Journal of Engineering Research and Reports

Volume 25, Issue 11, Page 13-21, 2023; Article no.JERR.108942
ISSN: 2582-2926

Modelling and Prediction of Water Quality Index of Selected Pond Water in Aboh Mbaise Local Government Area, Imo State, Nigeria

Frank C. Mbachu [a] and Ify L. Nwaogazie [b*]

[a] *Centre for Occupational Health, Safety and Environment (COHSE), University of Port-Harcourt, Choba, Rivers State, Nigeria.*
[b] *Department of Civil and Environmental Engineering, University of Port-Harcourt, Choba, Rivers State, Nigeria.*

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JERR/2023/v25i111016

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/108942

Original Research Article

Received: 09/09/2023
Accepted: 14/11/2023
Published: 18/11/2023

ABSTRACT

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

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 92.9, 99.0 and 96.6% as well as root mean square error (RMSE) of 66.673, 0.672 and 51.968 respectively.

**Corresponding author: Email: Ifynwaogazie@yahoo.com;*

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 with the following independent variables-sulphate, TSS, phosphate, turbidity, total solid and nitrates.
The model output is relevant in WQI prediction given the applicable water quality characteristics.
This predictive model will find wide application in selecting water treatment options for pond water in the study area.

Keywords: Modelling, prediction; water quality index; principal component analysis; multiple regression; aboh-mbaise.

1. 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 (physicochemical, 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].

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 Enyiogugu, 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 (Mbachu et al) 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, [11] and implemented in Microsoft Excel. The detailed formula is as presented in an earlier publication by the authors (Mbachu et al) [10] and Babatunde et al. [12].

Similar application of water quality index in surface and ground water are available in literature [13,14,15].

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.

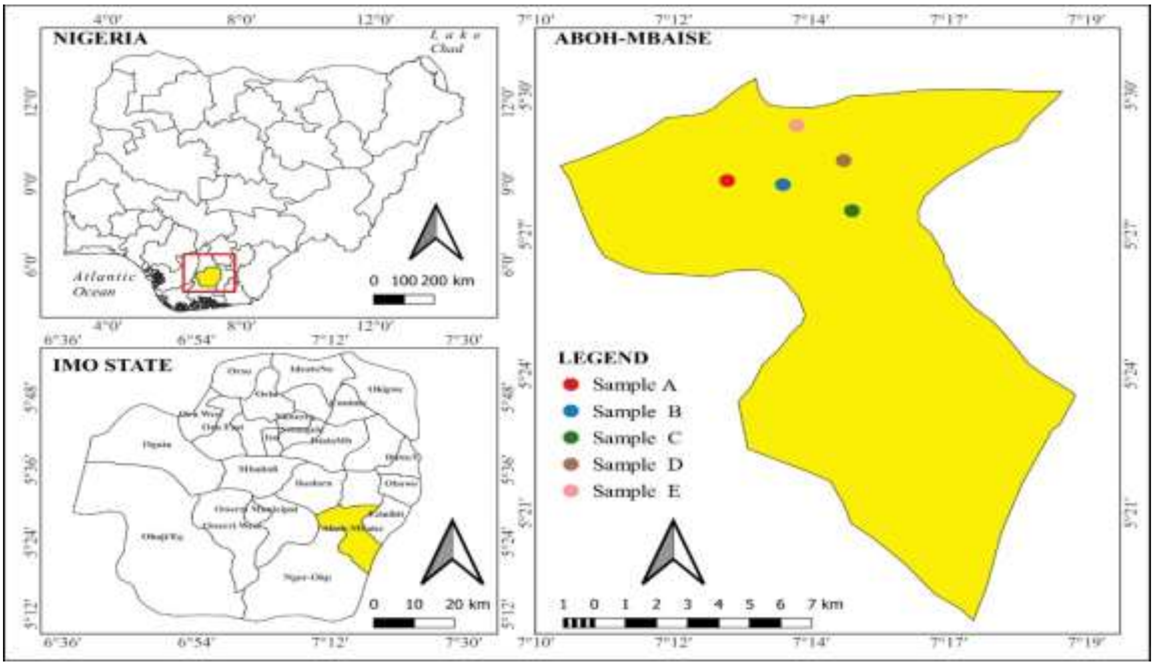

Fig. 1. The study area

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.

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 physicochemical parameters of the pond water. [8, 11] The PCA analysis revealed distinct relationships between various physicochemical parameters, providing valuable insights into the complex dynamics of water quality (Fig. 2).

The three principal components were retained after varimax rotation. The factor loading of the physicochemical 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 (-0.650), 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.

