Heliyon 9 (2023) e16196

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

A multivariate statistical approach and water quality index for water quality assessment for the Rokel river in Sierra Leone

A. Barrie [a,*], S.K. Agodzo [b], R. Frazer-Williams [c], E. Awuah [d], E. Bessah [b]

[a] *Regional Water and Environmental Sanitation Centre Kumasi (RWESCK), Department of Civil Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana*
[b] *Department of Agricultural and Biosystems Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana*
[c] *Department of Chemistry, Fourah Bay College, The University of Sierra Leone, Freetown, Sierra Leone*
[d] *Department of Civil Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana*

ARTICLE INFO

Keywords:
Rokel river
Principal component analysis
Cluster analysis
Analysis of variance and
Water quality index

ABSTRACT

The main aim of the study was to use multivariate statistical approach to determine the relationship between parameters, identify the factors affecting the quality of water and interpret and group the water quality parameters. Water quality data was collected during two seasons; wet season spanning from June to August 2019 and dry season spanning from February to April 2019. The physiochemical and microbial parameters measured from the sampling process were turbidity, temperature, pH, electric conductivity, total hardness, calcium carbonates, total dissolved solid (TDS), dissolved oxygen (DO), total suspended solids (TSS), iron, nitrate, phosphate, potassium, sulphate, chromium, fluoride, e. coli and coliform. A total of 406 data set were collected and analysed using Principal Component Analysis, water quality index, cluster analysis and analysis of variance (ANOVA). These data sets were tested for sampling adequacy using Kaiser-Meyer-Olkin and Bartlett's Test and the result on the Kaiser-Meyer-Olkin Measure of Sampling Adequacy obtained was 0.615. The analysis yields Five PCs extraction with eigenvalues >1. These components explained 82.628% of the total variance of the entire components. The maximum water quality index 13 which indicated a grade A and can be treated for water supply. The following parameters Chromium 0.39 mg/l, Iron 1.88 mg/l, turbidity 18.66NTU, Phosphates 26.00 mg/l and fluorides 1.75 mg/l exceeded the WHO guidelines for drinking water. The mean values electrical conductivity is 12.26 μS/cm, 31.8 μS/cm for rain and dry seasons respectively., The following parameters Turbidity, Total Dissolved Solid, Total Suspended solids, Iron, Phosphate, Fluoride and Sulphate shows variation with High during the rain and low during the dry season with significant statical difference with a p value < 0.05. Whereas there is difference between the seasonal values of chromium, Nitrate and Potassium. The ANOVA resulted in P-value >0.05 which indicated no statistically significant different for chromium, Nitrate and Potassium. The seasonal variation was corroborated by cluster analysis with two clusters of C1 and C2. The PCs analysis, cluster analysis and ANOVA gave detailed characterization of the source and group correlation amongst the physiochemical and microbial parameters.

* Corresponding author.
E-mail address: barrieamadu12@gmail.com (A. Barrie).

https://doi.org/10.1016/j.heliyon.2023.e16196
Received 12 November 2022; Received in revised form 30 April 2023; Accepted 9 May 2023
Available online 8 June 2023
2405-8440/

1. Introduction

Water quality play a key role in maintaining a well-balanced environment [1]. Activities like mining, agricultures and land use changes does influence surface water quality [2]. Water is indispensable for all forms of life on Earth. According to the World Health Organization (WHO) a single human being requires between 50 and 100 L of water to ensure that its most basic needs are met. In fact, currently, approximately 884 million people still lack access to safe drinking water and more than 2.6 billion do not have access to basic sanitation. It is also estimated that approximately 1.5 million children under 5 years of age die each year as a result of water and sanitation-related diseases [3] and according to WHO, 88% of the diarrhoeal deaths are due to unsafe water, inappropriate sanitation, and lack of hygiene. In fact, an estimated 1 billion people (15% of the world population) still practice open defecation. Moreover, the majority (71%) of those without sanitation live in rural areas and 90% of all open defecation takes place in rural areas. This resolution acknowledges the importance of equitable access to safe and clean drinking water and sanitation as an integral component of the realization of all human rights [4].

It's reported that Africa is in dire situation and is only getting worse in terms of water scarcity. With Africa growing population, its predicted that by 2025, Africa will be close to 230 million and facing water scarcity and up to 460 million will be living in water stress areas (Africa Water Vision for 2025 [5].

It has also been reported that Sierra Leone has insubstantial water storage to last through their dry season. It withdraws only one-third the amount of freshwater of other countries in similar size. Chemicals used during agriculture production are polluting surface waters where many rural citizens collect their drinking water. Mining has caused land degradation and water pollution. Deforestation by mining has depleted water resources, as well as slash-and-burn farming, urbanization, and infrastructure building [6].

The Rokel river serves as a major source for water supply for villages along its flow parts. Two major district head quarter towns which are Lunser Town and Tonkolili District get supply from this river. Due to the catchment degradation and the increase in population of Freetown, the only alternative sources for water supply to Freetown is the Rokel River. Its flows from the north through major towns and drain to the ocean through the estuary in the Western area. The river drains an area of 10, 620 sq.-km with a length of 400 km along its course. There are two iron mining sites and the Bumbuna hydropower station along its drainage course carrying significant load of materials (dissolved and particulate phase) from natural and anthropogenic sources that cause negative effect on the river [7].

