Journal of Environmental Statistics

September 2018, Volume 8, Issue 5. http://www.jenvstat.org

Statistical Models for Evaluating Water Pollution: The Case of Asejire and Eleyele Reservoirs in Nigeria

K.O.Obisesan
Department of Statistics
University of Ibadan, Nigeria
ko.obisesan@ui.edu.ng

P. Christopher
African Institute for Mathematical Sciences (AIMS)
Tanzania
privatus@aims.ac.tz

Abstract

Water pollution is a major environmental problem due to rapid population growth that over exploit and pollute the water resources. In this work the physico-chemical study of Asejire and Eleyele reservoirs are carried out to examine the water pollution levels. Eleyele and Asejire reservoirs are the two major sources of pipe-borne water in Ibadan with a population of about four million people. Water samples were collected from both sites from January 2003-December 2007 and analysed for 13 physico-chemical parameters. The data were subjected to Principal Component Analysis (PCA) to define the parameters responsible for the main variability in water pollution. The PCA produces 5 significant main components explaining 66.6% and 69.8% variance in Asejire and Eleyele reservoir, respectively. Generalized Linear Model (GLM) is applied to study the variability in turbidity level which shows that four parameters in each reservoir are important to explain the turbidity variation. Also many parameters in Asejire lie within the SON and WHO permissible limits while in Eleyele reservoir many parameters lie out. This therefore is an indication that water in Eleyele reservoir is more polluted than in Asejire reservoir.

Keywords: Principal Component Analysis, Generalized Linear Model, Water pollution, Water Quality.

1. Introduction

Water is an essential need for all living organisms. However this valued resource is increasingly being threatened as its demand increasing due to human population growth, Urbanization, industrialization and anthropogenic activities. The biggest water threat is water pollution which is the serious and growing problem (Kolawole, Ajayi, Olayemi, and Okoh 2011). Water pollution is observed at two levels: surface and ground. The contamination is caused by leaching from waste dumps, agricultural chemicals and industrial wastes (Awoyemi, Achudume, and Okoya 2014). The composition of both surface and ground water depends on the natural factors such as topographical, meteorological, hydrological and biological in the drainage basin which vary depending on weather conditions, seasonal variation in the run off volume and water levels (Mueller, Newton, Holly, and Preston-Martin 2001).

The most important fresh water resources are rivers (Kolawole *et al.* 2011). Rivers play a basic role in assimilating the urban waste water, industrial wastes and surface run off from agricultural fields (Basu and Lokesh 2014). Human being and other living organisms depend on water for their survival. Therefore consumption of polluted water put human health and aquatic organisms in most countries at risk (Kolawole *et al.* 2011). According to Zeman, Kross, and Vlad (2002) ground water is less susceptible to pollution and contamination than surface water bodies.

Water quality of different water resources is subjected to ongoing consequences of water pollution and this results in the increase in demand for monitoring its quality. Any water pollution may results in a problem of survival for living organisms and we can assess the water quality by studying its physico-chemical and microbiological parameters. In irrigation area, the quality of water is an important in assessing the salinity and alkalinity conditions (Gupta, Choudhary, and Vishwakarma 2009).

Before using water for any domestic, agricultural or industrial purpose it is very essential and important to test it (Tiwari 2015). Water quality testing is an essential part of environmental monitoring. The aquatic life as well as surrounding ecosystem is affected when water quality is poor [1]. Water quality must be tested with different physico-chemical parameters and the selection of parameter for testing should depend on the water uses, quality and purity Tiwari (2015). Water quality Monitoring enables managers to maintain good water quality. By doing so, managers make appropriate decision and take prior actions to ecosystem degrading (Damanik-Ambarita, Everaert, Forio, Nguyen, Lock, Musonge, Suhareva, Dominguez-Granda, Bennetsen, Boets *et al.* 2016). Monitoring help researchers to predict, learn and use information from natural processes in the environment and determine the effects of human activities in an ecosystem.

Water quality parameters provide the information on what is going on at the site. These measurement efforts can also assist in restoration projects or ensure environmental standards are being met. Monitoring of water quality involves different tests such as physical and chemical conditions, sediments and the biological composition of aquatic system Damanik-Ambarita *et al.* (2016). Physical test is applicable for physical appearance such as color, temperature, pH, turbidity, Total Suspended Solids (TSS) and Total Dissolved Solids (TDS). Chemical test is performed for Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), alkalinity, dissolved oxygen, and hardness. According to Tiwari (2015), only developed countries managed to monitor these criteria due to the availability of sophisticated analytical instruments, technology and trained manpower.