Table 1. Descriptive statistic of physiochemical parameters

Pysiochemical Parameters	Statistic	Amakwu Umuelemem	Ibeku Okwuato	Umuabaza Okwuato	Umuanuma Nguru	Urakwo Enyiogugu
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
	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
Iron	mean	1.80	1.87	1.89	1.80	1.84
	std	0.00	0.00	0.00	0.00	0.00
Copper	mean	0.00	0.02	0.01	0.01	0.02
	std	0.00	0.00	0.00	0.00	0.01

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

Locations	WQI
Urakwo Enyiogugu	830.6944
Urakwo Enyiogugu	836.1406
Umuabaza Okwuato	1255.045
Umuabaza Okwuato	1255.663
Ibeku Okwuato	744.5951
Ibeku Okwuato	745.1515
Amakwu Umuelemem	1053.428
Amakwu Umuelemem	1044.023
Umuanuma Nguru	1006.742
Umuanuma Nguru	976.9529

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]

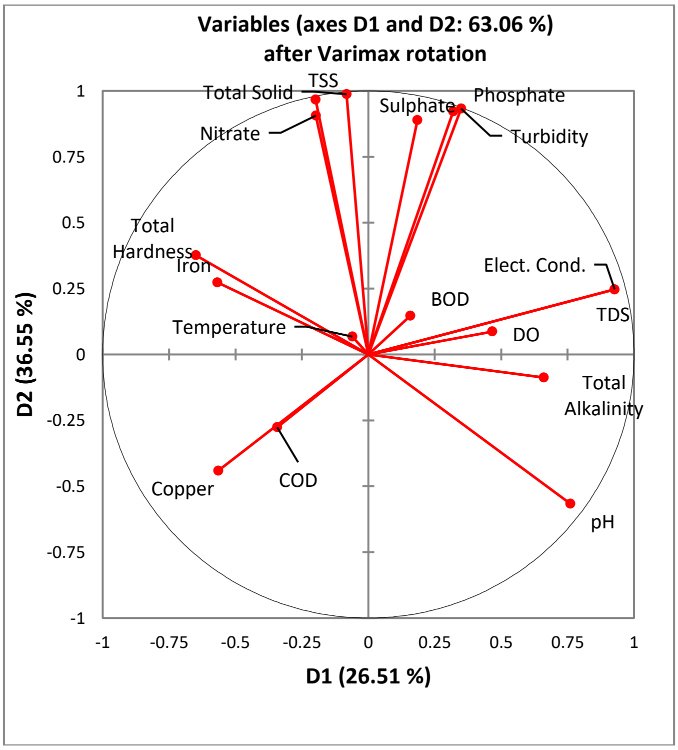

Fig. 2. Biplot showing the physicochemical parameters on principal component 1 and 2

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

PC3 was characterized by loadings from variables such as Chemical Oxygen Demand (COD) (-0.745), Iron (-0.768), Dissolved Oxygen (DO) (0.841), Total hardness (0.637) 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.

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].

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

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*Electrical Cond. -1.19519*Total Hardness -10.57513*Total Alkalinity…… (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

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

$$WQ1=a_0 +a_1x_1+a_2x_2+\dots+a_6x_6 \quad (2).$$

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*Sulphate + 0.70969*TSS + 4.77412*Phosphate + 0.83703*Turbidity - 0.69243*Total Solids - 1.21455*Nitrate (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. The 2 basic parameters for model acceptance are Goodness of Fit and RMSE and they are equally applicable in mathematical model development and have been adopted in model development in Geotechnical engineering [16].

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.	10.090	1.425	7.079	0.001	6.426	13.755
Total Hardness	-1.195	1.373	-0.870	0.424	-4.725	2.335
Total Alkalinity	-10.575	5.862	-1.804	0.131	-25.645	4.495

*WQI = 6100 - 955.91808*pH + 10.09049*Electrical Cond. -1.19519*Total Hardness -10.57513*Total Alkalinity*

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	454.994	12.483	36.450	< 0.0001	415.268	494.720
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:

$WQI=a_0 +a_1x_1+a_2x_2+\ldots+a_5x_5$ (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:

WQI = -34433 - 2.70898*COD + 12684*Iron - 3.60639*Total Hardness + 1669*DO - 689.53228*BOD (3b)

The Goodness of Fit associated with Equation 3b is 96.6% and an RMSE of 51.968 which implies that an error of 3.4% 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

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 principal components of D1, D2 and D3 is a case of factor reduction via principal component analysis which are exemplified in literature [17,18].

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 Fig. 3.

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 unknown values of WQI using obtained values of selected independent variables of Equation 2b.

Table 9. Model parameters

Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)
Intercept	-34433.161	5334.648	-6.455	0.003	-49244.737	-19621.585
COD	-2.709	0.756	-3.585	0.023	-4.807	-0.611
Iron	12683.835	1298.194	9.770	0.001	9079.417	16288.252
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

*WQI = -34433 - 2.70898*COD + 12684*Iron - 3.60639*Total Hardness + 1669*DO - 689.53228*BOD*