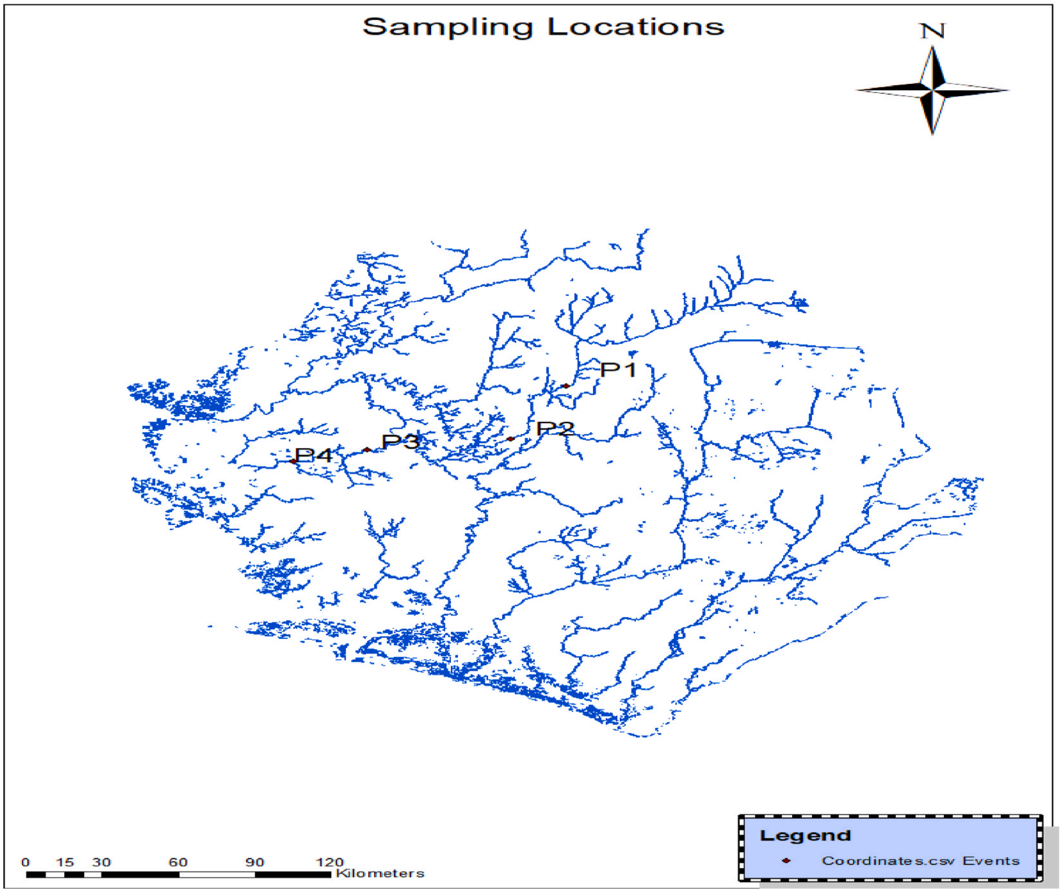

Fig. 1. Sampling points along the Rokel River.

The difficulty in the evaluation of surface water quality lies in the complexity in the analysis of large quantity of parameters. Usually, these parameters contain valuable information about the characterizer of the source and the interpretation of the parameters and these are the important step. Hence, the Principal Component Analysis can be used to provide a unique solution in the analysis of large quantity of parameters so that the original data can be reconstructed from the results. Principal components (PCs) actually take the cloud of data points and rotate it in such a way that maximum variability is visible. In other words, it identifies the most important gradients [8,9]. Many research has applied this method to evaluate water quality and have set the baseline information for management of environmental issues [10–12]. The Sample Adequacy for Principal Component Analysis is based on the *Kaiser-Meyer-Olkin and Bartlett's Test* which values range from a scale of 1–10. In this method, the total amount of variance is equal to the observed variables being analysed. The observed variables are standardized. Eigenvalues defines the amount of variance for each component and the Eigen vectors are used to determine component scores [10].

Further to evaluate the water quality, hierarchical cluster analysis and water quality index was applied. These three methods used would provide a detailed information and interpretation of the data sets for this research [13].

The main objective of this study was to characterize the source of water supply in order to determine the relationship between parameters, and identify the factors affecting the quality of water resources. The data obtained would be used for the design of constructed Wetland for pre-treatment in a water supply system.

2. Approach and methodology

2.1. Study area

The sampling location P1 is at Bumbuna close to the Hydropower (site coordinate: Longitude −11.7532 and latitude 9.052325). The second sampling location P2 is at Magburaka Town, which is downstream to Leone Rock iron mining site and close to the water supply treatment work with site coordinate longitude −11.9488 and latitude 8.72852. These two sites are upstream to the Rokel river as shown in Fig. 1. The two downstream locations are P3 and P4. The coordinates for these sites are P3 which is close to the Lunsar Iron mine site, is longitude (−12.4561) and latitude (8.6676). The site P4 located at the bridge linking the western area to northern region of the country is at longitude (−12.7173) and latitude (8.595607).

2.2. Data collection

The procedure involved in the collection of samples at four locations in the Rokel River Fig. 1 showed the sampling points in the river. These samples were collected from the subsurface at a depth of 20 cm for analysis. Monthly seasonal sample were collected, preserved and transported in accordance with standard method [14]. The dry seasonal period samples were collected during February, March and April 2019 whilst the wet period samples were collected in June July and August 2019. The *physiochemical and microbial parameters measured* from the samples were turbidity, temperature, pH, electric conductivity, total hardness, calcium carbonates, total dissolved solid (TDS), dissolved oxygen (DO), total suspended solids (TSS), iron, nitrate, Phosphate, potassium, sulphate, chromium, fluoride, *E. coli* and total Coliform. All *physiochemical* tests were done on site according to standard methods [14] using Hydro Kit HK3000. The bacteriological test was done at the laboratory and qualitative analysis was carried by multiple tube fermentation technique [14].

2.3. Data analysis

Principal Component Analysis (PCA) is a data mining Technique for data exploration which is used to observed pattern in the data with the aim to identify transformation of the data that define these patterns. This technique is built on an orthogonal, linear transformation of the data into new representation in space. This can be understood as the changing of original attributes by new attributes', referred to principal components. The principal component can be represented with r^{th} sample and k^{th} Component Y as shown below:

$$Y_{rk} = a_{1k}x_{r1} + a_{2k}x_{r2}\dots\dots\dots\dots a_{ik}x_{rp}$$

The correlation of variable X_i and principal component Y_{rk} is represented by:

$$r_{ij} = \sqrt{\left(a_{ij}^2 Var\left(Y_j\right)\right.} \Big/ S_{ij}$$

Table 1
Water quality rating as per weight arithmetic water quality index method.

WQI Value	Rating of Water Quality	Grading
0–25	Excellent water quality	A
26–50	Good water quality	B
51–75	Poor water quality	C
76–100	Very Poor water quality	D
Above 100	Unsuitable for drinking purpose	E

The data was imputed into SPSS 16.0 for computation using principal component analysis. With the parameters of different scale and magnitude, they were standardized for normalization of all parameters. To determine the sampling adequacy, the Kaiser -Meyer -Olkin (KMO) and Bartlett's test were used to determine the conformity of the samples to adequacy. These tests determined the proportioned variance which is common and might be affected by underlying factors. The range for this factor is (0–1). If the determined value is between 1 and 5, factor analysis would be useful. Value less than 0.5 indicate that PCA cannot be applied. In this study, the data was tested for sampling adequacy using KMO and Bartlett's Test. The results are presented on Table 3. Kaiser-Meyer-Olkin measure of sampling adequacy gave a value of 0.615 which is an indicator for PCA to be used. Table 6 Analysis of variance was carried out to check the effect of seasonality on the water quality parameters. ANOVA have been reported to be applied in water quality evaluation by many researchers [15,16].

