

Research Article

Data-Driven Analysis Based on Graphical and Statistical Modelling of the Water Quality of Niger Delta Region of Nigeria

Davidson E. Egirani^{1*}, Mohd T. Latif², Ifeoma M. Ugwu³, Alfred W. Opukumo¹

¹Faculty of Science, Niger Delta University, Wilberforce Island, Nigeria

²School of Environmental and Natural Resource Sciences, The National University of Malaysia, Bangi, Malaysia

³Faculty of Science, Enugu State University of Science and Technology, Enugu, Nigeria

Email: eenonidavidson@yahoo.com

Received: 12 June 2020; **Revised:** 30 October 2020; **Accepted:** 20 November 2020

Abstract: Data-driven models derived from data science tools have been used to investigate water quality in some parts of the Niger Delta Region (NDR) of Nigeria. Eleven communities were affected in this study. Eleven water samples obtained from twenty-five available sources, collected from January 2019 to December 2019, included rainwater, surface water, and groundwater. These samples were analysed for their physicochemical and bacteriological parameters. The physical characteristics of the water points ranged from a pH of 6.61-7.2, Electrical Conductivity (EC) of 450-1,742 units, turbidity of 0.72-13.65 units, and Total Dissolved Solids (TDS) of 225-794. The chemical dataset generated was subjected to several scientific data models such as Principal Component Analysis (PCA), Piper, Pie, Collins, and Schoeller interpretations. There is evidence that the water resources are potable in sections where *Escherichia coli* and total coliforms do not exceed the international and regional recommended limits of 0 per 100 ml of sample. In addition, the community water points are good for livestock and excellent for both recreation and irrigation purposes. Possible water contamination sources include faecal pollution from shallow wells and unconfined aquifers. Land use planning, as well as the enactment and implementation of environmental laws, is necessary in this region to ensure effective surface water and groundwater resource management.

Keywords: data-driven analysis, data science, contamination, modelling, Niger Delta, water quality

1. Introduction

The quality of water is a function of the physical, chemical and bacteriological characterization of available water resources. These parameters are better expressed using data-driven tools such as modelling, data pipelines, and visualization. These are all components of data science used in modern water quality assessment and prediction. Lack of access to potable water is not linked to the quantity of availability all the time but accessing quality water remains an area of research interest worldwide [1]. Effective legislation on proper land use and planning are required to achieve human health and ecosystem devoid of risk from contaminated water [2]. Also, global food production is enhanced on this platform. The main public concern in recent years has been the damage to human health that is caused by contaminated water through environmental pollution [3].

Several countries are yet to deal with water contamination arising from reckless use and disposal of agricultural

Copyright ©2020 Davidson E. Egirani, et al.

DOI: <https://doi.org/10.37256/ccds.212021478>

This is an open-access article distributed under a CC BY license
(Creative Commons Attribution 4.0 International License)

<https://creativecommons.org/licenses/by/4.0/>

waste [4]. Therefore, governance that promotes clean water is critical in water resource management. The biggest recipients of contaminated water runoff remain lakes, rivers, and infiltration into groundwater [5]. The Water Quality Index (WQI) and other theoretical models are used to monitor water quality changes in water supply over time. Assessments of water quality are very important for knowing its suitability for various purposes. Water quality studies have been widely considered in the literature [6].

Several studies on water quality in this region have been presented. However, these studies were not based on statistical and graphical models. These analyses were based on simple descriptive and multivariate statistical modelling [7-8]. Despite the region's abundance of water, water quality is threatened by deteriorating environmental conditions and water management that are not being sufficiently addressed [9]. These contaminants arise from anthropogenic sources [9]. Water wells close to effluent discharge is a potential source of contamination. Drinking water supply in many of these communities could be unhealthy and it does not meet the regulatory standards [7]. Bacterial presence in groundwater makes it unfit and it is a major contributor to water-borne diseases such as diarrhoea, nausea, gastroenteritis, typhoid fever and dysentery. Total Bacterial Count (TBC), total coliforms and *Escherichia coli* are common indicators of water contamination with disease-causing pathogens [8]. Sources of total and faecal coliform in groundwater can include infiltration of domestic or wild animal faecal matter, effluent from leaking septic systems or sewage discharges and vegetal matter runoff [8, 10-11]. According to the World Health Organization's (WHO) standard and Canadian drinking water quality, total coliforms and faecal coliforms in 100 mL of water must both be below detectable levels. The guideline for total bacteria count is less than 100 CFU/mL⁻¹ [10].

It is well recognized that the failure to protect water sources and inadequate water treatment are the primary reasons for drinking water contamination with bacteria in this region. The present study centers on using statistical and graphical models such as Principal Component Analysis (PCA), piper, Pie, Collins, and Schoeller and PHREEQC to describe the water quality in the NDR of Nigeria. Herein, surface water and groundwater quality assessment in this region using data-driven tools are presented.

2. Water resource modelling in NDR

Several water quality models exist. These include model-driven techniques such as spectrophotometry [12], electrochemical analyses [13] or chromatography [14]. Ngouna et al [15] used a data-driven method in detecting and diagnosing causes of water quality contamination from trazine and nitrate. These models have not been applied in this region since the sources of water contamination are diffused and not from a point-source. The NDR consists of organic-peaty marine mud in the tidal flat, saline mangrove swamp, the clay of the freshwater back swamps and the deltaic lateritic soils of the dry plains. The Benin Formation consists of coastal plain sands. Majority of aquifers located within the region are unconfined [16-17].

The clay thickness in the study area is over 6 m in most locations. This thick clay does not constitute a good aquitard due to the waterlogged nature of the study area almost all year round. Also, the degree of surface contamination of groundwater increases towards the southern fringe of this region [18].

An analysis of the physical, bacteriological and chemical composition of the surface water and groundwater in this region provide information on the spatial pollution and provides a strategic framework for water resource management. The major physicochemical and bacteriological properties of interest include pH, temperature, electrical conductivity, total bacterial count, *E.coli*, *T. coli*, turbidity, salinity, colour, odour, anions and cations [19]. The study aims to provide information on the water resource quality in parts of the NDR using scientific data models. The objective of this study includes the physicochemical and bacterial analyses of the surface water and groundwater. Herein, the results have provided a management strategy to reduce the ingress of contaminants into the water resources. Since there are variable sources of potential water contamination, this study used statistical and graphical tools to address the problem of water quality in the NDR of Nigeria.

2.1 Location and geology of the study area

The selected area of study is coordinated as Eastings 600-6300E and Northings 500-5150N located in the NDR of Nigeria (Figure 1-3). The topography is low-lying but steadily rising to about 40 m inland. Thus, flooding and

subsequent infiltration of water contaminants into water sources can be attributed to anthropogenic and natural causes. Sources of flooding may be due to the high level of rainfall and release of water from Niger and Benue Rivers [20]. Agricultural activities and exploration from oil companies have contributed to the gradual shift from forested vegetation to open grassland thus affecting the distribution of rainfall in the study area.

The vegetation remains a freshwater swamp with several rivers and creeks sustaining the eco-belt. These rivers include the River Niger as a major contributor to the Niger Delta, Escravos, Nuns, Orashi, Ramos, and the Kolo creek [21].

Units of sandstone and shale intercalating the Agbada Formation and the Akata Formation have emerged. These three units cut across the whole NDR and each ranges in age from early tertiary to recent. The Benin Formation consists predominantly of freshwater continental sands and gravel with intercalations of shale. This Formation constitutes an unconfined unit that is most prolific in the water supply to the region. The Benin Formation is the most prolific aquifer with a hydraulic conductivity of $6.16-108 \text{ m}^3$ [22].

