model output is relevant in WQI prediction given the applicable water quality characteristics. The computed WQI will guide the decision on choice of water source for domestic purposes.

Key words: Modelling, Prediction, Water Quality Index, PrincipalComponent Analysis, Multiple regression, Aboh-Mbaise.

Introduction

Water is used for different purposes-domestic, industrial, commercial and recreational purposes, etc. Water use is dependent on its quality which may be transcribed by water quality index. Anthropogenic activities adversely affect water quality including waste disposal, mining, soil erosion, [1] and heavy metal pollution.[2]WQI metrics identifies categories of water in terms of excellent, good, fair and poor quality. WQI models have been developed in the last 5 decades and a variety of them have been published in literature.

The first WQI model was developed by Horton et al [3] following which, a more robust model was developed by Brown in 1970 [4] and later modified by Brown to incorporate about 13 parameters with support from the National Sanitation Foundation NSF-WQI.

Presently, over 35 WQI models have been introduced by various countries to evaluate surface water quality around the world [5, 6, 7, 8]. That is to say, there is no uniformity in the development of WQI for general use. The choice among the existing WQI models is researcher dependent. In a situation where water quality models are selected as dependent variables against water quality parameters (physiochemical, heavy metals and microbials) as independent variables, there is always need to adopt factor reduction with respect to the independent variables using principal component analysis [9] et al. The multiple linear regression model of WQI against the water quality parameters is usually calibrated using laboratory results [9] and the model verification is actualized by simple comparison of laboratory results and water quality predicted output. These procedures of calibration and model verification were adopted in the course of this research work.

2. MATERIALS AND METHODS

2.1 Study Area

The study area of this research is made up of three communities in Aboh-Mbaise Local Government Area in Imo State where five water ponds were randomly selected (Fig. 1). The five selected ponds are located in the following villages: Olakwo in Enyioyugu, Umuabazu in Okwuato, Ibeku in Okwuato, Ama-Ukwu in Umuelem and Umuanuma in Nguru. Additional details on study area is as given in the authors earlier paper [10].

2.2 Water Quality Index (WQI)

The Weighted Quality Index was determined for this study by using the Weighted Arithmetic Index Method created by Brown et al., [12] and implemented in Microsoft Excel. The detailed formula is as presented in an earlier publication by the authors [9] and Babatunde et al [11].

Comment [H5]: According to whose authority is this method? and there should be a basis for selection for example may be because of the location of the pond, proximity to anthropogenic activities and so on

Comment [H6]:

Comment [H7]: Mention the name of the author, its better that way

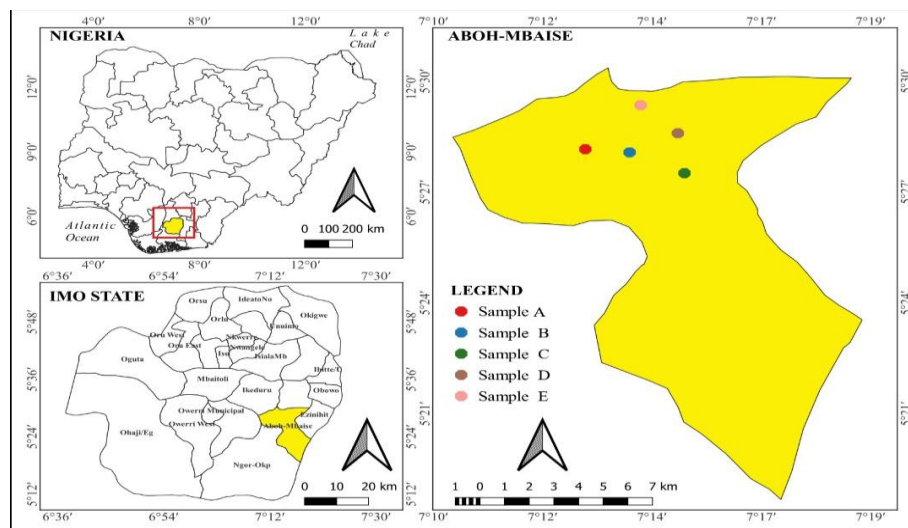


Fig 1. The study area

2.3 Method of Data Analysis

Water samples from the five pond locations were analyzed for physiochemical parameters and the descriptive statistics (mean and SD) of the parameters were computed. Pearson correlation was used to understand the relationship between the physiochemical parameter and understand physiochemical parameters that are highly correlated. Highly correlated independent variables result to violation of multicollinearity when used in developing predictive models. Principal Component Analysis (PCA) was utilized to reduce the dimension of the physiochemical parameters and eliminated multicollinearity due to orthogonal rotation of the final solution. Water Quality Index (WQI) was calculated and regressed against

PCA factor loading scores. Multiple linear regression was used in developing the predictive model and model performance was assessed with Mean Squared Error and goodness of fit R^2 .