2.4. Cluster analysis

Hierarchical cluster analysis (HCA) is another method of multivariate analysis used for the classification of physiochemical parameters without making any prior assumption about the data to classify the objects of the system into categories or clusters based on their nearness or similarity [16,17]. The Word Method was used for the classification since it used the approach of analysis of variance to determine the distances between clusters [17]. reported that using the Euclidean distance as distance measure and the Word Method as the linkage rule, this combination can result in the most distinctive groups in a dendrogram. The dendrograms provided the visual interpretation, summaries for the clusters, their proximity and with dramatic reduction of the originality of the data sets [18, 19].

2.5. Water quality index

In order to process the data, z-score normalization method was used. This phase is the key to improve the data for analysis after which the water quality index was calculated from the most significant parameters of the dataset. Then the water sample were classified on the basis of the water quality index as shown in Table 1 [20–22]. Many researchers have reported on the use of WQI [23–26].

3. Results and discussionss

The maximum readings of chromium, Iron, turbidity, phosphates and fluorides are higher than the accepted limits provided [27]. Presented in Table 2 are results of descriptive statistics of water physiochemical parameters and heavy metals analysed. The guideline for the eigenvalue-one rule only the principal components with values range from one and above are considered important [8,28,29]. These components explained 82.628% of the variance of the entire components. The scree plot Fig. 2 showed the five PCs which are the most significant components i.e. PC1 (40.762%), PC2 (14.726%), PC3 (11.858%), PC4 (8.471%) and PC5 (6.811%).

3.1. Principal Component Analysis

Results of *Kaiser-Meyer-Olkin (KMO)* and Bartlett's test are presented in Table 2.

Sampling adequacy for principal component analysis to be applicable. Component loading and communalities for each variable in five selected components before direct oblimin rotation were described in Table 4 and Fig. 3 showed Structural matrix for rotation and radar graph respectively. PC1 is correlated with (Phosphate, fluoride, Iron, Total Suspended solid Turbidity and Hardness) while Total dissolved Solid is negatively correlated (Table 5). PC2 correlated with (Coliform and *E-Coli*) were as Dissolved Oxygen is negative. PC3

Table 2
Descriptive statistics for water quality parameters.

Description	Units	Range	Minimum	Maximum	Mean	Std. Deviation
Turbidity	NTU	13.26	5.40	18.66	10.56	4.28
Temperature	^{o}C	6.30	23.90	30.20	27.84	1.60
pH		1.50	6.50	8.00	7.23	0.34
Electrical Conductivity	μS/cm	31.90	9.70	41.60	22.03	10.62
Calcium carbonate	mg/l	85.00	25.00	110.00	70.96	21.91
Total dissolved Solid	mg/l	8.39	7.83	16.22	12.01	2.48
Dissolved oxygen	mg/l	1.63	2.07	3.70	2.87	0.48
Total suspended solid	mg/l	6.80	7.00	13.80	9.66	2.41
Iron	mg/l	1.85	0.03	1.88	0.91	0.87
Nitrate	mg/l	0.84	0.01	0.85	0.22	0.30
Phosphate	mg/l	25.98	0.02	26.00	12.04	12.21
Potassium	mg/l	7.50	0.70	8.20	3.36	2.21
Sulphate	mg/l	6.00	10.00	16.00	12.19	1.62
Chromium	mg/l	0.27	0.12	0.39	0.22	0.06
Fluoride	mg/l	1.68	0.07	1.75	0.83	0.75
E. coli	CFU/ml	10.00	10.00	20.00	12.21	2.32
Total Coliform	CFU/ml	19.00	20.00	39.00	32.79	4.47

Table 3
Sampling adequacy for principal component analysis.

Kaiser-Meyer-Olkin (KMO) and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.615
Bartlett's Test of Sphericity	Approx. Chi-Square	474.075
df		136
Sig.		0.00

Table 4
Total variance explained.

Component		Initial Eigenvalues		Rotation (sum of squared loading
		Total % of Variance	Cumulative (%)	Total
1	6.930	40.762	40.762	6.817
2	2.503	14.726	55.488	2.344
3	2.016	11.858	67.346	2.454
4	1.440	8.471	75.817	1.699
5	1.158	6.811	82.628	1.268
6	0.947	5.570	88.198	
7	0.617	3.630	91.828	
8	0.498	2.932	94.760	
9	0.369	2.169	96.930	
10	0.244	1.436	98.366	
11	0.135	0.795	99.161	
12	0.060	0.356	99.516	
13	0.039	0.229	99.746	
14	0.026	0.156	99.902	
15	0.009	0.055	99.957	
16	0.007	0.041	99.998	
17	0	0.002	100	

Table 5
Structural matrix for rotation.

	PC1	PC2	PC3	PC4	PC5
Turbidity	**0.904**	−0.080	0.239	−0.113	−0.106
Temperature	−0.098	0.296	**0.606**	**0.558**	−0.090
pH	0.277	**0.758**	0.270	−0.291	0.214
EC	−0.916	0.033	0.226	−0.105	−0.150
CaCO3	0.528	−0.297	0.112	0.551	−0.474
TDS	−0.786	−0.124	0.335	−0.063	−0.038
DO	−0.184	−0.746	0.447	−0.231	0.237
TSS	**0.933**	−0.132	0.191	−0.075	−0.066
Iron	**0.980**	0.003	−0.150	0.048	0.030
Nitrate	−0.025	−0.167	0.126	**0.522**	**0.742**
Phosphate	**0.985**	−0.029	−0.091	0.056	0.027
Potassium	−0.231	0.219	0.293	0.481	0.036
Sulphate	**0.519**	−0.037	**0.627**	−0.132	0.157
Chromium	0.394	0.434	**0.578**	−0.317	0.041
Flouride	**0.979**	−0.036	−0.141	0.053	0.044
E. coli	−0.127	**0.773**	0.050	0.098	−0.266
Coliform	0.037	**0.540**	−0.473	0.120	0.361

Correlated with pH, Chromium and Sulphate PC4 correlated with Temperature and Potassium and PC 5 Correlated with Nitrate. Further Table 6 defined the relationships between variables and the results of correlations between parameters.