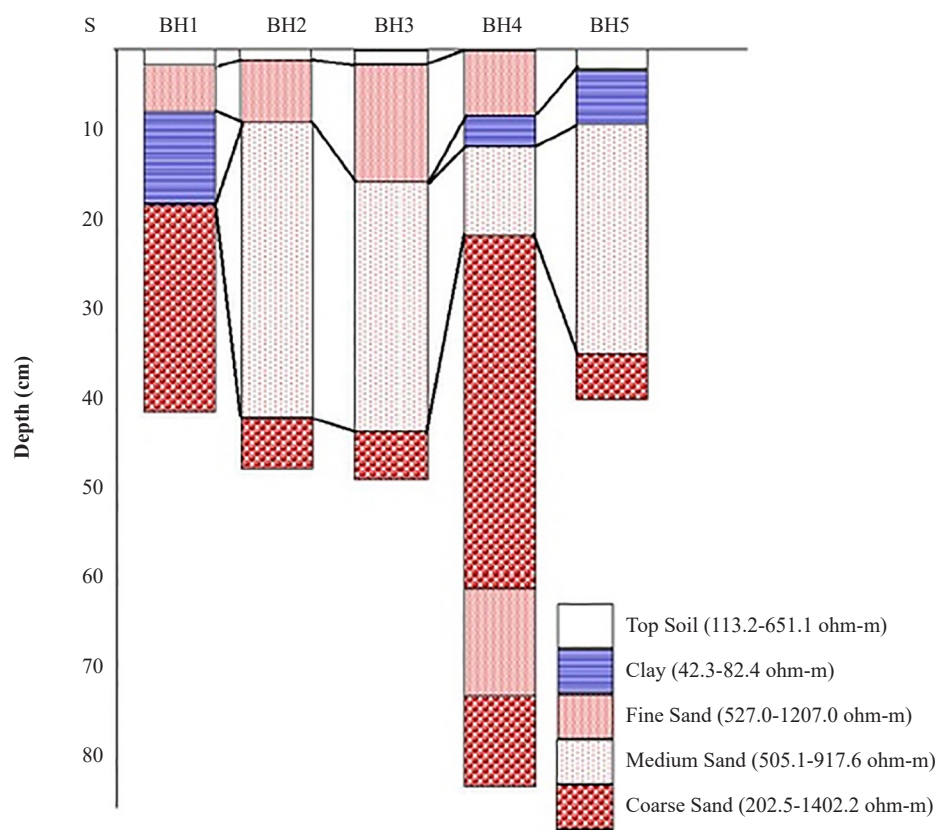


Figure 1. Borehole lithology in the NDR of Nigeria [22]

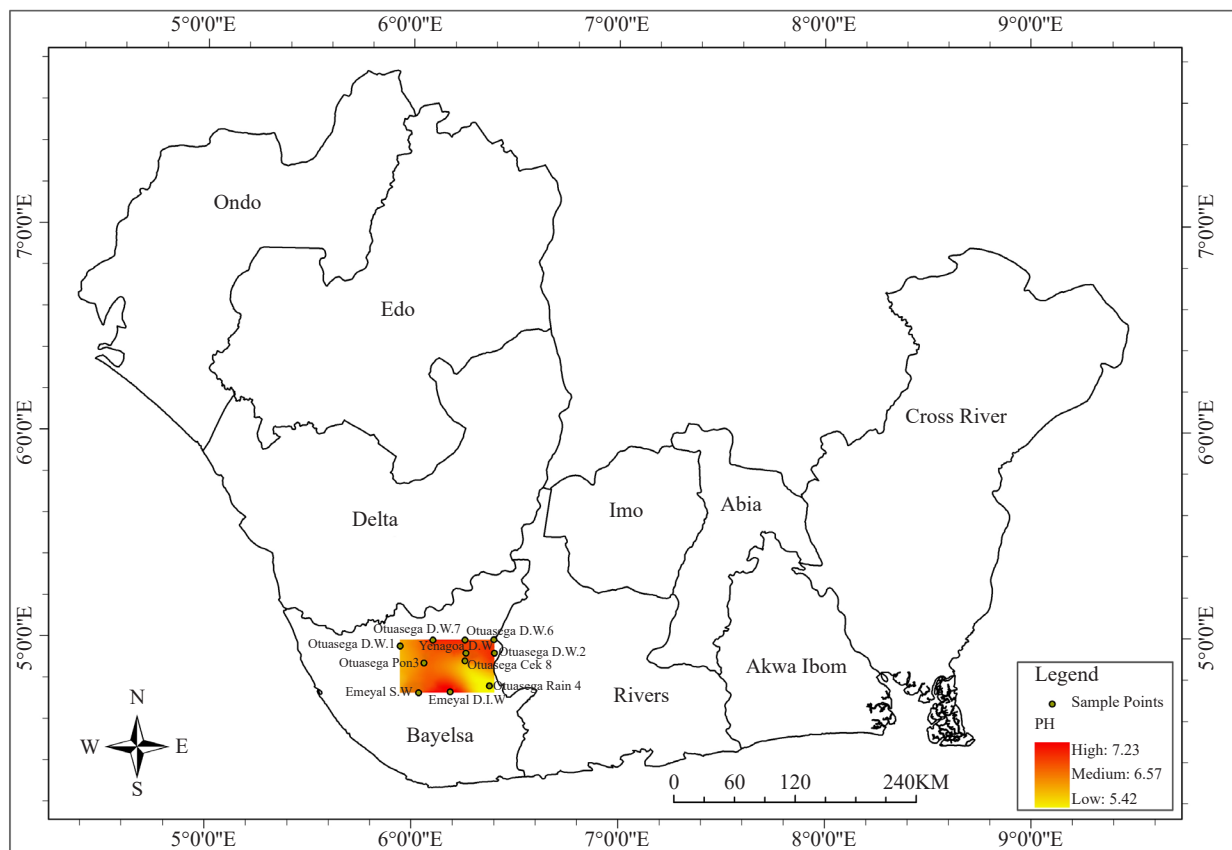


Figure 2. Map of the NDR showing the study location and pH variation differentiated [After modification by the author]

2.2 Data-driven model concepts of water quality

Several scientific data tools have emerged to understand the quality of water resources. Data obtained from the chemical analysis provides the framework for further scientific data modelling. A considerable number of techniques and methods based on differences and chemical properties of water are available to classify, compare and summarize large volumes of data. Common procedures include simple inspections and comparison of chemical analyses, the preparation of graphs and maps and more extensive statistical analysis [23].

2.3 Principal component analysis (PCA) scientific data model

PCA is a scientific data model employed to highlight the sources of physicochemical properties. Water quality datasets were subjected to principal component analysis (PCA) [19]. Statistical computations were executed using the statistical software package, IBM SPSS Statistics 20. PCA seeks to transform original data and set of variables into a new group of uncorrelated axes (the principal components). These components are linear groups of the original data [19]. PCA is thus concerned only with establishing which linear component exists within the data and how a particular variable might contribute to that component [24].

2.4 Piper scientific data model

This model as used in Figure 4 presents the abundance or relative abundance of component ions in surface water and groundwater. For groundwater, it reveals the groundwater facies. This model provides a combination of several plots used in water classification and identification of water mixing [25].

2.5 Pie scientific data model

This model presents concentration ratios for individual samples. This model is useful in the easy comparison of concentrations of different samples. Circular but simple diagrams are mostly used. The subdivisions of the area represent proportions of different ions in the percentage of meq/L [26].

2.6 Collins scientific data model

This model is designed to display concentrations (not ratios) of individual water samples. It is a cumulative chart and the values are not readily apparent [26].

2.7 Schoeller scientific data model

This model is designed to display the logarithm of major ion in meq/l. It demonstrates different water types in the same diagram. Sample concentrations (not ratios) are displayed and compared. Similar waters exhibit similar “fingerprints” [27].

2.8 PHREEQC scientific data model

The geochemical program (PHREEQC Interactive version 3.0.6.7757 from the United States Geological Survey) is a tool used in the geochemical calculations of aqueous species [28]. Also, PHREEQC provides information on the degree of precipitation of solutions containing different ions based on the saturation index (SI) (Table 1). The SI defines the ratio of ion activity product (IAP) in log form. Negative and positive SI values represent the potential for dissolution and precipitation, respectively [23].