3. Results and Discussion

3.1 Water Quality Index Analysis

The WQI values were computed using data from table 1 and presented in table 2.

The summary of the water quality index values (table 2) were computed using water quality data of the 5 selected ponds in table 1.

Table 1: Descriptive Statistic of Physiochemical Parameters

Physiochemical Parameters	Statistic	AmakwuUmuele mem	IbekuOkw uato	UmuabazaOk wuato	UmuanumaN guru	UrakwoEnyio gugu
pH	mean	6.40	6.35	6.00	6.45	6.05
	std	0.00	0.07	0.00	0.07	0.07
Temperature	mean	25.80	27.30	27.25	26.35	26.05
	std	0.99	0.57	0.49	0.35	1.20
Colour, PCU	mean	936.00	606.00	2570.00	980.00	1988.00
	std	0.00	0.00	14.14	0.00	0.00
Electrical Conductivity	mean	132.50	88.00	107.50	138.50	81.00
	std	0.71	1.41	0.71	0.71	1.41
DO	mean	8.55	7.70	7.80	8.20	8.05

Formatted Table

Comment [H8]: This should be Physicochemical

Comment [H9]: Why don't you have a table like this

Comment [H10]:

Comment [H11]:

Comment [H12]:

	std	0.07	0.00	0.00	0.00	0.07
BOD	mean	1.15	0.40	0.50	0.65	0.80
	std	0.07	0.00	0.00	0.07	0.14
COD	mean	148.00	292.00	196.00	104.00	148.00
	std	5.66	5.66	5.66	0.00	5.66
Turbidity	mean	700.20	410.75	891.50	640.50	474.15
	std	0.28	0.35	0.71	0.71	0.21
Total Solid	mean	356.00	198.00	552.50	271.00	304.00
	std	7.07	39.60	12.02	25.46	2.83
Total Alkalinity	mean	12.00	6.00	8.00	20.00	10.00
	std	0.00	2.83	0.00	5.66	2.83
TDS	mean	86.12	57.20	69.88	90.03	52.65
	std	0.46	0.92	0.46	0.46	0.92
TSS	mean	269.88	140.15	482.73	180.98	251.35
	std	7.53	39.60	12.34	25.00	3.75
Nitrate	mean	47.64	29.19	53.11	37.50	46.15
	std	0.09	0.00	0.00	0.10	0.48
Phosphate	mean	23.00	4.00	31.00	21.00	13.50
	std	1.41	0.00	0.00	4.24	0.71
Total Hardness	mean	119.14	88.06	119.14	98.42	150.22
	std	0.00	0.00	0.00	0.00	3.66
Sulphate	mean	100.00	0.00	100.00	50.00	50.00
	std	0.00	0.00	0.00	0.00	0.00

Comment [H13]: Is it only BOD you determined in physicochemical determination? If yes, its not enough

Table 2: Water quality index summary for the five sampling locations

Locations	WQI
UrakwoEnyiogugu	830.6944
UrakwoEnyiogugu	836.1406
UmuabazaOkwuato	1255.045
UmuabazaOkwuato	1255.663
IbekuOkwuato	744.5951
IbekuOkwuato	745.1515
AmakwuUmuelemem	1053.428
AmakwuUmuelemem	1044.023
UmuanumaNguru	1006.742
UmuanumaNguru	976.9529

Comment [H14]: What are the parameters you examined to arrive at this WQI? Don't assume your audience knows, give details to afford repeatability of the research

WQI rating: 0-25=Excellent water quality, 26-50=Good water quality, 51-75=Poor water quality, 76-100=Very poor water quality, >100 unfit for consumption. Source: Brown *et al.* 1972 [12]

The WQI from all the sampling locations indicate poor to very poor water quality. None of the locations fall within the range of good water quality, as all WQI values are above 100, making the water unfit for consumption according to the established criteria. This alarming result underscores a significant issue with the overall water quality in these areas, necessitating urgent attention and remedial measures to ensure safe and healthy water sources for the communities.