3.2. Output from cluster analysis

The output from the Ward Method using the Euclidean distance as a distance measure resulted in the following dendrograms in Fig. 4a–c. According to the cluster analysis shown in the dendrogram Fig. 4a for the sampling points represented as C1 formed by P1 and P2 which are upstream. Whereas cluster C2 formed by P3 and P4 which are downstream. This showed a clear distinction between the location of the point in the river and their closest distance. With Fig. 4b it was observed that the dendrogram is clustered seasonally that is C1 formed the wet months of August, July and June. Whereas cluster C2 formed the dry months of April March and February.

Table 6
Correlation of water quality parameters.

	Turbidity	Temp	pH	EC	$CaCO_3$	TDS	DO	TSS	Fe	NO_3-	PO_4^{3-}	K	SO_4^{2}-	Cr	F^-	*E.* coli	Coliform
Turbidity	1																
Temp	−0.05	1															
pH	0.27	0.13	1														
EC	−0.76	0.25	−0.15	1													
CaCO3	0.52	0.30	−0.29	−0.41	1												
TDS	−0.50	0.20	−0.16	0.79	−0.31	1											
DO	0.02	−0.08	−0.44	0.21	−0.10	0.35	1										
TSS	0.97	−0.05	0.21	−0.80	0.55	−0.57	0.02	1									
Fe	0.84	−0.16	0.24	−0.94	0.50	−0.83	−0.26	0.87	1								
NO3-	−0.08	0.24	−0.07	−0.12	0.00	0.11	0.19	−0.05	0.02	1							
PO_4^{3-}	0.87	−0.13	0.23	−0.93	0.52	−0.79	−0.21	0.89	0.99	0.03	1						
K	−0.20	0.26	0.10	0.15	0.00	0.13	−0.04	−0.22	−0.25	0.15	−0.24	1					
SO_4^{2}-	0.50	0.21	0.34	−0.37	0.20	−0.33	0.29	0.53	0.40	0.02	0.44	0.05	1				
Cr	0.51	0.32	0.61	−0.14	0.01	−0.15	−0.06	0.47	0.28	−0.13	0.28	−0.07	0.46	1			
F^-	0.84	−0.17	0.21	−0.94	0.50	−0.81	−0.21	0.87	0.99	0.03	1	−0.24	0.42	0.25	1		
E-coli	−0.11	0.27	0.42	0.09	−0.20	0.04	−0.54	−0.21	−0.12	−0.18	−0.13	0.25	−0.09	0.18	−0.15	1	
Coliform	−0.15	0.03	0.23	−0.14	−0.21	−0.25	−0.50	−0.11	0.09	0.07	0.04	−0.06	−0.25	0.07	0.08	0.17	1

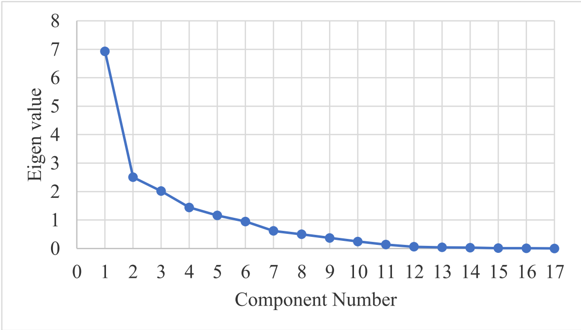

Fig. 2. Scree plot of eigenvalues.

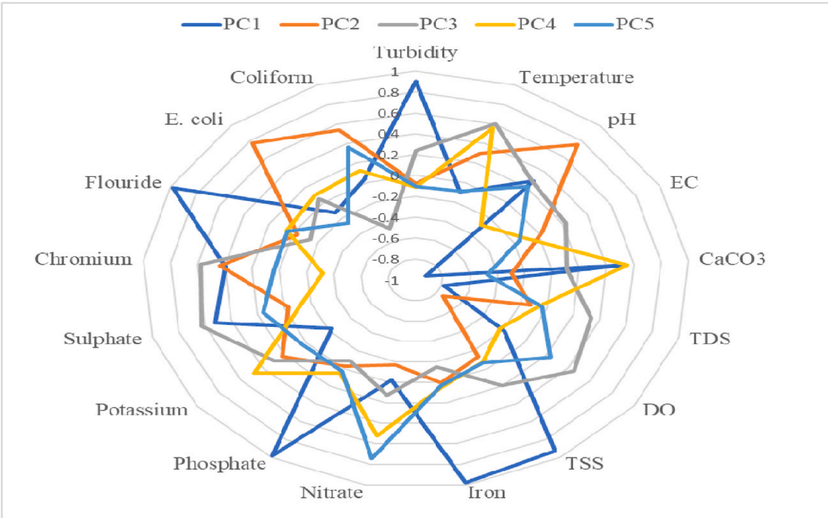

Fig. 3. Structural matrix for rotation radar graph.

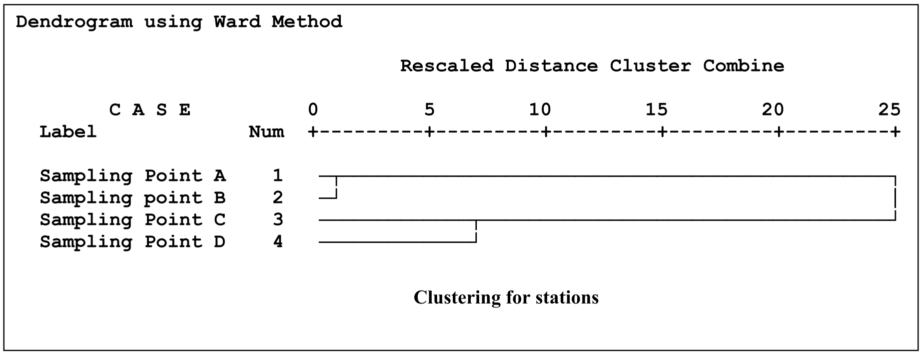

Fig. 4a. Clustering for stations.

This indicated that seasonality effect in the clustering.

Water quality index calculation.