Table 1. Parameters used in calculating saturation index from PHREEQC

Ion	Conc. (ppm)	Ion	Conc. (ppm)
Ca	variable	Cl	Variable
Mg	variable	NO ₃	Variable
Na	variable	Fe	Variable
K	variable	pH	Variable
HCO ₃	variable	T	25°C
SO ₄	variable	Pe	4
Density	1	Cl	Variable
Alkalinity	variable	Mn	Vatiable

2.9 Sodium absorption ratio scientific data model

The sodium adsorption ratio (SAR) model is a measure of the relative amount of sodium to calcium and magnesium in a water sample which indicates the relative proportion of sodium ions in a water sample to those of calcium and magnesium. In calculating the SAR from water analysis data, it is essential to convert the units from parts per million or milligrams per Litre to milliequivalents per Litre (Equation 1):

$$SAR = \frac{Na}{\sqrt{\frac{Ca+Mg}{2}}} \quad (1)$$

This model provides information on the ratio of sodium to calcium and magnesium-based on the dominance of sodium in a soil-water system. A low SAR means that the water when used could cause structural engineering failure (Table 2) [29].

Table 2. Hazard levels for SAR [29]

SAR	Hazard
< 10	Safe to irrigate with limited structural decay but salt-sensitive plants may be affected
10-18	Hazard on fine-textured soils with a high cation exchange capacity. Suitable on coarse-textured soil
18-26	Hazard on most soils. Need to manage with amendments and drainage (i.e. leaching)
26	Not suitable for irrigation

2.9.1 Saturation index (SI) scientific data model

The saturation index (SI) links pH, salinity, alkalinity, and hardness (Table 3) and provides information on the scaling and corrosion properties of the water. Also, the Langelier Saturation Index (LSI) calculator links the scaling properties of the water [30].

Table 3. Interpreting the Langelier Saturation Index (LSI) [25]

SI Range	Likelihood of scaling	Likelihood of corrosion
-0.5 to 0.5	Not likely	Not likely
0.5 to 1.5	Moderate risk	Not likely
> 1.5	Strong risk	Not likely
-0.5 to -1.5	Not likely	Moderate risk
< -1.5	Not likely	Strong risk

2.9.2 Residual sodium carbonate (RSC) scientific data model

Residual sodium carbonate model provides information on the sodium carbonate and sodium bicarbonate content. Residual alkalinity is a milli-equivalent per Litre (mEq/L) expression of sodium carbonate water on clay soils with exchangeable calcium and magnesium content. Sodium when charged with high residual sodium water lead to calcium replacing magnesium thus causing structural failure of engineering works (Equation 2) [31]:

$$RSC = (CO_3 + HCO_3) - (Ca + Mg) \quad (2)$$

When RSC is negative then the water passes the engineering works failed the test. An RSC greater than 1.25 indicates a potential hazard to soil structure. Additions of a calcium source such as gypsum or acidification of the water before use may be required [32].

2.9.3 Canadian water quality index (CWQI) scientific data model

WQI model provides a non-technical answer about the quality of water resources. It is unitless and has a number ranging from 1 to 100. The higher the number, the better the water quality [33]. The World Health Organization standard guidelines are presented (Table 4).

Table 4. World Health Organization standards for drinking water [28]

Constituent	Upper limit (ppm)
Appearance	Colourless 15 units
Odour	Odourless
pH	6.5-8.5
Turbidity	5 FTU
Alkalinity	200 mg/L
Oxygen	4-12 mg/L
Total Hardness	50-200 mg/L
Ammonium as NH ₄	0.04 mg/L
Nitrite	0.005 mg/L
Nitrate	50 mg/L
Phosphate	0.1 mg/L
Chloride	200 mg/L
Iron	0.1 mg/L
Manganese	0.05 mg/L
Sulphate	200 mg/L
Calcium	200 mg/L

2.9.4 Physicochemical concept of surface water and groundwater scientific tools

PH is used to describe the acidity and alkalinity of a solution. The World Health Organization Guide level [34] for pH is provided in Table 4. The major sources of TDS are all solid materials in solution whether ionic or not. TDS excludes suspended sediment or dissolved gas [35]. Turbidity provides information on the optical ability of inorganic materials. Most industries require water with a turbidity of 10 or less, particularly food and brewage manufacturers. Colour is caused largely by organic materials in solution.

These include metallic substances in industrial wastewater and organic decay products in swampy waters. A limit of 15 units has been set for portable water supply. Food and beverage industries prefer water with the colour of fewer than 10 units. Odour in water indicates contamination and can arise from the following sources [35]:

- presence of hydrogen sulphide in water due to septic sewage or chemical and biological reduction of sulphides, especially in deep groundwater;
- contamination by chemicals from industrial source
- growth of algae, protozoa and fungi in the water.

Iron is one of the most objectionable species in water supplies. It is removed from water supplies for aesthetic and economic effect. About 0.3 mg/L is the portable limit but most industries prefer 0.1 to 0.2 mg/L as an upper limit.

All natural waters contain some sodium concentrations ranging from less than 0.5 mg/L in rainwater to more than 1000 mg/L in brine.

The concentration of calcium in natural waters commonly range between 10 mg/L and 100 mg/L limits. Calcium in water would not be desirable for domestic supplies because of the disadvantage of washing, bathing and laundry [35].

Magnesium is usually found in lower concentration than calcium due to the greater hardness and like calcium, form scales and deposits on heating. The concentration of magnesium above 150 mg/L, especially if presented with sulphate can cause gastrointestinal irritation and diarrhoea. High values of sulphate above 200 mg/L can lead to the attack of diarrhoea, especially in infants. The World Health Organization Guide Level is 400 mg/L [28]. High nitrate concentration in drinking water has been implicated in deaths of infants by cyanosis (methaemoglobin anaemia) and possible cancer forming agents [35].

The major sources of chloride in water are from sedimentary evaporite. It may also be found in igneous rocks. The limits of chloride have been set at 250 mg/L for reasons of portability. The taste of chloride in water is probably influenced by the concentration of calcium and magnesium. There is no physiological danger for chloride in this range. Its high concentration is not a problem in the industrial water supply. The irrigation water criteria are dependent on the types of plant, the amount of irrigation water use, soil and climate [35]. The spatial distribution map of the study area is presented (Figure 3).