3.2 Principal Component Analysis (PCA)

PCA was used to analyze the input data from the physiochemical parameters of the pond water. [8, 12]The PCA analysis revealed distinct relationships between various physiochemical parameters, providing valuable insights into the complex dynamics of water quality (Figure 2).

The three principal components were retained after varimax rotation. The factor loading of the physiochemical parameters on the

principal component are presented in Table 3. Factor loading with score greater than 0.5 loaded strongly on a particular component.

Principal Component 1 (PC1), with prominent loadings from variables such as pH (0.761), Electric Conductivity (0.927), Total Hardness, Total Dissolved Solids (TDS) (0.927), and Total Alkalinity (0.661) represents fundamental water characteristics. It reflects the mineral content, electrical conductivity, and basic chemical composition of the water samples. Higher scores on PC1 indicate water with elevated mineral content and specific chemical properties.

PC2 was influenced significantly by Sulphate (0.890), Total Suspended Solids (TSS) (0.967), Phosphate (0.922), Turbidity (0.933), Total Solid (0.987), and Nitrate (0.906) which emphasizes the presence of pollutants and suspended matter in the water. Locations with higher PC2 scores exhibit higher levels of suspended solids, pollutants, and reduced water clarity due to turbidity.

PC3 was characterized by loadings from variables such as Chemical Oxygen Demand (COD) (-0.745), Iron (-0.768), Dissolved Oxygen (DO) (0.841), and Biochemical Oxygen Demand (BOD) (0.869) which captures the balance between organic and inorganic pollutants and oxygen levels in the water. Higher PC3 scores indicate higher pollution levels and potential oxygen depletion in the water bodies.

Comment [H15]: Adjust the chart to go down a little

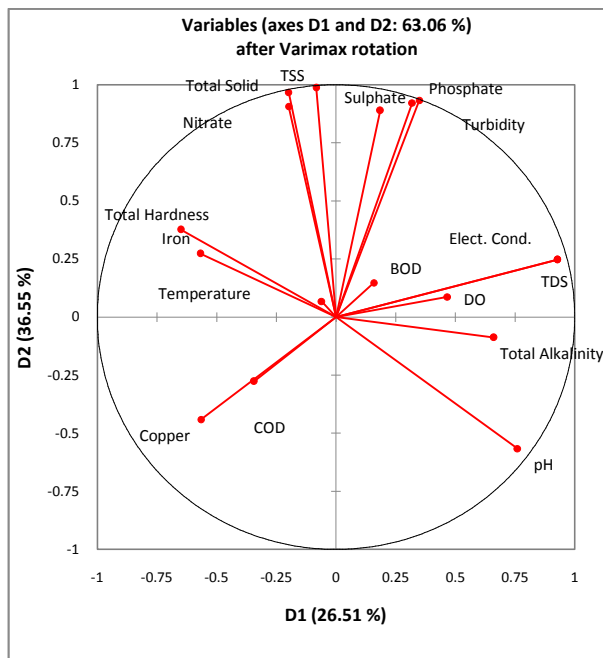


Figure 2: Biplot showing the physiochemical parameters on principal component 1 and 2

Table 3: Factor Loading

	D1	D2	D3
pH	0.761	-0.566	0.143
Elect. Cond.	0.927	0.247	0.276
COD	-0.343	-0.276	-0.745
Sulphate	0.185	0.890	0.387
Temperature	-0.060	0.068	-0.804
Iron	-0.568	0.273	-0.768
Copper	-0.565	-0.441	-0.533
TSS	-0.199	0.967	-0.123
Total Hardness	-0.650	0.377	0.637
Phosphate	0.319	0.922	0.143
Turbidity	0.349	0.933	-0.056
DO	0.467	0.087	0.841
Total Solid	-0.082	0.987	-0.089
TDS	0.927	0.247	0.276

Comment [H16]: Give standard units for every parameters you have examined
This is an academic paper

Nitrate	-0.198	0.906	0.365
BOD	0.159	0.147	0.869
Total Alkalinity	0.661	-0.088	0.403

3.3 Principal Component Regression (PCR)

The variables retained on PC1 (pH, Electric conductivity, Total Hardness, TDS, and Total alkalinity) which were regressed against the WQI as Equation (1) [9]

$$WQI = a_0 + a_1x_1 + a_2x_2 + \dots + a_5x_5 \quad (1).$$

Where: x_1 =pH, x_2 =Electric conductivity; x_3 =Total hardness, x_4 =TDS; x_5 =Total alkalinity