The water quality index was calculated using the parameters obtained from the river Rokel at the locations defined in Fig. 1 WQI was calculated using the following relationship

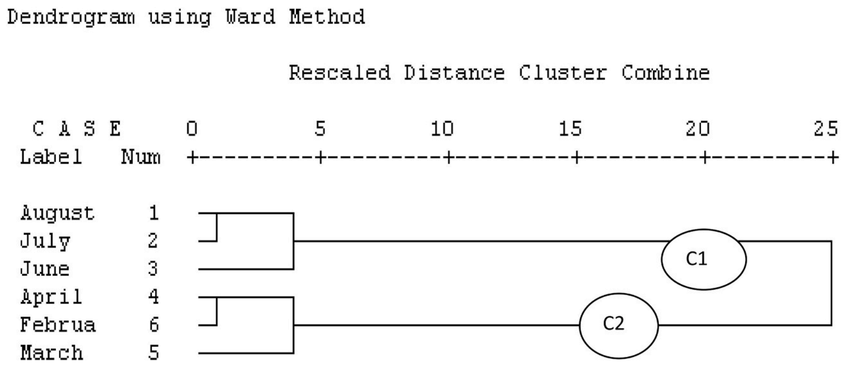

Fig. 4b. Clustering for months.

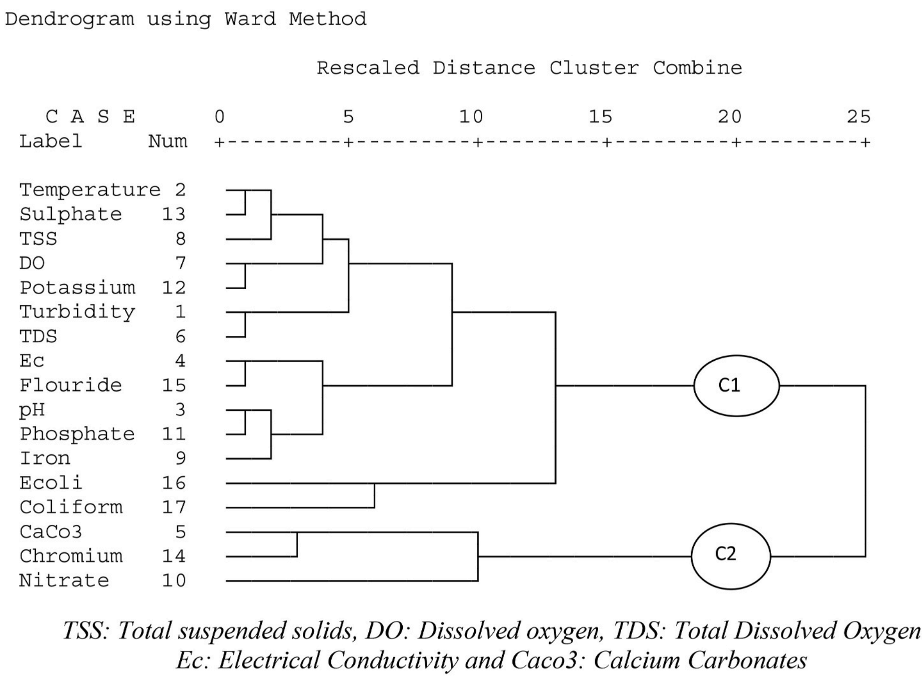

Fig. 4c. Clustering for physio chemical parameters. TSS: Total suspended solids, DO: Dissolved oxygen, TDS: Total Dissolved Oxygen. Ec: Electrical Conductivity and Caco3: Calcium Carbonates.

$$WQI = \frac{\sum_{j=1}^{n} q_j x w_j}{\sum_{j}^{n} wj} \quad (1)$$

Where n is the total number of parameters included in WQI in the computation q_j which is the quality rating parameter j computed by equation (1). w_j the unit weight for each parameter in the calculated equation

$$q_j = 100\, x\, \frac{\{V_j - V_{ideal}\}}{\{S_i - V_{ideal}\}} \quad (2)$$

Where Vj was measured value of parameter j in the tested water sample V_{ideal} was the ideal value parameter of parameter j in pure water the value is zero for all parameters except for Do = 14.6 mg/l and pH = 7.0 and s_j is recommended WHO standard value.

Where also

$$w_j = \frac{K}{S_j} \quad (3)$$

Where K is the proportionality constant that can be computed from

$$K = \frac{1}{\sum_{j=1}^{n} S_j} \tag{4}$$

These values were computed from equation (3,4)

Where mean (μ) and standard deviation α

Sampling point 4 showed the highest quality index whilst sampling station 1 showed the least index value Fig. 5.

3.3. Seasonal variation of water quality parameters

3.3.1. Total suspended solids (TSS)

The mean TSS for the wet season concentration was 11.74 mg/l and the dry season concentration was 7.50 mg/l. The analysis of variance (ANOVA) indicated significant difference between the mean of the rain season and dry season, with test of significant $p < 0.05$. TSS is one of the variables associated with PC 1. These results are consistent with the finding [29] which showed statically significant different $p < 0.05$, with mean values of 38 mg/l for the dry period and of 69.7 mg/l for the rainy period. Further a mean value of 23.5 mg/l and 2423.65 mg/l for the dry and rainy season respectively was reported [30].

3.3.2. Turbidity

The mean turbidity for the wet season concentration was 14.12 NTU and the dry season concentration was 6.99 NTU. The analysis of variance indicated significant difference between the means of the rain season and dry season, with test of significant $p < 0.05$. Turbidity was one of the variables associated with PC1. The increase in turbidity during the rains was due to runoff [29]. The mean wet season values of 196.5 NTU and mean dry season value of 0.57 was reported [30]. Similarly, mean values of 29 NTU for the dry period and 96.2 NTU for the rainy period was recorded [29]. Also, mean values of 28.5NTU for the dry period and 63NTU for the rainy period was reported [31]. These findings corroborated the outcome of this research. But $p > 0.05$ which is contrary to this finding was also reported [31].

3.3.3. Total dissolved solid (TDS) and electrical conductivity

The mean total dissolved solid for the wet season concentration was 10.04 mg/l and the dry season concentration was 13.98 mg/l. The analysis of variance (ANOVA) indicated significant difference between the mean of the rainy season and dry season, with test of significant $p < 0.05$. TDS was one of the variables associated with PC 3 but a very weak correlation of 0.335.

Mean values of 1760 mg/l and 2 mg/l for the dry and rainy seasons respectively was reported [30]. Mean seasonal TDS values which are statically significantly different with $p < 0.05$ was reported but, these values which are 2304.5 mg/l and 840.3 mg/l for the rainy and dry season respectively are contrary to this finding [31]. Also, mean TDS values of 53 mg/l and 90 mg/l for the dry and rainy season respectively was recorded [29].