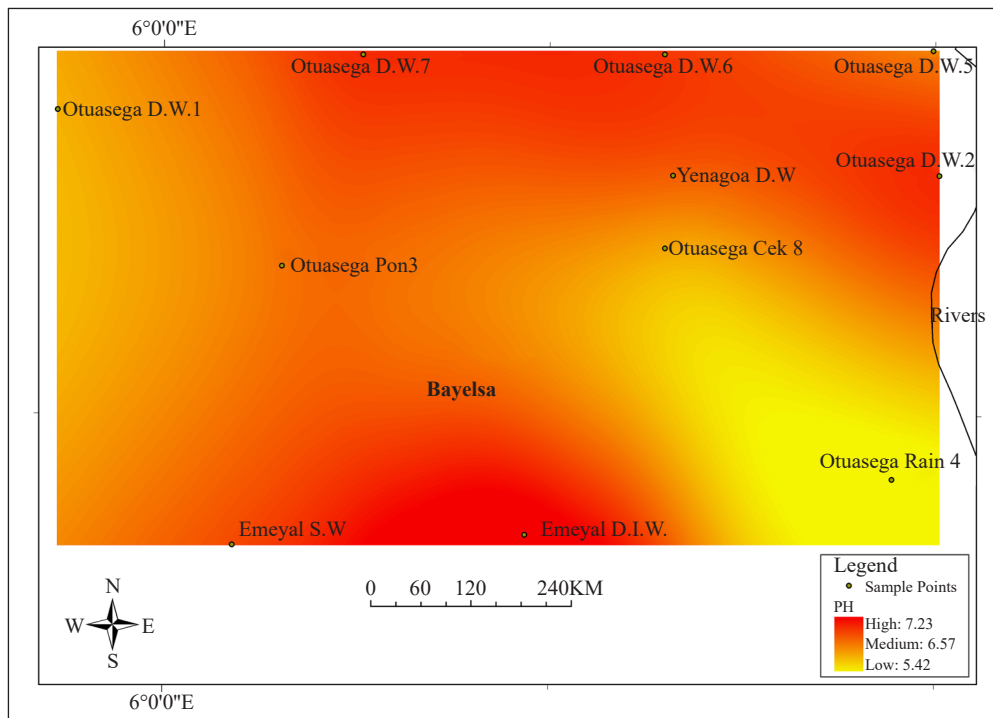


Figure 3. Spatial distribution of sampled communities and pH variation

3. Materials and analytical methods

3.1 Physicochemical analysis

The American Society for Testing and Materials (ASTM) was used in this study (Table 5-6). Only analytical grade reagent and chemicals were used in preparing reagents and standards. Collected surface water and groundwater samples were analyzed for pH, temperature, TDS, per standard methods. Insitu measurements were carried out for the rain, surface water and groundwater collected. Unstable field parameters namely pH, conductivity, TDS, and temperature were analyzed in the field and recorded. TDS was determined by using the HACH TDS meter. pH and conductivity measurements were performed with a mercury-in-glass thermometer, portable Orion Model 290 pH meter and Oakton Model 35607 conductivity meter respectively. For colour determination, Lovibond colour disc was used. Determination of total hardness was by Titrimetric Method. Chloride and iron were determined using a portable data logging spectrophotometer (Atomic Absorption Spectrophotometer HACH DR/2010). The major cations were analyzed using the atomic absorption spectroscopy (AAS). Salinity was determined using Mohr's method. Turbidity was determined using the HACH turbidity. Nitrate was determined by Hach DR. 4000 spectrophotometer using cadmium reduction method. The sulphate content of the sample was determined by turbidimetric method and bicarbonate was determined by the titrimetric method [36].

3.2 Bacteriological analysis

The total bacterial count, total coliforms and *Escherichia coli* were determined quantitatively. The standard plate counting (SPC) method was used to enumerate the total bacterial count [37]. Total coliforms were detected and quantified with the use of Eosin methylene blue (EMB) agar and their incubation at 37°C. Their counts were expressed in CFU/100 mL of the water [36].

Table 5. Physical characterization of water resources in the study area

Community Serial number	Community water point Location Name	pH	EC (µs/cm)	Salinity (µs/cm)	Turbidity FTU	TDS ppm
1	Emeyal Drilled Well	7.2	1545	1.08	4.25	772
2	Emeyal Surface Well	6.98	1241	0.65	6.42	620
3	Otuasega Drilled Well 1	7.04	1273	2.1	3.40	639
4	Otuasega Drilled Well 2	6.87	1344	0.66	4.60	672
5	Otuasega Drilled Well 3	7.02	450	0.24	6.50	225
6	Otuasega Drilled Well 4	7.04	1278	2.5	5.60	649
7	Otuasega Creek Water	6.8	1742	0.85	13.65	320
8	Otuasega Pond Water	6.9	1587	1.47	12.27	794
9	Otuasega Rainwater	6.61	1416	0.27	0.72	708
10	Otuasega Drilled Well 5	6.78	632	0.33	0.80	316
11	Yenagoa Drilled Well	6.94	815	0.5	1.84	407

Table 6. Descriptive statistics of water resources in the study area

	N	Minimum	Maximum	Mean	Std. Deviation
	Statistic	Statistic	Statistic	Statistic	Statistic
pH	11	6.61	7.20	6.8991	0.04898
EC	11	450.00	1742.00	1141.3636	122.99817
Salinity	11	0.24	13.65	2.1314	1.17427
Turbidity	11	0.72	13.65	5.4591	1.27639
TDS	11	225.00	794.00	556.5455	60.64417
Valid N (listwise)	11				

4. Results and discussion

4.1 Physico-chemical characteristics

Except for six water points (i.e. 2, 6, 7, 8, 10 and 11), the rest of the community water points were clear and odourless. The pH ranged from 6.61 community water point 9 to a maximum of 7.2 in community water point 1 (Table 7). Therefore, it is within the acceptable limits for drinking water. The colour of the groundwater ranged from 4 to a maximum of 12. Also, this is within the acceptable limits for portable, domestic, and industrial water. The calcium concentration ranged from 3.2 mg/L for community water point 9 to 6.0 mg/L for community water point 11. This concentration range is within the recommended guide level for water meant for drinking, irrigation, domestic,

and industrial uses. The sodium concentration ranged from 3.6 mg/L for community water point 5 to 8.4 mg/L for community water point 1. The concentrations of iron ranged from 0.02 mg/L for community water point 9 to 0.34 mg/L for community water point 2. This upper range is above recommended guidelines for domestic and industrial water and could cause colouration.

Table 7. Chemical characterization of water resources in the study area

Community No	NO ₃ ppm	SO ₄ ppm	Cl ppm	HCO ₃ ppm	Na ppm	K ppm	Ca ppm	Mg ppm	Fe ppm	Mn ppm	TH ppm
1	0.38	3.6	10.6	6.7	8.4	5.8	4.5	1.6	0.2	0.02	27
2	0.32	4.2	7.8	5.2	6.5	4.6	4.0	3.4	0.34	0.1	20
3	0.35	6.65	12.8	0.8	4.3	6.5	4.5	4.2	0.15	0.04	18.5
4	0.55	3.2	6.8	7.8	5.2	5.6	6.5	5.8	0.12	0.02	22.2
5	0.26	6.3	8.4	1.3	3.6	2.4	3.2	2.7	0.06	0.1	10.5
6	0.34	8.4	9.5	6.6	7.6	6.3	4.6	3.8	0.21	0.02	15.5
7	0.54	6.4	14.8	6.5	8.2	9.5	6.8	7.4	0.2	0.2	28.5
8	0.42	2.4	15.6	6.8	7.5	9.4	8.3	8.6	0.3	0.1	20.0
9	0.20	2.05	6.2	1.2	1.65	2.4	3.2	1.1	0.02	0.00	11.6
10	0.33	6.8	8.6	5.1	6.7	5.6	4.4	6.5	0.28	0.02	5.0
11	0.25	8.4	6.8	4.6	4.5	3.8	6.0	6.8	0.3	0.02	12.0

The chloride concentration ranged from 15.6 mg/L for community water point 8 to 6.2 mg/L for community water point 9 (Table 7). The range falls within the recommended guide level for drinking, domestic, and irrigation purposes. Nitrate values were appreciably low ranging from 0.2 mg/L for community water point 9 to 0.55 mg/L for community water point 4. This value still falls within the recommended guide level for water meant for drinking, irrigation, and domestic purposes. Sulphate concentration ranged from 2.05 mg/L for community water point 9 to a maximum concentration of 8.4 mg/L for community water point 11. This value is still within the acceptable limits for water meant for drinking, irrigation, and domestic uses. Bicarbonate concentration ranged from 0.8 mg/L for community water point 3 to a maximum of 6.8 mg/L for community water point 8. This value is still within the acceptable limits for water meant for drinking, irrigation, and domestic purposes. Differences in physicochemical and bacteriological characterization as recorded previously [32] in selected areas of the NDR may be attributed to regional variation in anthropogenic and natural sources of water contaminants in the region.

4.2 Statistical analysis

The factor analysis using principal component method is presented (Tables 8-13).