The input data for WQI and independent variables listed in Equation 1 were extracted from Tables 1 and 2 to calibrate Equation 1. The resulting Goodness of Fit after calibration is presented in table 4 and the corresponding values of the constants (Table 5) in Equation 1 are incorporated and given as Equation 1b:

$$WQI = 6100 - 955.91808 * pH + 10.09049 * \text{Electrical Cond.} - 1.19519 * \text{Total Hardness} - 10.57513 * \text{Total Alkalinity} \dots \dots \quad (\text{Equation 1b})$$

The Goodness of Fit associated with Equation 1b is 92.9% and the Root Mean Square Error (RMSE) is 66.673 which implies that an error of 7.1% is expected when Equation 1b is used to predict WQI for any of the 5 ponds.

Table 4: Goodness of Fit

Observations	10.000
Sum of weights	10.000
DF	5.000
R ²	0.929
Adjusted R ²	0.873
MSE	4445.228
RMSE	66.673
DW	2.379

Table 5: Model Parameters

Source	Value	Standard error	t	Pr> t	Lower bound (95%)	Upper bound (95%)
Intercept	6100.306	1174.307	5.195	0.003	3081.403	9119.209
pH	-955.918	178.520	-5.355	0.003	-1414.856	-496.980
Elect. Cond. Total	10.090	1.425	7.079	0.001	6.426	13.755
Hardness Total	-1.195	1.373	-0.870	0.424	-4.725	2.335
Alkalinity	-10.575	5.862	-1.804	0.131	-25.645	4.495

$$WQI = 6100 - 955.91808 * pH + 10.09049 * \text{Electrical Cond.} - 1.19519 * \text{Total Hardness} - 10.57513 * \text{Total Alkalinity}$$

PC2 which retained sulphate, TSS, phosphate, turbidity, total solids, and nitrate which were regressed against WQI as Equation 2.

$$WQI = a_0 + a_1x_1 + a_2x_2 + \dots + a_6x_6 \quad (2)$$

Where: x_1 =Sulphate; x_2 =TSS; x_3 =Phosphate, x_4 =Turbidity; x_5 =Total solid; x_6 : Nitrate

The procedure for model calibration for Equation 1 is repeated for Equation 2. The corresponding Goodness of Fit and model constants are presented in Tables 6 and 7. In effect, the calibrated version of Equation 2 is presented as Equation 2b:

$$WQI = 454.99394 + 0.10619 * \text{Sulphate} + 0.70969 * \text{TSS} + 4.77412 * \text{Phosphate} + 0.83703 * \text{Turbidity} - 0.69243 * \text{Total Solid} - 1.21455 * \text{Nitrate} \quad (\text{Equation 2b})$$

The goodness of fit associated with Equation 2b is 99.0% and the RMSE is 0.672 which implies that an error of 1% is expected when Equation 2b is used to predict WQI for any of the 5 ponds.

Model 2 Development

Table 6: Goodness of Fit

Observations	10.000
Sum of weights	10.000
DF	3.000
R ²	0.990
Adjusted R ²	0.990
MSE	0.452
RMSE	0.672
DW	2.383

Table 7: Model Parameters

Source	Value	Standard error	t	Pr> t	Lower bound (95%)	Upper bound (95%)
Intercept	454.99	12.483	36.450	< 0.0001	415.26	494.72
Sulphate	0.106	0.052	2.037	0.134	-0.060	0.272
TSS	0.710	0.095	7.438	0.005	0.406	1.013
Phosphate	4.774	0.169	28.312	< 0.0001	4.237	5.311
Turbidity	0.837	0.010	86.642	< 0.0001	0.806	0.868
Total Solid	-0.692	0.108	-6.424	0.008	-1.035	-0.349
Nitrate	-1.215	0.228	-5.320	0.013	-1.941	-0.488

PC3 which retained COD, iron, total hardness, DO and BOD which were regressed against WQI as Equation 3:

$$WQI = a_0 + a_1X_1 + a_2X_2 + \dots + a_5X_5 \quad (3).$$

Where: x_1 =COD; x_2 =Iron; x_3 =total hardness, x_4 =DO; x_5 =BOD

The procedure for model calibration for Equation 2 is again repeated for Equation 3. The corresponding Goodness of Fit and model constants are presented in Tables 8 and 9. In effect, the calibrated version of Equation 3 is presented as Equation 3b:

$$\text{WQI} = -34433 - 2.70898 \cdot \text{COD} + 12684 \cdot \text{Iron} - 3.60639 \cdot \text{Total Hardness} + 1669 \cdot \text{DO} - 689.53228 \cdot \text{BOD} \quad (\text{Equation 3b})$$

The Goodness of Fit associated with Equation 2b is 92.3% and an RMSE of 51.968 which implies that an error of 7.7% is expected when Equation 3b is used to predict WQI for any of the 5 ponds.