The mean electrical conductivity for the wet season concentration was 12.26 μS/cm and the dry season concentration was 31.80 μS/cm. Mean values of 3520 μS/cm and 466 μS/cm for the dry and rainy season respectively was recorded [30]. Also reported was 784 μS/cm and 295.2 μS/cm for the dry and rainy reasons respectively [32]. Mean values of 252 μS/cm and 235 μS/cm for rain and dry seasons respectively was reported which are contrary to this study [29].

The analysis of variance (ANOVA) indicated no significant difference between the mean of the rain season and dry season, with test of significant $p > 0.05$ Electrical conductivity was one of the variables associated with PC 1 but a negative correlation

3.3.4. pH, temperature and Dissolved Oxygen (DO)

The mean wet season pH was 7.31 and dry season was 7.23. There was no significant different between the means of the two seasons with the p – value been 0.305 which is great than 0.05 which is an indicator for difference between means. Most researchers reported the pH range from 6.5 to 8.65 for both seasons [29,30,33]. The mean wet season temperature was 27.62 °C and the dry season was 28.07 °C. The analysis of variance (ANOVA) indicated no significant difference between the mean of the rain season and dry season, with test of significant $p > 0.05$ Temperature was one of the variables associated with PC 3. The mean wet season dissolved oxygen concentration was 2.75 mg/l and the dry season concentration was 2.98 mg/l. Dissolved oxygen was one of the variables associated with PC 3 but very weak correlation. Most researchers reported the dissolve oxygen range from 1.2 to 6.8 mg/l for both seasons [29,30,33,34].

3.3.5. Calcium carbonates ($Caco_3$)

The mean wet season $Caco_3$ concentration was 82.29 mg/l and the dry season concentration was 59.67 mg/l. The analysis of variance (ANOVA) indicated significant difference between the mean of the rain season and dry season, with test of significant $p < 0.05$ CaCo3, one of the variables associated with PC 4 but a weak correlation of 0.551. Mean values of 145.2 mg/l and 170.3 mg/l for the dry and wet seasons respectively was reported which is consistent with this finding [32]. A contrary mean values of 180 mg/l and 88 mg/l for dry and wet reasons was reported [34].

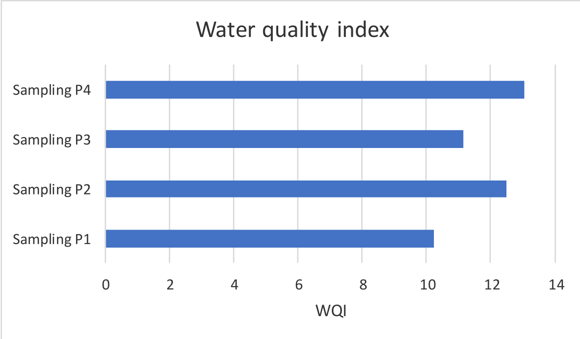

Fig. 5. Water quality along the sampling sites.

3.3.6. *Fluoride*

The mean wet season fluoride concentration was 1.56 mg/l and the dry season concentration was 0.11 mg/l. The analysis of variance (ANOVA) indicated no significant difference between the mean of the rain season and dry season, with test of significant $p > 0.05$ was one of the variables associated with PC 1 The level of fluoride can be attributed to the erosion fluoride-bearing minerals in the geologic substrate in a simple linear fashion and there is no significant difference between the mean values [35,36]. This finding differs from recorded mean values of 0.14 mg/l and 5.45 mg/l for the rainy and dry seasons respectively, which showed a higher value for the dry season and lower for the raining season. But regarding the statistically difference the finding are corroborated [31,37].

3.3.7. *Chromium and iron*

The mean wet season chromium concentration was 0.24 mg/l and the dry season concentration was 0.21 mg/l. The analysis of variance (ANOVA) indicated no significant difference between the mean of the rain season and dry season, with test of significant $p > 0.05$. Chromium was one of the variables associated with PC 1. The mean wet season iron concentration was 1.76 mg/l and dry season concentration was 0.07 mg/l. The analysis of variance (ANOVA) indicated significant difference between the mean of the rain season and dry season, with test of significant $p < 0.05$. Iron was one of the variables associated with PC 1. The iron and chromium variations can be attributed to the erosion metals from bearing minerals in the geologic substrate in a simple linear fashion and mining runoff into the tributaries of the river [35,36]. There are statically significant difference between the mean values for both iron and chromium. Mean values of 0.6 mg/l and 1.02 mg/l for the dry and wet period for iron respectively was reported [33,36]. Also mean values of 0.01 mg/l and 0.018 mg/l for the dry and wet period for chromium reported which corroborated this study [38].

3.3.8. *Nitrate and potassium*

The mean wet season nitrate concentration was 0.22 mg/l and the dry season concentration 0.21 mg/l. The analysis of variance (ANOVA) indicated no significant difference between the mean of the rain season and dry season, with test of significant $p > 0.05$. Nitrate was one of the variables associated with PC 4 but with a weak correlation. Researches have reported values of 46.85 mg/l, 3.16 mg/l for the wet season and 37.40 mg/l and 1.2 mg/l for the dry season which corroborated this study [32,34].

The mean wet season potassium concentration was 2.84 mg/l and the dry season concentration was 3.89 mg/l. Potassium was one of the variables associated with PC 4 but with a weak correlation., Mean values of 0.55 mg/l and 1.74 mg/l for the wet and dry seasons respectively was reported which corroborated this study [31]. A contrary mean values of 15 mg/l and 11.5 mg/l was reported with no statically significant mean difference recorded [32].

3.3.9. *Phosphate and sulphate*

The mean wet season phosphate concentration was 23.97 mg/l and the dry season concentration was 0.12 mg/l. The analysis of variance (ANOVA) indicated significant difference between the mean of the rain season and dry season, with test of significant $p < 0.05$. Phosphate was one of the variables associated with PC 1. Mean values of 0.02 mg/l and 16.20 mg/l was reported for the dry and rainy seasons respectively [31]. But also, mean values 2.32 mg/l and 1.15 mg/l for the dry and rainy seasons respectively was reported which was contrary to this findings [34].

The mean wet season sulphate concentration was 12.83 mg/l and the dry season concentration was 11.55 mg/l. The analysis of variance (ANOVA) indicated significant difference between the mean of the rain season and dry season, with test of significant $p < 0.05$ sulphate was one of the variables associated with PC 3. Mean values of 10.50 mg/l and 70 mg/l for the dry and the rainy seasons respectively was reported [31]. But a contrary report in terms of seasonal variation 79.8 mg/l and 45 mg/l for the dry and rainy seasons respectively was reported [32].