In recent years, many studies have been done using different multivariate statistical techniques such as Principal Component Analysis (PCA), Cluster Analysis (CA), analysis of variance (ANOVA), discriminant analysis (DA), factor analysis (FA) and multiple regression analy-

[1] http://www.fondriest.com/environmental-measurements/parameters/water-quality/

sis (MRA) in analysing and interpreting water pollution level. These Includes the works of Mustapha and Abdu (2012) used PCA and MRA to assess the surface water quality in Jakara basin, Gajbhiye, Sharma, and Awasthi (2015) used PCA for interpretation and grouping of water quality parameters in Jabalpur city using Moti Nala and Urdana Nala water samples, Kolawole *et al.* (2011) on Asa river, Obisesan, Lawal, and Adelakun (2013) on Eleyele and Asejire in Oyo state, Basu and Lokesh (2014) employed MRA and MANOVA on Srirangapatna, Koklu, Sengorur, and Topal (2010) employed the PCA, FA, MRA, and DA on Melen river system (Turkey), The other study was done by Awoyemi *et al.* (2014) who used ANOVA on Majidun-Ilaje area of Ikorodu, Lagos state.

This work was therefore carried out to investigate the application of Principal Component Analysis (PCA) on assessing and analysing the variables that contributes to water pollution, Generalized Linear Model (GLM) on measuring the level of turbidity of water resources and to study the impact of the discussed variables on leading case of water pollution on Eleyele and Asejire reservoirs and make appropriate recommendation.

2. Material and Methods

2.1. Study Area

The study was conducted in Eleyele and Asejire reservoirs in Ibadan Oyo State Nigeria. Ibadan is the capital city of Oyo state in Nigeria. Eleyele reservoir is the second largest reservoir in Oyo State located at latitude 07^0 25'00"N - 07^0 27'00"N, longitude 03^0 50'00"E - 03^0 53'00"E and elevation between 190 m along the river channel and 230 m on the surrounding slopes. This reservoir is used as the source of fishery development, flood control and domestic water supply. The sources of pollution in Eleyele are domestic wastes, industrial effluent, agricultural run-offs and bad fishing practices (Olanrewaju, Ajani, and Kareem 2017).

Asejire Reservoir is also in Oyo state on Osun river about 30 km East of Ibadan in the South West of Nigeria. It is located at latitude 07^0 23'35"N and longitude 04^0 08'14"E [2]. It maintains a constant depth of 81 meters throughout, even during the dry season.

2.2. Data Preparation

The data set of Asejire and Eleyele reservoirs were obtained from the water quality monitoring work conducted by water corporation of Oyo state Ibadan, Nigeria. The data comprising 13 physico- chemical parameters monitored monthly over 5 years from January 2003 to December 2007, includes Turbidity, Color, pH, Dissolved Oxygen, Alkalinity, Total Hardness, Calcium Hardness, Iron, Silica, Total Solids, Dissolved Solids and Total Suspended Solids and these determine the impact of pollution with respect to water pollution. The government was able to selected 13 Physico-chemical parameters at this time. The government has a laboratory close to Eleyele reservoir. They make sure that samples were collected and immediately analysed in the laboratory. All water samples from both reservoirs were collected, preserved and stored for analysis as outlined in standard methods for examination of water quality and wastewater Gajbhiye *et al.* (2015).

[2] http://wikimapia.org/22828441/Asejire-Reservoir

2.3. Principal Component Analysis (PCA)

Principal Component Analysis is one of the oldest and widely technique used for multivariate analysis. In field some data are collected from the single population on the large number of variables. The fundamental idea of PCA is to describe the variation of a set of of uncorrelated variables in which each is a particular linear combination of the original variable (Everitt and Dunn 2001). By using PCA the decreasing order of importance derive a new variables, this means that the variation in the original is much explained by the first principal component (PC1). The second principal component (PC2) accounts for the remaining variation which is uncorrelated with the PC1, and this procedure continues. The Principal components can be expressed using the following equation:

$$Z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + \cdots + a_{ip}x_{pj} \tag{1}$$

Where Z is the component score, a is the component loading, x is the measured value of a variable, i is the component number, j is the sample number, and p is the total number of variables.