Table 8. Principal component analysis showing four communalities

	Initial	Extraction
EC	1.000	0.818
Salinity	1.000	0.556
Turbidity	1.000	0.971
TDS	1.000	0.896

Table 9. Principal component analysis showing total variance (Table 8 explained)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.211	55.271	55.271	2.211	55.271	55.271
2	1.030	25.750	81.020	1.030	25.750	81.020
3	0.611	15.272	96.293			
4	0.148	3.707	100.000			

Table 10. Principal component analysis showing 11 communalities

	Initial	Extraction
pH	1.000	0.853
Nitrate	1.000	0.782
Sulphate	1.000	0.793
Chloride	1.000	0.883
Bicarbonate	1.000	0.929
Sodium	1.000	0.906
Potassium	1.000	0.920
Calcium	1.000	0.865
Magnesium	1.000	0.949
Iron	1.000	0.761
Maganese	1.000	0.712

Table 11. Principal component analysis showing initial eigen values and extractive sum of square loadings

Component	Initial Eigen values			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.32	48.42	48.42	5.32	48.42	48.42
2	1.63	14.89	63.31	1.63	14.89	63.31
3	1.21	11.01	74.33	1.21	11.01	74.33
4	1.17	10.69	85.02	1.17	10.69	85.02
5	0.61	5.58	90.60			
6	0.53	4.85	95.46			
7	0.33	3.06	98.52			
8	0.11	1.07	99.60			
9	0.03	0.35	99.95			
10	0.00	0.04	100.00			
11	0.00	0.00	100.00			

The extraction method was used to perform principal component analysis (Tables 8-11). The communalities for

all of the variables included on the components were greater than 0.5 and had simple structures. Cumulative % of component explains the values of the total variance in variables which are included on the component. The result of PCA indicates four main controlling factors underlying the physical characterization of the water in the study area (Table 7). 81.02% of the total variance obtained from the component analysis suggests a strong loading of EC, salinity, turbidity and TDS. These parameters are strongly correlated and it is a suggestion of a common source [24]. 85.03% of the cumulative variance is observed in component 4 of the analysis thus suggesting a strong loading of pH, nitrate, sulphate, chloride, bicarbonate, sodium, potassium, calcium, magnesium, iron and manganese. Herein, these parameters are strongly correlated and are the suggestion of a common source of these ions.

4.3 Graphical analyses

The graphical analyses of the chemical data are presented (Figure 4-7).

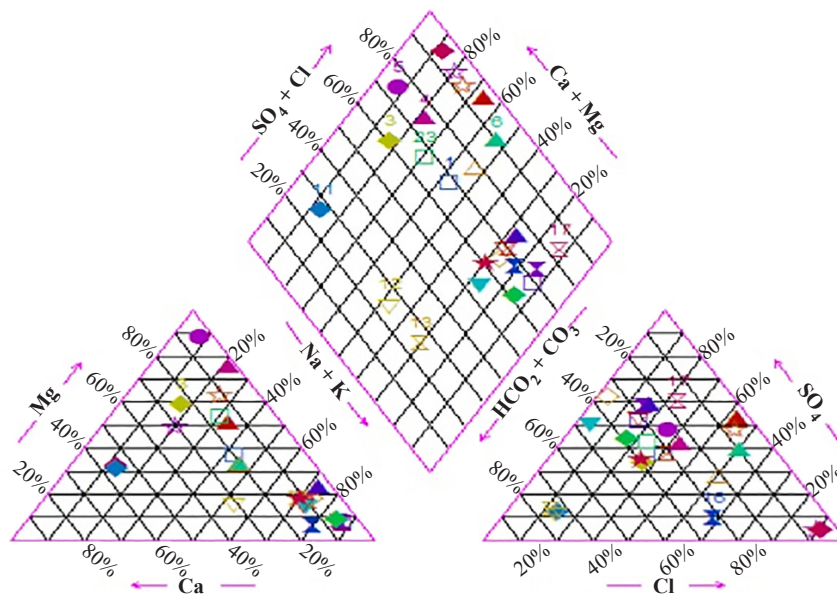


Figure 4. Piper diagram for groundwater in the study area

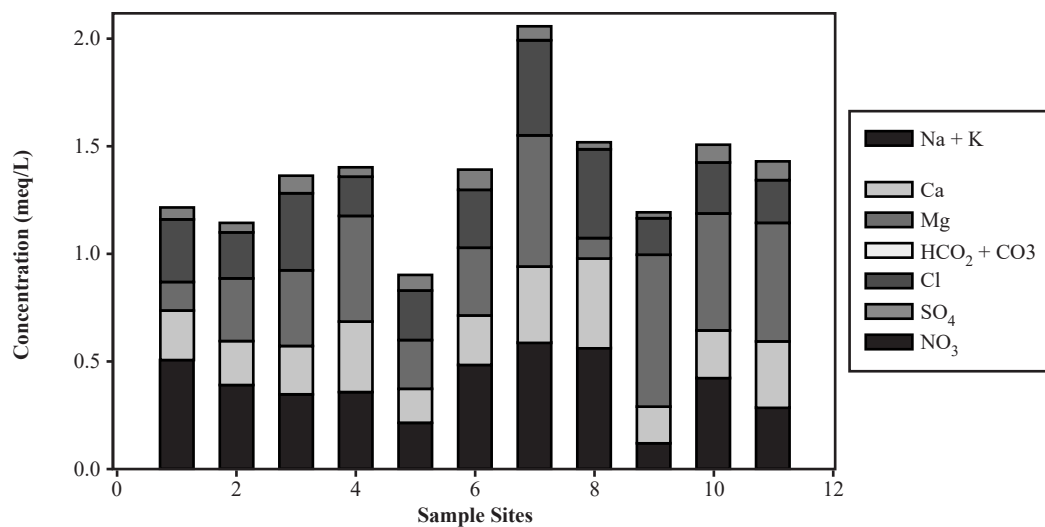


Figure 5. Collins diagram

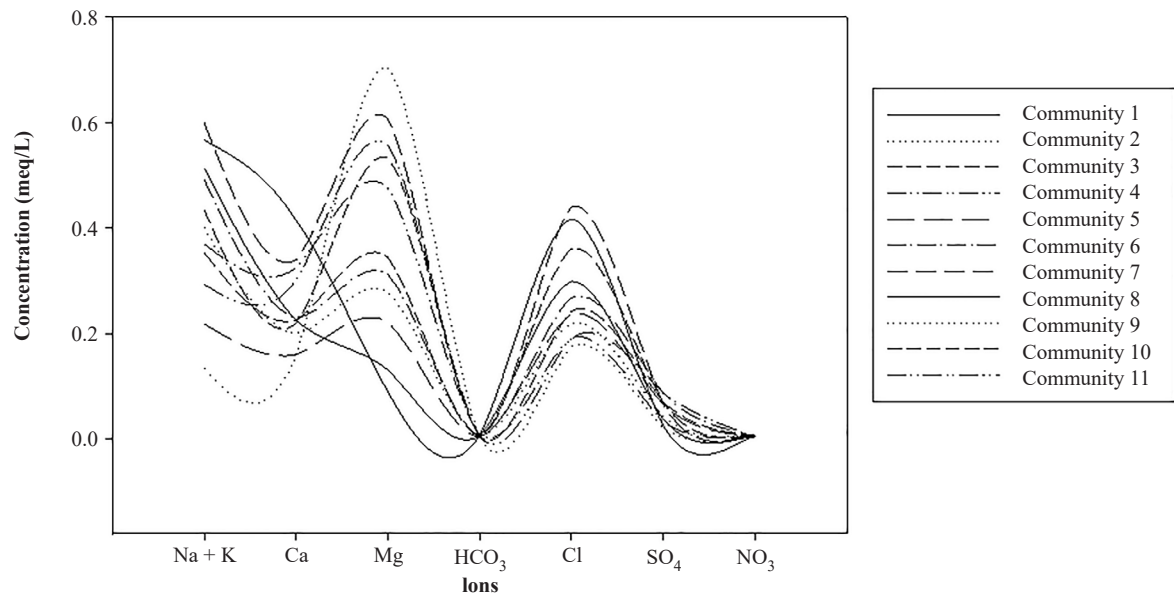


Figure 6. Schoeller diagram

The Schoeller diagram model revealed water quality with similar fingerprints with peaks of Mg, and Cl and a common trough for HCO₃ except for water points in community 1 and 8. Collins diagram revealed a TDS of about 2 meq/L for water from community 7 and the lowest TDS for water from community 5. These are surface water and groundwater respectively (Figure 5-6).