Model 3 Development

Table 8: Goodness of Fit

Observations	10.000
Sum of weights	10.000
DF	4.000
R ²	0.966
Adjusted R ²	0.923
MSE	2700.684
RMSE	51.968
DW	1.947

Table 9: Model Parameters

Source	Value	Standard error	t	Pr> t	Lower bound (95%)	Upper bound (95%)
Intercept	34433.16	5334.64	-6.455	0.003	49244.73	19621.58
COD	-2.709	0.756	-3.585	0.023	-4.807	-0.611
Iron	12683.83	1298.19	9.770	0.001	9079.417	16288.25

	5	4				2
Total Hardness	-3.606	2.055	-1.755	0.154	-9.311	2.098
DO	1668.518	478.587	3.486	0.025	339.729	2997.308
BOD	-689.532	442.202	-1.559	0.194	1917.301	538.236

$$\text{WQI} = -34433 - 2.70898 \cdot \text{COD} + 12684 \cdot \text{Iron} - 3.60639 \cdot \text{Total Hardness} + 1669 \cdot \text{DO} - 689.53228 \cdot \text{BOD}$$

Comparing the 3 fitted multiple regression models of Equations 1b, 2b and 3b, the best model is that of Equation 2b. This is possible given that the Goodness of Fit is the highest (99.0%) with the least RMSE value of 0.672. Therefore, this water quality model of Equation 2b will be handy in predicting WQI once the associated water quality parameters are known and the outcome will be valuable in water quality recommendation for domestic use for treatment.

This predictive model will find wide application in selecting water treatment options for pond water in the study area.

3.4 Model Verification

Customarily, model development involves 3 processes, namely: Model calibration, model verification and model prediction.

In this study, model calibration was actualized with WQI computed from Brown's formula and used as dependent variable while the parameters highlighted in bold as factor loading (table 3) under D1 D2 D3 as various independent variables and the final outcomes are represented as Equations 1B, 2B and 3B respectively. The model verification is carried out by using the various values of the applicable independent variables (see Table 1) to compute applicable values of WQI, the plot of which are as presented in Figure 3 below.

The distribution of the computed WQ values are closest to the fitted line for Fig 3b and mildly dispersed for Figs 3c and 3a respectively

and the plotted result confirms the Goodness of Fit R^2 of 99.0% of Fig 3b as the best model for the study area.

The verified model of Fig 3b can be used to actualize model prediction in the study area, that is predicting known values of WQI using obtained values of selected independent variables of Equation 2B.

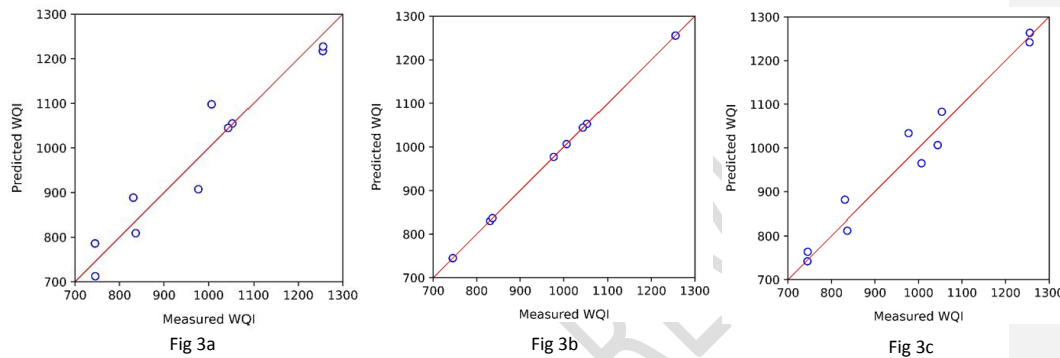


Fig 3: Model Prediction Comparison

4. Conclusion

Water quality index modelling yielded a 6 parameter multilinear regression linear model with Goodness of Fit of 99.0% and RMSE of 0.672.