The main objective of using the PCA is to see if the first few components can explain better the variation in the original data. This reduces the number of variables to be few, more meaningful and interpretable linear combination of the data, in which each linear combination will correspond to a principal component (PC).

2.4. Generalized Linear Model (GLM)

GLM extend the concept of ordinary regression model. It is specified by three components, a *random component* identifies the probability distribution of a response variable Y. The *systematic component* specifies the explanatory variables $(X_1, X_2, \ldots, X_n)$ used in linear predictor function, and a *link function* which specify the link between random and systematic component. The random component of GLM consists of a response variable Y with independent observations $(y_1, y_2, \ldots, y_n)$ of the response variable Y from a distribution in the natural exponential family with the probability density function or mass function of the form

$$f(y_i; \theta_i) = a(\theta_i)b(y_i)\exp\left[y_i Q(\theta_i)\right] \tag{2}$$

where $Q(\theta_i)$ is the natural parameter. The systematic component relates a vector $(\eta_i, \ldots, \eta_n)$ to the covariates variables through linear model. Let x_{ij} represent the value of covariates j $(j = 1, \ldots, p)$ for subject i then

$$\eta_i = \sum_{j=1}^{n} \beta_j x_{ij}, \qquad i = 1, \ldots, n. \tag{3}$$

The linear combination of covariates is called linear predictors. The link function connects the random and systematic components. Let $\mu_i = \mathrm{E}(Y_i),\ i, \ldots, n$. The model link μ_i to η_i by $\eta_i = g(\mu_i)$, thus g link $\mathrm{E}(Y_i)$ to the covariates variables through the formula

$$g(\eta_i) = \sum_{j=1}^{n} \beta_j x_{ij}, \qquad i = 1, \ldots, n. \tag{4}$$

.

2.5. Randomization

Randomization is one of the resampling techniques. It uses permutation tests as a way to calculate distribution free p-values for any dataset under the study McCue, Carruthers, Dawe, Liu, Robar, and Johnson (2008). Randomization effectively removes the homogeneity, normality, independence of residual and dispersion assumptions. When the assumptions are reasonable, the randomization p-values should be equal to the model p-values. The randomization is applied for the model with the fixed covariates.

3. Results and Discussion

3.1. Data Description and Visualization

The minimum (Min), maximum (Max), mean, standard deviation (S.D) and median (Med) values of each Water parameter (Par) (Table 1) are presented in Table 2 for both Asejire and Eleyele reservoir.

Table 1: Definition of water quality parameters

Abbreviation	Parameter Name	Units of Measurement
Tur	Turbidity	Nephelometric Turbidity Units (NTU)
Col	Color	Hazen Units (HU)
PH	pH	Logarithmic Units (LU)
DO	Dissolved Oxygen	Milligram per litre (mg/L)
Alk	Alkalinity	Milligram per litre (mg/L)
TH	Total Hardness	Milligram per litre (mg/L)
CaH	Calcium Hardness	Milligram per litre (mg/L)
Cl	Chloride	Milligram per litre (mg/L)
Fe	Iron	Milligram per litre (mg/L)
Si	Silica	Milligram per litre (mg/L)
Sol	Total Solids	Milligram per litre (mg/L)
DS	Dissolved Solids	Milligram per litre (mg/L)
SS	Total Suspended Solids	Milligram per litre (mg/L)