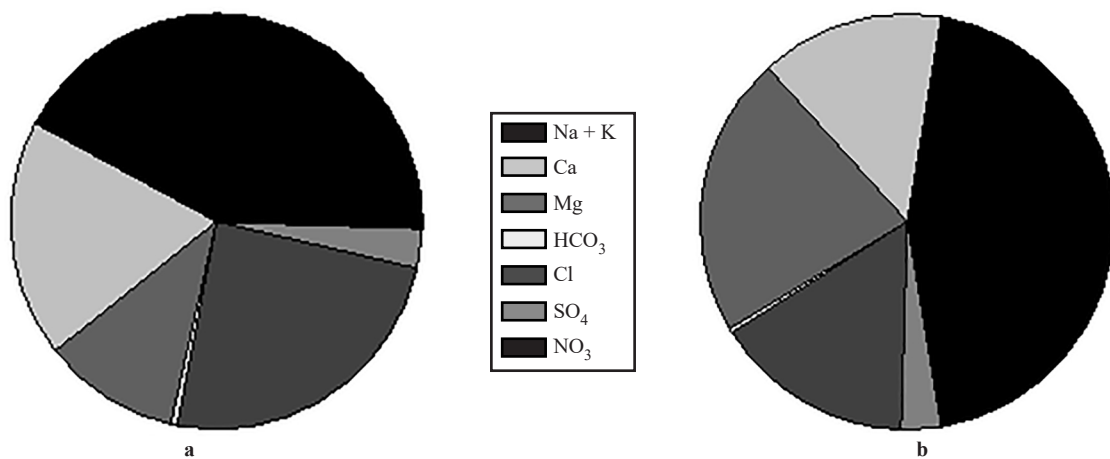


Figure 7. (a) Pie diagram for Community water point 1; (b) Pie diagram for overall water samples

The Pie diagram model revealed the highest percentage of Na + K ions, followed by Mg for the cations, and chloride followed by SO₄ for the anions (Figure 7). Variations of Na + K, Chloride, TDS from pie diagrams and Collins diagram indicate seawater intrusion on the water supply in the study area. There is an indication of the dominance of Na + K cations to other ions in the studied area (Figure 4).

The number of *Escherichia coli* and total coliforms are seen per 100 ml of some of the water samples provided (Table 12) and exceeds the recommended international and regional limits of 0 per 100 ml of the sample.

Table 12. Matrix of the bacteriology of water supply in the study area

Samples	E. coli	T. coli	TBC	Comments
Community 1	0	1	1	Not acceptable
Community 2	3	5	8	Not acceptable
Community 3	0	0	0	Acceptable
Community 4	0	0	0	Acceptable
Community 5	0	0	0	Acceptable
Community 6	0	0	0	Acceptable
Community 7	2	8	10	Not acceptable
Community 8	6	12	18	Not acceptable
Community 9	0	0	0	Acceptable
Community 10	1	2	3	Not acceptable
Community 11	2	5	7	Not acceptable

Therefore, some of these water supply in these communities do not meet the drinking water standards set by regulators. Water sources have influenced bacterial contamination significantly ($p < 0.01$). Bacterial contamination was common with surface water and groundwater from an unconfined aquifer. Therefore, wellhead protection is another important factor influencing bacterial contamination. The conditions which enhance the entrance of contaminants into drill water well include, limited extension of the wellhead about the ground surface, broken well cap, and no well housing. Therefore, there is need for proper land use planning and firm enactment and implementation of environmental laws in this region, to effect robust surface water and groundwater resource management.

The speciation of aqueous species and mineral saturation calculations were performed using PHREEQC and saturated index plotted against pH (Table 13 and Figure 8).

Table 13. pH versus PHREEQC saturation index of some mineral species

Samples	pH	SI		
		Ochre	Goethite	Hematite
Community 1	7.2	1.80	7.69	17.39
Community 2	6.98	3.99	9.88	21.76
Community 3	7.04	3.43	9.32	20.65
Community 4	6.87	1.66	7.55	17.11
Community 5	7.02	1.59	7.48	16.96
Community 6	7.04	1.95	7.84	17.68
Community 7	6.8	1.95	7.84	17.68
Community 8	6.9	3.94	9.84	21.68
Community 9	6.61	1.09	6.99	15.98
Community 10	6.78	2.15	8.04	18.09
Community 11	6.94	3.92	9.81	21.64

With increasing pH, the saturation index was in the order: hematite > goethite > ochre. The major minerals were

goethite, hematite and yellow iron oxide. Water points in community 1, 2, 4, 5, 6, 7, 9, 10 and 11 were under-saturated with siderite and calcite and dolomite, indicating the potential dissolution of these minerals. Community water point 2 and 8 were saturated with calcite, dolomite and siderite.

Furthermore, community water points 3 was saturated with siderite. Besides, all community groundwater was supersaturated with yellow iron oxide (Ochre), goethite and hematite indicating potential yellow colouration of the groundwaters. Harvested rainwater from community 9 apart from its dark colouration due to gas flaring possessed excellent drinking water criteria. Therefore, rainwater harvesting can readily substitute for groundwater exploration.

The overall water quality based on the Canadian Water Quality Index (CWQI) is marginal (Figure 9). Lowering of water quality is attributed to drinking and aquatic water criteria. The water quality is excellent for recreational and irrigation purposes and good for livestock utilization (Table 14 and Figure 9).

All sodium adsorption ratios were < 10 indicating community water point safe for irrigation with limited structural failure (Table 15). However, salt-sensitive plants may be affected. Residual sodium carbonate for all community water point revealed negative values. This pattern is suggestive of the fact that the water passes the civil engineering test of causing structural failure of engineering works. Also, the LSI for all community water point exhibited a negative pattern thus characterizing the water points as corrosive but no scaling [28].

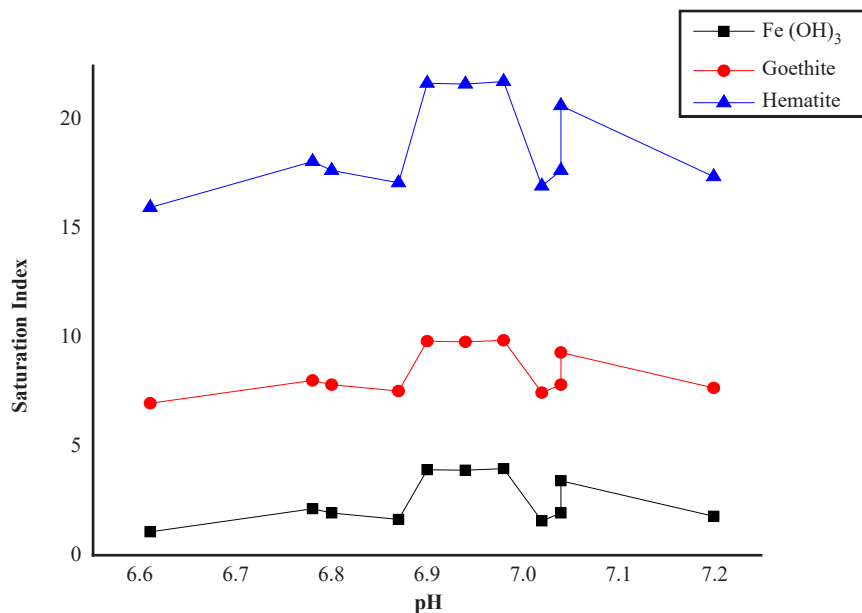


Figure 8. Plots of PHREEQC saturation index versus pH for community water points in the study area

Table 14. Data summary of CWQI in the study area

Data Summary	Overall	Drinking	Aquatic	Recreation	Irrigation	Livestock
CWQI	53	75	43	100	95	92
Categorization	Marginal	Fair	Poor	Excellent	Excellent	Good
F1 (Scope)	47	27	60	0	8	8
F2 (Frequency)	27	17	31	0	1	9
F3 (Amplitude)	60	29	71	0	0	6
Minimal Dataset Requirement of 4 Variables	Met	Met	Met	Not Met	Met	Met
Contaminant Analysis of Last Sample	Not Tested	Not Tested	Not Tested	Not Tested	Not Tested	Not Tested