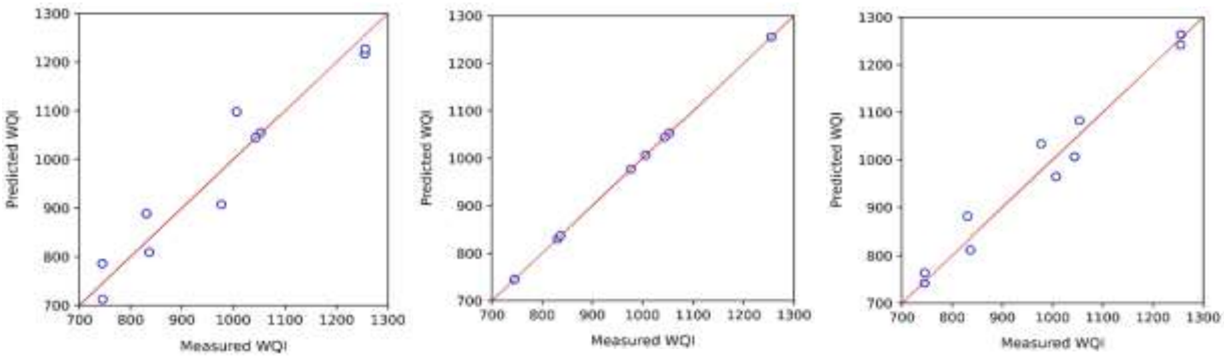

Fig 3a Fig 3b Fig 3c

Fig. 3. Model prediction comparison

4. CONCLUSION

Water quality index modelling yielded a 6 parameter multilinear regression 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, that is sulphate, TSS, phosphate, turbidity, total solid and nitrates out of 25 independent variables.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Lobato TC, Hauser-Davis RA, Oliveira TF, Silveira AM, Silva HAN, Tavares MRM, Saraiva ACF. Construction of a novel water quality index and quality indicator for reservoir water quality evaluation: A case study in the Amazon region. J. Hydrol. 2015;522:674–683. Available:https://doi.org/10.1016/j.jhydrol.2015.01.021.
2. Sanchez E, Colmenarejo MF, Vicente J, Rubio A, García MG, Travieso L, Borja R, Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution. Ecol. Indic. 2007;7:315–328. Available: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. A water quality index-do we dare. Water Sew Work. 1970;117:339–343.
5. Abbasi T, Abbasi SA. Water-quality indices. Water Quality Indices. Elsevier. 2012;353–356. Available:https://doi.org/10.1016/B978-0-444-54304-2.00016-6. Abrahao, ˜ R., Carvalho M, Da Silva WR, Machado TTV, Gadelha CLM.
6. Dadolahi-Sohrab A, Arjomand F, Fadaei-Nasab M. Water quality index as a simple indicator of watersheds pollution in

southwestern part of Iran. Water Environ. J. 2012;26:445–454.
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.
Horton RK. An index number system for rating water quality. Journal of the Water Pollution Control Federation. 1965;37:300-306.

7. Kannel PR, Lee S, Lee YS, Kanel SR, Khan SP. Application of water quality indices and dissolved oxygen as indicators for river water classification and urban impact assessment. Environ. Monit. Assess. 2007;132:93–110. Available:https://doi. org/10.1007/s10661-006-9505-1.
8. Stoner JD. Water-quality indices for specific water uses. Geol. Surv. Circ; 1978.
9. Nwaogazie IL. 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. 2021;367.
10. Mbachu FC, Nwaogazie IL, Amah V. Water Quality Assessment on Physicochemical and Biological Parameters of Selected Lentic Ecosystem in Aboh Mbaise Local Government Area, Imo State, Nigeria. International Journal of Environment and Climate Change. 2023;13(11):110–122. Available:https://doi.org/10.9734/ijecc/2023/v13i113150.
11. 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.
12. Awosanya Babatunde, 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. Ava. 2023;12(1):110-120. DOI: 10.9734/JERR/2023/v25i1875.
13. Al-Kareem SA, ALKzwini RS. Statistical analysis for water quality index for Shatt-Al-Hilla river in Babel city. Water Pract Technol. 2022;17:567–586. DOI: 10.2166/wpt.2022.004
14. Akhtar N, Ishak MIS, Ahmad MI, Umar K, Md Yusuff MS, Anees MT, Qadir A, Ali Almanasir YK. Modification of the Water Quality Index (WQI) process for simple calculation using the Multi-Criteria Decision-Making (MCDM) Method: a review. Water. 2021;13:905. DOI: 10.3390/w13070905.
15. Banda T, Kumarasamy M. Development of a universal water quality index (UWQI) for South African River Catchments. Water. 2020;12:1534. DOI: 10.3390/w12061534
16. Irokwe J, Nwaogazie I, Sule S. Development of empirical models for the estimation of CBR value of soil from their index properties: A case study of the ogbia-nembe road in Niger Delta Region of Nigeria. Open Journal of Civil Engineering. 2022;12:648-664. DOI: 10.4236/ojce.2022.124036
17. Hasan BM, Adnan Mohsin Abdulazeez. A review of principal component analysis algorithm for dimensionality reduction. Journal of Soft Computing and Data Mining. 2021;2(1):20-30. DOI:https://doi.org/10.30880/jscdm.2021.02.01.003
18. Odia Mamuyovwi, Nwaogazie Ify L, Nwachukwu Eunice O, Avwiri 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;18(6):1-16. Article no. BJAST.31198. ISSN: 2231-0843, NLM ID: 101664541

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/108942