3.3.10. *E. coli and Total Coliform*

The mean wet season *E. coli* concentration was 11.92 CFU/ml and the dry season concentration was 12.50 CFU/ml. The analysis of variance (ANOVA) indicated no significant difference between the mean of the rain season and dry season, with test of significant $p > 0.05$ *E. coli* was one of the variables associated with PC 4 but with a weak correlation. Mean values of 0 CFU/100 ml and 9000 CFU/

100 ml for the dry and rainy seasons respectively was reported [29].

The mean wet season total coliform was 33.08 CFU/ml and dry season concentration was 32.79 CFU/ml. The analysis of variance (ANOVA) indicated no significant difference between the mean of the rainy season and dry seasons, with test of significant $p > 0.05$. Total Coliform was one of the variables associated with PC 2. Mean values of 53000 CFU/100 ml and 3 CFU/100 ml for the wet and dry seasons respectively was reported [30]. Also reported was mean values of 1584 CFU/ml and 1680 CFU/ml for the dry and rainy seasons respectively [30].

4. Conclusions

The principal component was used to characterized the quality of water in Rokel River. The detailed output from the analysis showed five components that defined *82.628%* of the total variances.

The water quality index was used to define the quality at the specified location along the course of the river. The highest water quality index was 13 for sampling location P3 which showed it's a grade A source and the source is good for portable water supply upon treatment to remove the heavy metal and reduction of the nutrients.

The water quality showed seasonal variation in quality due to increase in runoff into the river tributary. The results indicated chromium, iron, turbidity, phosphates and fluorides gave values which are higher than guideline provided by WHO.

The following parameters turbidity, total dissolved solid, total suspended solids, iron, phosphate, fluoride and sulphate showed seasonal variations with high values during the rain and low during the dry season with significant statistical difference, it's as a result of direct runoff and human anthropogenic action link to seasonality.

Whereas chromium, Nitrate and Potassium showed seasonal variation but the result did not show any statistically significant different. The variations are corroborated by the cluster analysis.

The component technique, water quality index, analysis of variance and the cluster provided a reliable classification of surface water quality and in the future can be applied in evaluations of water quality in our river systems The study would be a monitoring tool for environmental management.

Author contribution statement

Amadu Barrie: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

S. K. Agodzo; R. Frazer-Williams; E. Awuah; E. Bessah: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Funding statement

This work was supported byRegional Water and Environmental Sanitation Centre Kumasi (RWESCK) (ACE I - P126974-Cr. No. 54230).

Data availability statement

Data associated with this study has been deposited at Reserved DOI under the accession number 10.17632/b46hs74yjf.1.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was funded by the Regional Water and Environmental Sanitation Centre Kumasi (RWESCK) at the Kwame Nkrumah University of Science and Technology (KNUST), Kumasi with funding from Ghana Government through the World Bank under the Africa Centre of Excellence project.' The views expressed in this paper do not reflect those of the World Bank, Ghana Government and KNUST.''

References

[1] Fikret Ustaoğlu, Beyhan Taş, Yalçın Tepe, Halim Topaldemir, Comprehensive assessment of water quality and associated health risk by using physicochemical quality indices and multivariate analysis in Terme River, Environ. Sci. Pollut. Control Ser. (2021), https://doi.org/10.1007/s11356-021-15135-3 Turkey.

[2] B. Tas¸, Y. Tepe, F. Ustao glu, S. Alptekin, Benthic algal diversity and water quality evaluation by biological approach of Turnasuyu Creek, NE Turkey, Desalination Water Treat. 155 (2019) 402–415.

[3] V. Amiri, S. Kamrani, A. Ahmad, P. Bhattacharya, J. Mansoori, Groundwater quality evaluation using Shannon information theory and human health risk assessment in Yazd province, central plateau of Iran, Environ. Sci. Pollut. Res. 28 (1) (2021) 1108–1130.