This predictive model will find wide application in selecting water treatment options for pond water in the study area. Principal component analysis played a role of factor reduction that gives credence to 6 parameter models out of 25 independent variables.

Comment [H17]: List this out because this is a contribution to knowledge

References

1. Lobato, T.C., Hauser-Davis, R.A., Oliveira, T.F., Silveira, A.M., Silva, H.A.N., Tavares, M. R.M., Saraiva, A.C.F., 2015. Construction of a novel water quality index and quality indicator for reservoir water quality evaluation: A case study in the Amazon region. *J. Hydrol.* 522, 674–683. <https://doi.org/10.1016/j.jhydrol.2015.01.021>.

Comment [H18]: You need to arrange your references alphabetically and use more recent reference
Do you mean to say that there are no recent researches in this aspect of research?

2. Sanchez, ´ E., Colmenarejo, M.F., Vicente, J., Rubio, A., García, M.G., Travieso, L., Borja, R., 2007. Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution. *Ecol. Indic.* 7, 315–328. <https://doi.org/10.1016/j.ecolind.2006.02.005>.
3. Horton RK. An index number system for rating water quality. *Journal of the Water Pollution Control Federation.* 1965;37:300- 306.
4. Brown RM, McClelland NI, Deininger RA, Tozer RG (1970) A water quality index-do we dare. *Water Sew Work* 117:339–343
5. Abbasi, T., Abbasi, S.A., 2012. Water-Quality Indices. *Water Quality Indices.* Elsevier, pp. 353–356. <https://doi.org/10.1016/B978-0-444-54304-2.00016-6>.
Abrahao, ~ R., Carvalho, M., Da Silva, W.R., Machado, T.T.V., Gadelha, C.L.M.
6. Dadolahi-Sohrab, A., Arjomand, F., Fadaei-Nasab, M., 2012. Water quality index as a simple indicator of watersheds pollution in southwestern part of Iran. *Water Environ. J.* 26, 445–454. Brown RM, McClelland NJ, Deininger RA, O’Connor MF. A water quality index— crossing the psychological barrier. *Proceedings of the International Conference on Water Pollution Research, Jerusalem.* 1972;787-797. 13. Horton RK. An index number system for rating water quality. *Journal of the Water Pollution Control Federation.* 1965;37:300- 306.
7. Kannel, P.R., Lee, S., Lee, Y.-S., Kanel, S.R., Khan, S.P., 2007. Application of Water Quality Indices and Dissolved Oxygen as Indicators for River Water Classification and Urban Impact Assessment. *Environ. Monit. Assess.* 132, 93–110. <https://doi.org/10.1007/s10661-006-9505-1>.
8. Stoner, J.D., 1978. *Water-Quality Indices for Specific Water Uses.* Geol. Surv. Circ.
9. Nwaogazie, I. L. *Probability and Statistics for Science and Engineering Practice*, 1st Edition, Published by Prints Konzults, Lagos, pp.252, 1999; 2nd Edition, University of Port Harcourt

Press, pp293, 2006; 4th Edition, De-Adroit Innovation, Enugu, pp 367, 2021.

10. Mbachu, F. C., Nwaogazie , I. L., & Amah , V. (2023). Water Quality Assessment on Physicochemical and Biological Parameters of Selected Lentic Ecosystem in AbohMbaise Local Government Area, Imo State, Nigeria. *International Journal of Environment and ClimateChange*, 13(11), 110–122. <https://doi.org/10.9734/ijecc/2023/v13i113150>.
11. AwosanyaBabatunde, Ify L. Nwaogazie and Chiedozie F. Ikebude: Water Quality Assessment of Nigerian Port Authority Waterway in Port Harcourt, Rivers State Nigeria. *Journal of Engineering Research and Report JERR*. Ava2023;12(1):110-120.. DOI: 10.9734/JERR/2023/v25i1875.
12. Brown RM, McCleiland NJ, Deininger RA, O'Connor MF. A water quality index— crossing the psychological barrier. *Proceedings of the International Conference on Water Pollution Research, Jerusalem*. 1972;787-797.
13. Odia Mamuyovwi, Nwaogazielfy L., Nwachukwu Eunice O. and Awiri Gregory O."Modeling Leachate Pollution Index and Potential for Selected Municipal Solid Waste Dump Sites: A Case Study. *British Journal of Applied Science and Technology*. 2016. Vol. 18, Issue 6, pp. 1-16. Article no. BJASt.31198. ISSN: 2231-0843, NLM ID: 101664541.