Table 2: Descriptive Statistics of the Parameters under study

	Asejire Reservoir				Eleyele Reservoir				Limits	
Par	Min	Max	Mean±S.D	Med	Min	Max	Mean±S.D	Med	SON	WHO
Tur	1	4	2.68 ± 0.81	3	2	32	8.24 ± 5.61	8.1	5	5
Col	4	7	5.03 ± 0.41	5	5	25	7.62 ± 4.17	6	15	15
PH	6.4	8.8	7.49 ± 0.54	7.4	6.2	8	7.19 ± 0.39	7.3	6.5-8.5	-
DO	6.4	11	7.6 ± 0.87	7.5	5.5	54	11.2 ± 9.7	10.75	-	500
Alk	22	78	38.68 ± 9.37	38	8	100	48.37 ± 17.14	47.5	-	500
TH	36	88	57.03 ± 10.35	58	64	112	90.5 ± 10.82	92	150	500
CaH	9	70	38.7 ± 10.62	39.5	34	100	65.47 ± 12.06	64	-	-
Cl	10.4	35	19.45 ± 5.59	18	23.5	56	36.22 ± 5.81	35.4	250	250
Fe	0.21	0.28	0.24 ± 0.02	0.25	2	2.6	2.33 ± 0.18	2.3	0.3	0.3
Si	4	14	10.33 ± 2.17	10	4	17	12.88 ± 2.59	14	-	-
Sol	34	217	138.87 ± 25.95	138	178	365	246.77 ± 28.26	244.5	-	-
DS	48	190	109.5 ± 2.61	112.5	138	245	173.38 ± 14.77	173	500	500
SS	12	86	39.75 ± 14.76	35.5	36	92	69.92 ± 10.93	72.5	-	-

Asejire reservoir : From Table 2, it is clear that Sol and DS are dominant parameter with high mean concentration of 138.87 mg/L and 109.50 mg/L respectively. This show that, these variables have a common source of origin (Mustapha and Abdu 2012). The average value of PH is 7.49 LU which is slightly above neutral level. The average concentration of TH, Alk, CaH, and DO are 57.03, 38.68, 38.7 and 7.6 mg/L respectively.

Eleyele reservoir: From Table 2 we observe that in Eleyele reservoir Sol, DS and TH are dominant parameters with the high mean concentration of 246.77 mg/L, 173.38 mg/L and 90.50 mg/L respectively. This also show that these variables have the common anthropogenic source of origin (Mustapha and Abdu 2012; Awoyemi *et al.* 2014). The PH range from 6.20 LU to 8.0 LU with average value of 7.19 which is slightly above neutral level. The average concentration of SS, Alk, CaH and DO are 69.92, 48.37, 65.47 and 11.2 mg/L respectively.

Thus Sol and DS are the dominant parameters with high mean concentration in both Asejire and Eleyele reservoir.

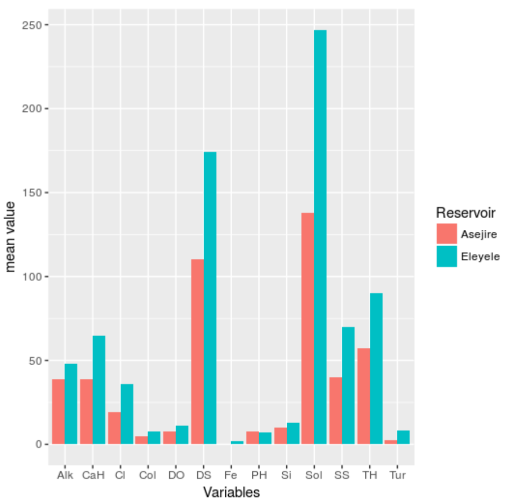

Figure 1: Bar plot for mean value of variables for Asejire and Eleyele reservoirs.

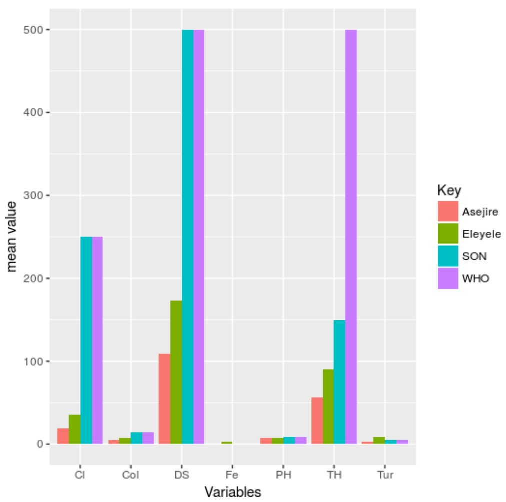

Figure 2: Bar plot for variable comparison with SON and WHO permissible limits.

Figure 1 shows that the average concentration of variables in Eleyele reseirvoir are higher than that of Asejire reservoir. Table 2 and Figure 2 show that the concentration of Tur, Col and Fe in Eleyele reservoir are greater than SON and WHO permissible limits. The concentrations of Tur, Col, Fe, DS and Cl in Asejire reservoir are within SON and WHO prescribed limits.