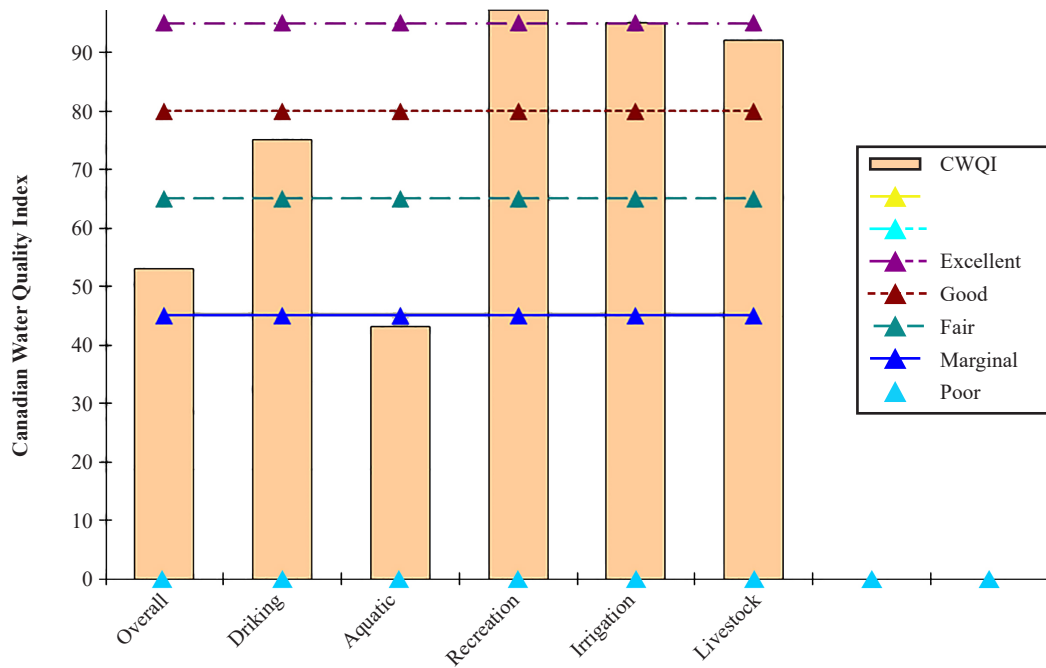


Figure 9. CWQI for water supply in the study area

Table 15. Sodium adsorption ratio (SAR), residual sodium carbonate (RSC) and Langelier Saturation Index (LSI) of community water points

Location/Code	SAR	RSC	LSI
Community 1	0.86	-0.34	-3.0
Community 2	0.57	-0.47	-3.4
Community 3	0.35	-0.56	-4.1
Community 4	0.35	-0.79	-3.1
Community 5	0.35	-0.37	-3.9
Community 6	0.63	-0.53	-3.2
Community 7	0.51	-0.93	-3.2
Community 8	0.43	-1.11	-3.0
Community 9	0.20	-0.24	-4.5
Community 10	0.47	-0.74	-3.5
Community 11	0.29	-0.85	-3.2

On a comparative study, Egunjobi [37] studied factors controlling water resources development in the NDR of Nigeria. Also, Idodo-Umeh [38] studied water Quality Assessment in the same region. All these studies failed to use data-driven graphical tools.

5. Summary and conclusions

Scientific data-driven models have been used to qualify the water resources in the NDR (i.e. South-Southern Niger

Delta Region of Nigeria). Graphical tools include PCA, descriptive statistics, piper diagram, Collins diagram, Schoeller diagram, pie diagram, phreeqc saturation indices, and the CWQI.

The study revealed that these community water points are fair for drinking (where the *Escherichia coli* and total coliforms are within the limits of 0 per 100 ml of the sample). However, these water points are poor for aquatic life, but good for livestock and excellent for both recreational and irrigation purposes. The application of principal component analysis techniques reflects that the physicochemical variables are contamination activities from a common source. It is recommended that well heads be properly built above ground level and provided with well cap. This precaution would limit contamination arising from domestic, soil erosion and vegetal effects. Besides, rainwater harvesting may be a good alternative to groundwater exploration due to the high annual rainfall in the studied area.

Acknowledgments

The authors are indebted to the authority of the corresponding author's institutional affiliation for the release of research allowances used in sponsoring this research.

Conflict of interest

The authors declare no competing financial interest.

References

- [1] Osei AS, Newman MJ, Patrick JAAM, Ayeh-Kumi F, Kwasi MO. Microbiological quality of packaged water sold in Accra, Ghana. *Food Control*. 2013; 31: 172-175. Available from: <https://doi.org/10.1016/j.foodcont.2012.08.025>.
- [2] Kolawole OM, Ajayi KT, Olayemi AB, Okoh AI. Assessment of water quality in Asa river (Nigeria) and its indigenous *clarias gariepinus* fish. *International Journal of Environmental Research and Public Health*. 2011; 8: 4332-4352. Available from: <https://doi.org/10.3390/ijerph8114332>.
- [3] Khan AW, Ahmad SA, Sayed MHSU. Arsenic contamination in groundwater and its effect on human health with particular reference to Bangladesh. *Journal of Preventive and Social Medicine*. 1997; 16: 65-73.
- [4] Azizullah A, Khattak MN, Richter P, Häder DP. Water pollution in Pakistan and its impact on public health-A review. *Environmental International*. 2011; 37: 479-497. Available from: <https://doi.org/10.1016/j.envint.2010.10.007>.
- [5] Dassenakis M, Scoullos M, Krasakopoulou E, Plavidou A, Kloukiniotou M. Effects of multiple source pollution on a small Mediterranean river. *Applied Geochemistry*. 1998; 13: 197-211. Available from: [https://doi.org/10.1016/S0883-2927\(97\)00065-6](https://doi.org/10.1016/S0883-2927(97)00065-6).
- [6] Nwankwoala HO, Abadom ND, Oborie E. Geochemical assessment and modeling of water quality for irrigation and industrial purposes in otuoke and environs, Bayelsa State, Nigeria. *Water Conservation & Management*. 2018; 2: 13-17. Available from: <https://doi.org/10.26480/wcm.01.2018.13.17>.
- [7] Amadi AN, Nwankwoala HO, Jimoh MO, Dan-Hasssan MA, Aminu T. Modelling the groundwater quality in parts of eastern Niger delta, Nigeria using multivariate statistical technique. *American Journal of Water Resources*. 2014; 2(5): 118-125. Available from: <https://doi.org/10.12691/ajwr-2-5-3>.
- [8] Maduka O, Ephraim-Emmanuel B. The quality of public sources of drinking water in oil-bearing communities in the Niger Delta region of Nigeria. *AAS Open Research*. 2019; 2(23): 23. Available from: <https://doi.org/10.12688/aasopenres.12964.1>.
- [9] Toriman ME, Hashim N, Hassan AJ, Mokhtar M, Juahir H, Gasim MB, Abdullah MP. Study on the impact of tidal effects on water quality modelling of Juru River, Malaysia. *Asian Journal of Scientific Research*. 2011; 4(2): 129-138. Available from: <https://doi.org/10.3923/ajsr.2011.129.138>.
- [10] Ye B, Yang L, Li Y, Wang W, Li H. Water sources and their protection from the impact of microbial contamination in rural areas of Beijing, China. *International Journal of Environmental Research and Public Health*. 2013; 10: 879-891. Available from: <https://doi.org/10.3390/ijerph10030879>.