[4] Roman Rollnick, Thierry Naudin, Annual Report, Publishing Services Section, ISO 14001:2004-certified, Copyright © United Nations Human Settlements Programme (UN-HABITAT), Business Media, LLC, Nairobi, 2010, 2007.
[5] UNECA, Africa Water Vision for 2025: Equitable and Sustainable Use of Water for Socioeconomic Development, 2000.
[6] I. Diana, Ortiz 1, Marta Piche-Ovares, M. Luis, Romero-vega Joseph wagman, Adriana troyo, The Impact of Deforestation, Urbanization, and Changing Land Use Patterns on the Ecology of Mosquito and Tick-Borne Diseases in Central America, Insects 13 (2022) 20, https://doi.org/10.3390/insects13010020.
[7] S. Gao, Z. Wang, Q. Wu, J. Zeng, Multivariate statistical evaluation of dissolved heavy metals and a water quality assessment in the Lake Aha watershed, Southwest China, PeerJ 8 (2020), e9660, https://doi.org/10.7717/peerj.9660.
[8] A. Cicek, E. Kose, C. Tokatl, Using factor ana lysis to evaluate sediment quality of a significant mining area in Turkey. Polish Journal of Environmental Studies 28:2021–2025. classification, Water Poll. Contr. 77 (2019) 271–284.
[9] F. Rakotondrabe, Water quality assessment in the Bétaré-Oya gold mining area (East Cameroon): multivariate statistical analysis approach, 2018, Sci. Total Environ. 610–611 (2018) 831–844.
[10] R. Kumari, R.C. Sharma, Assessment of water quality index and multivariate analysis of high altitude sacred Lake Prashar, Himachal Pradesh, India, Int. J. Environ. Sci. Technol. 16 (2019) 6125–6134 ([Google Scholar] [CrossRef).
[11] W. Yang, Y. Zhao, D. Wang, H. Wu, A. Lin, L. He, Using principal components analysis and IDW interpolation to determine spatial and temporal changes of surface water quality of xin'anjiang river in huangshan, China, Int. J. Environ. Res. Public Health 17 (2020) 2942 ([Google Scholar] [CrossRef]).
[12] G. Ioele, M. De Luca, F. Grande, G. Durante, R. Trozzo, C. Crupi, G. Ragno, Assessment of surface water quality using multivariate analysis: case study of the crati river, Italy, 2020, Water 12 (2020) 2214 ([Google Scholar] [CrossRef]).
[13] Z. Shahrizim, Work Shop on Statistical Analysis for Ecological Data, 2018 [supplementary materials], http://www.inspem.upm.edu.my/activities/workshop_on_statistical_analysis_for_ecologcal_data_iii_2018-17265?L=eng.
[14] APHA, Standard Methods for the Examination of Water and Wastewa Ter, twentieth ed., American Public Health Association (APHA), American Water Works Association (AWWA) and Water Environment Federation (WEF), Washington, DC, 1998.
[15] I. Sewanu, A. Isaac, I. Balogun, A.S. Okinawan, Integrated approaches to groundwater quality assessment and hydro chemical processes in Lagos, Nigeria, Appl. Water Sci. 8 (2018) 1–19, https://doi.org/10.1007/s13201-018-0847.
[16] a Bernard Parinet, b Antoine Lhote a, Legube Bernard, Principal Component Analysis: an Appropriate Tool for Water Quality Evaluation and Management—Application to a Tropical Lake System, 0304-3800/$ – See Front Matter © 2004, Elsevier B.V. All rights reserved, 2004, https://doi.org/10.1016/j.ecolmodel.2004.03.007.
[17] C. Guler, G. Thyne, J. McCray, K. Turner, Evaluation of graphical and multivariate statistical methods for classification of water chemistry data, Hydrogeol. J. 10 (4) (2002) 455–474.
[18] Handan Aydin, Fikret Ustaoglu, Yalc¸ın Tepe, Elif Neyran Soylu, Assessment of Water Quality of Streams in Northeast Turkey by Water Quality Index and Multiple Statistical Methods, Environmental Forensics, 2020, https://doi.org/10.1080/15275922.2020.1836074.
[19] S. Shrestha, F. Kazama, Assessment of Surface Water Quality Using Multivariate Statistical Techniques: A Case Study of the Fuji River Basin; Japan vol. 22, Environmental Modelling & Software, 2007, pp. 464–475.
[20] K.P. Singh, A. Malik, D. Mohan, S. Sinha, Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India): a case study, Water Res. 38 (2004), 3980e3992.
[21] J.H. Kim, R.H. Kim, J. Lee, T.J. Cheong, B.W. Yum, H.W. Chang, Multivariate statistical analysis to identify the major factors governing groundwater quality in the coastal area of Kimje, South Korea, Hydrol. Process. 19 (2005), 1261e1276.
[22] S. Shrestha*, F. Kazama, Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, 2007, Japan Environmental Modelling & Software 22 (2006), 464e475.
[23] S.H. Ewaid, S.A.A. Al-Hamzawi, Water quality index for Al-Gharraf River, southern Iraq, Egyp. J. Aqu. Res. 43 (2017) 117–122.
[24] E. Mutlu, Evaluation of spatio-temporal variations in water quality of Zerveli stream (northern Turkey) based on water quality index and multivariate statistical analy ses, Environ. Monit. Assess. 191 (6) (2019) 335.
[25] G.A. Aliyu, N.R.B. Jamil, M.B. Adam, Z. Zulkefee, Assessment of Guinea Savanna River system to evaluate water quality and water monitoring networks, Glob. J. Environ. Sci. Manag. 5 (3) (2019) 345–356.
[26] H. Aydın, F. Ustaoğlu, Y. Tepe, E.N. Soylu, Assessment of water quality of streams in northeast Turkey by water quality index and multiple statistical methods, Environ. Forensics 22 (1–2) (2021) 270–287.
[27] WHO, Guidelines for Drinking-Water Quality, fourth ed., 2021.
[28] H. Ghahremanzadeh, R. Noori, A. Baghvand, T. Nasrabadi, Evaluating the main sources of groundwater pollution in the southern Tehran aquifer using principal component factor analysis, Environ. Geochem. Health 40 (4) (2018) 1317–1328.
[29] Jackson Adiyiah Nyantakyi, Bernard Fei-Baffoe1, Osei Akoto, Seasonal variations in physicochemical and nutrient water quality of river tano in Ghana, Int. J. Environ. Chem. 4 (1) (2020) 1–12, https://doi.org/10.11648/j.ijec.20200401.11. http://www.sciencepublishinggroup.com/j/ijec.
[30] D. Pullanikkatil, L.G. Palamuleni, T.M. Ruhiiga, Impact of land use on water quality in the Likangala catchment, southern Malawi, Afr. J. Aqu. Sci. 40 (3) (2015) 277–286 (Printed in South Africa — All rights reserved).
[31] A. Umer, B. Assefa, J. Fito, Spatial and seasonal variation of lake water quality: beseka in the Rift Valley of Oromia region, Ethiopia, Int. J. Ener. Water Res. (2019), https://doi.org/10.1007/s42108-019-00050-8.
[32] J.Y. Magaji, P.E. Adakayi, Assessment of seasonal variation in surface water quality in and around Mpape Dumpsite, Federal Capital Territory, Am. J. Clim. Stud. ISSN 2520-0471 (Online) 2 (Issue 1) (2021) 1–15, 2021.
[33] D. Meera, S. Kumar, M.G. Sherly, P. Anand, A study on hydro chemical characteristics of fresh water lentic ecosystems in chavara industrial area-south west coast of India, Int. J. Sci. Res. Pub. 5 (12) (2015) 87–093 2019. Issue 2.
[34] B.A. Anhwange, E.B. Agbaji, E. C Gimba, Impact Assessment of Human Activities and Seasonal Variation on River Benue, within Makurdi Metropolis International Journal of Science and Technology, 2012. ISSN 2224-3577.
[35] S.J. Cronin, V.E. Neall, J.A. Lecointre, M.J. Hedley, P. Loganathan, Environmental hazards of fluoride in volcanic ash: A case study from Ruapehu volcano, New Zealand, Journal of Volcanology and Geothermal Research 121 (3-4) (2003 March) 271–291, https://doi.org/10.1016/S0377-0273(02)00465-1.
[36] V. Saxena, S. Ahmed, Inferring chemical parameters for the dissolution of fluoride in groundwater, Environ. Geol.Japan Environ. Modell. Software 43 (6) (2003) 731–736.
[37] R. Tekle-Haimanot, Z. Melaku, H. Kloos, C. Reimann, W. Fantaye, L. Zerihun, K. Bjorvatn, The geographic distribution of fluoride in surface and groundwater in Ethiopia with an emphasis on the Rift Valley, Sci. Total Environ. 367 (2006) 182–190, https://doi.org/10.1016/j.scitotenv.2005.11.003.
[38] U. Nsikak, A. Benson, E. Usoro, M. Etesin, Metal Contamination of Surface Water, Sediment and Tympanotonus Fuscatus Var Radula of Iko River and Environmental Impact Due to Utapete Gas Flare Station, Nigeria, Springer Science, 2007. Published online: 2 September 2007.