3.2. Compositional Relation

The correlation between parameters can give more insight on the relationship between different variables as shown in Figure 3 and 4.

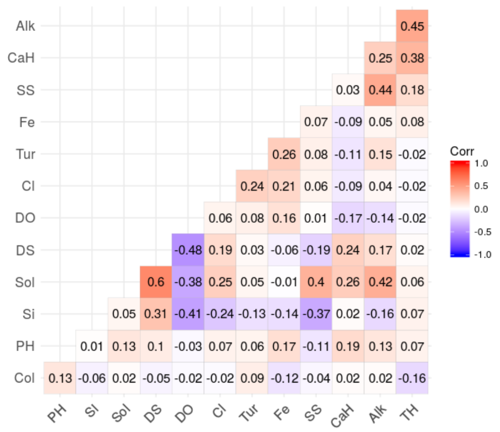

Figure 3: Asejire correlation matrix

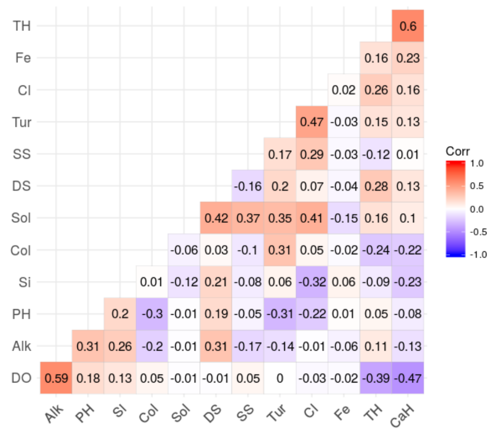

Figure 4: Eleyele correlation matrix

Asejire reservoir: The results of correlation coefficient matrix between physico- chemical parameter from Figure 3 shows that the very weak positive correlation was observed between Col and PH (0.13), Si and PH (0.01), Sol and Col (0.02), PH and Sol (0.13),Sol and Si (0.05), DS and PH (0.1), DS and Si (0.31), Cl and PH (0.07), DS and Cl (0.19), Tur and Sol (0.05), Tur and DS (0.03), DO and Tur (0.08), Fe and PH (0.17), Fe and DO (0.16), Fe and Tur (0.26), SS and DO (0.01), Tur and SS (0.08), CaH and Si (0.02), CaH and DS (0.24), CaH and SS (0.03), Alk and Col (0.02), PH and Alk (0.13), DS and Alk (0.17), TH and PH (0.07), SS and TH (0.18), TH and Fe (0.08).

The negative weak association was observed between Col and Si (−0.06), DS and Col (−0.05), DO and Col (−0.02), DO and PH (−0.03), Cl and Col (−0.02), Tur and Si (−0.13), Fe and Col (−0.12), Fe and DS (−0.06), SS and PH (−0.11), CaH and DO (−0.17), Tur and CaH (−0.11), Alk and DO (−0.14), Tur and TH (−0.02), Cl and TH (−0.02). The positive moderate correlation exists between DS and Sol (0.6) which indicate that they have the same source of origin.

Eleyele reservoir: The results of correlation coefficient matrix between physico- chemical parameter from Figure 4 shows that the positive and very weak correlation was observed between PH and DO (0.18), Alk and PH (0.31), DO and Si (0.13), Alk and Si (0.26), PH and Si (0.2), DO and Col (0.05), Si and Col (0.01), DS and Alk (0.31), PH and DS (0.19), DS and Si (0.21), Col and DS (0.03), Sol and DS (0.42), SS and DO (0.05), SS and Sol (0.37), Cl and Col (0.05), Cl and Sol (0.42), TH and DS (0.28), CaH and Sol (0.1), Fe and CaH (0.23).

The negative weak association was observed between Col and Alk (−0.2), PH and Col (−0.3), Col and Sol (−0.06), Sil and Sol (−0.12), PH and Sol (−0.01), Alk and Sol (−0.01), DS and

DO (-0.01), PH and SS (-0.05), SS and Col (-0.1), Alk and Tur (-0.14), Cl and Alk (-0.01), Cl and DO (-0.03), Fe and DO (-0.02), Fe and Alk (-0.06), Fe and Col (-0.02), TH and DO (-0.39), Si and TH (-0.09), CaH and Alk (-0.13), Si and CaH and (-0.23), Col and CaH (-0.22).