- [11] Ashbolt NJ. Microbial contamination of drinking water and disease outcomes in developing regions. *Toxicology*. 2004; 198(1-3): 229-238. Available from: <https://doi.org/10.1016/j.tox.2004.01.030>.
- [12] Abbas MN, Mostafa GA. Determination of traces of nitrite and nitrate in water by solid phase spectrophotometry. *Analytica Chimica Acta Journal*. 2000; 410: 185-192. Available from: [http://dx.doi.org/10.1016/s0003-2670\(00\)00736-4](http://dx.doi.org/10.1016/s0003-2670(00)00736-4).
- [13] Denuault G. Electrochemical techniques and sensors for ocean research. *Journal of Ocean Science Discuss*. 2009; 6: 1857-1893. Available from: <http://dx.doi.org/10.5194/osd-6-1857-2009>.
- [14] Lamb D, Simpson D, Jensen BD, Gardner JS, Peterson QP. Determination of perchlorate in drinking water by ion chromatography using macrocycle-based concentration and separation methods. *Journal of Chromatography A*. 2006; 1118: 100-105. Available from: <http://dx.doi.org/10.1016/j.chroma.2006.01.138>.
- [15] Ngouna RH, Ratolojanahary R, Medjaher K, Dauriac F, Sebilo M, Junca-Bourie J. A data-driven method for detecting and diagnosing causes of water quality contamination in a dataset with a high rate of missing values. *Engineering Applications of Artificial Intelligence*. 2020; 95: 103822. Available from: <https://doi.org/10.1016/j.engappai.2020.103822>.
- [16] Liersch S, Cools J, Kone B, Koch H, Diallo M, Reinhardt J, Fournet S, Aich V, Hattermann FF. Vulnerability of rice production in the Inner Niger Delta to water resources management under climate variability and change. *Journal of Environmental Science and Policy*. 2013; 34: 18-33. Available from: <https://doi.org/10.1016/j.envsci.2012.10.014>.
- [17] Omo-Irabor OO, Olobaniyi SB, Oduyemi K, Akunna J. Surface and groundwater water quality assessment using multivariate analytical methods: a case study of the Western Niger Delta, Nigeria. *Physics and Chemistry of the Earth, Parts A/B/C*. 2008; 33(8-13): 666-673. Available from: <https://doi.org/10.1016/j.pce.2008.06.019>.
- [18] Aweto KE, Akpoborie IA. Estimating aquifer parameters with geoelectric soundings: Case study from the shallow Benin Formation at Orokpe, Western Niger Delta, Nigeria. *British Journal of Applied Science and Technology*. 2015; 6(5): 486. Available from: <https://doi.org/10.9734/BJAST/2015/14541>.
- [19] Ologunorisa TE. An assessment of flood vulnerability zones in the Niger Delta, Nigeria. *International Journal of Environmental Studies*. 2007; 61(1): 356-369. Available from: <https://doi.org/10.1080/0020723032000130061>.
- [20] May R, Jinno K, Tsutsumi A. Influence of flooding on groundwater flow in central Cambodia. *Environmental Earth Science*. 2011; 63(1): 151-161. Available from: <https://doi.org/10.1007/s12665-010-0679-z>.
- [21] Leton TG, Olujide O. Landfill operations in the Niger delta region of Nigeria. *Engineering Geology*. 2004; 73: 171-177. Available from: <https://doi.org/10.1016/j.enggeo.2003.12.006>.
- [22] Astel A, Tsakovski S, Barbieri P, Simeonov V. Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. *Water Research*. 2007; 41(19): 4566-4578. Available from: <https://doi.org/10.1016/j.watres.2007.06.030>.
- [23] Shrestha S, Kazama F, Nakamura T. Use of principal component analysis, factor analysis and discriminant analysis to evaluate spatial and temporal variations in water quality of the Mekong River. *Journal of Hydroinformatics*. 2008; 10(1): 43-56. Available from: <https://doi.org/10.2166/hydro.2008.008>.
- [24] Güler C, Thyne GD, McCray JE, Turner KA. Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeology Journal*. 2002; 10(4): 455-474. Available from: <https://doi.org/10.1007/s10040-002-0196-6>.
- [25] Zaporozec A. Graphical interpretation of water quality data. *Groundwater*. 1972; 10(2): 32-43. Available from: <https://doi.org/10.1111/j.1745-6584.1972.tb02912.x>.
- [26] Özcan H, Ekinci H, Baba A, Kavdır Y, Yüksel O, Yiğini Y. Assessment of the water quality of Troia for the multipurpose usages. *Environmental Monitoring and Assessment*. 2007; 130(1-3): 389-402. Available from: <https://doi.org/10.1007/s10661-006-9406-3>.
- [27] Chidambaram S, Anandhan P, Prasanna MV, Ramanathan AL, Srinivasamoorthy K, Kumar GS. Hydrogeochemical modelling for groundwater in Neyveli aquifer, Tamil Nadu, India, using PHREEQC: A case study. *Natural Resources Research*. 2012; 21(3): 311-324. Available from: <https://doi.org/10.1007/s11053-012-9180-6>.
- [28] Takdastan A, Mirzabeygi M, Yousefi M, Khodadadia A, Abbasnia R, Soleimani H, Mahvi AH, Na-ghan DJ. Neuro-fuzzy inference system prediction of stability indices and sodium absorption ratio in Lordegan rural drinking water resources in west Iran. *Data in Brief*. 2018; 18: 255-261. Available from: <https://doi.org/10.1016/j.dib.2018.02.075>.
- [29] Aghazadeh N, Mogaddam AA. Assessment of groundwater quality and its suitability for drinking and agricultural uses in the Oshnavieh area, Northwest of Iran. *Journal of Environmental Protection*. 2010; 1(1): 30. Available from: <https://doi.org/10.4236/jep.2010.11005>.
- [30] Shabbir R, Ahmad SS. Use of geographic information system and water quality index to assess groundwater quality in Rawalpindi and Islamabad. *Arabian Journal for Science and Engineering*. 2015; 40(7): 2033-2047. Available

from: <https://doi.org/10.1007/s13369-015-1697-7>.

- [31] Baghapour MA, Nasseri S, Djahed B. Evaluation of Shiraz wastewater treatment plant effluent quality for agricultural irrigation by Canadian Water Quality Index (CWQI). *Iranian Journal of Environmental Health Science & Engineering*. 2013; 10(1): 27. Available from: <https://doi.org/10.1186/1735-2746-10-27>.
- [32] Hutton G. Global costs and benefits of reaching universal coverage of sanitation and drinking-water supply. *Journal of Water and Health*. 2013; 11(1): 1-12. Available from: <https://doi.org/10.2166/wh.2012.105>.
- [33] Bacha AA, Durrani MI, Paracha PI. Physical and bacteriological characteristics of drinking water of Peshawar. *Pakistan. Journal of Nutrition*. 2010; 9(10): 1028-1033. Available from: <https://doi.org/10.3923/pjn.2010.1028.1033>.
- [34] Buikema Jr AL, Niederlehner BR, Cairns Jr J. Biological monitoring part IV-toxicity testing. *Journal of Water Research*. 1982; 16(3): 239-262. Available from: [https://doi.org/10.1016/0043-1354\(82\)90188-9](https://doi.org/10.1016/0043-1354(82)90188-9).
- [35] Shi P, Jia S, Zhang XX, Zhang T, Cheng S, Li A. Metagenomic insights into chlorination effects on microbial antibiotic resistance in drinking water. *Journal of Water Research*. 2013; 47(1): 111-120. Available from: <https://doi.org/10.1016/j.watres.2012.09.046>.
- [36] Amadi AN, Olasehinde PI, Yisa J, Okosun EA, Nwankwoala HO, Alkali YB. Geostatistical assessment of groundwater quality from coastal aquifers of Eastern Niger Delta, Nigeria. *Geosciences*. 2012; 2(3): 51-59. Available from: <https://doi.org/10.5923/j.geo.20120203.03>.
- [37] Egunjobi L. Water resource as a factor in the development of the Niger Delta Region. *Environmentalist*. 2012; 8(2): 109-114. Available from: <https://doi.org/10.1007/bf02240276>.
- [38] Idodo-Umeh G. Water quality assessment of river Areba, Niger Delta, Nigeria using physical and chemical indices. *Tropical Freshwater Biology*. 2013; 22: 73.