The moderate positive correlation was observed between Alk and DO (0.56), CaH and TH (0.6) which indicate that they have the same source of origin. There is no correlation between DO and Tur which indicates that they have different source of origin. At this stage, it is difficult to group parameters into components and assign any physical significance. Hence, in the next step, the PCA has been applied. The correlation matrix is subjected to the PCA.

3.3. Water Pollution Level using Principal Component Analysis (PCA)

In this study PCA was applied to Asejire and Eleyele reservoir dataset each with $n \times p$ matrix, where $n = 60$ is the number of months from January 2003 to December 2007 and $p = 13$ is the number of variables (water quality parameters). This resulted to 60×13 matrix.

The number of selected PCs were selected by keeping the first few PCs that account for the most variation in data according to the following criteria. The first criterion was to select the PCs with eigenvalue which is greater than one (Kaiser's rule). The second was to determine the number of PCs required to explain the variation in each data set based on kaiser's rule. The PCs with the eigenvalues less than one were discarded and only the PCs with eigenvalue greater than on were retained. Also the interaction between parameters, months and components were inspected by using a biplot in order to visualize the sign and magnitude of each variable's contribution to the particular principal component. The main purpose of PCA is to reduce the contribution of the less significant variables to simplify more information coming from PCA. This is achieved by rotating the axis defined by PCA according to well established rules. The largest loading suggests the meaning of the dimension. The positive loading indicates that the variable contribution increases with increase in dimension and negative loading indicates the decrease. Tables 4 and 5 show the rotated component matrix, eigenvalue of each PC, percent and cumulative percent of the variance of Asejire and Eleyele respectively.

Table 3: Eigenvalues, Percentage Variance and Percentage Cummulative Variance for Eleyele and Asejire reservoir PCs.

	Eleyele Reservoir			Asejire reservoir		
PCs	Eigenvalue	% Variance	% Cum.variance	Eigenvalue	% Variance	% Cum.variance
PC1	2.6342	20.2631	20.2631	2.5626	19.7129	19.7129
PC2	2.0284	15.6035	35.8667	2.0931	16.1009	35.8139
PC3	1.9485	14.9888	50.8556	1.5479	11.9072	47.7212
PC4	1.3846	10.6508	61.5063	1.2717	9.7825	57.5037
PC5	1.0845	8.3423	69.8486	1.1769	9.0537	66.5574
PC6	0.9963	7.6641	77.5127	0.9124	7.0186	73.5761
PC7	0.7201	5.5390	83.0517	0.7925	6.0965	79.6726
PC8	0.5682	4.3705	87.4223	0.6272	4.8251	84.4977
PC9	0.4831	3.7162	91.1385	0.6127	4.7135	89.2112
PC10	0.4235	3.2583	94.3967	0.5035	3.8731	93.0842
PC11	0.3196	2.4591	96.8558	0.4544	3.4954	96.5797
PC12	0.2616	2.0127	98.8686	0.2846	2.1896	98.7694
PC13	0.1471	1.1313	100.000	0.1599	1.2306	100.000

Table 4: Rotated Component Matrix of Physico-Chemical Data (Asejire reservoir)

Parameter	PC1	PC2	PC3	PC4	PC5
Tur	-0.0370	0.2865	-0.3785	0.1145	-0.0247
Col	0.0134	-0.0101	-0.1235	-0.0645	-0.7957
PH	-0.1518	0.0351	-0.1460	0.55743	-0.3747
DO	0.3560	0.3448	0.0860	0.1648	-0.0223
Alk	-0.4088	0.2959	0.2116	-0.0153	-0.0716
TH	-0.2602	0.1480	0.4336	0.3637	0.2055
CaH	-0.3321	-0.0528	0.3069	0.2903	-0.1736
Cl	-0.1075	0.2557	-0.4849	-0.0053	0.1487
Fe	0.0122	0.3064	-0.2663	0.4007	0.2803
Si	-0.1061	-0.5136	-0.0165	0.1512	0.1828
Sol	-0.5099	0.0491	-0.1813	-0.2649	-0.0192
DS	-0.4268	-0.2680	-0.3165	-0.0228	0.1148
SS	-0.1957	0.4334	0.2151	-0.4169	0.0072
Eigenvalue	2.5627	2.0931	1.5479	1.2717	1.1770
% of variance component	19.7129	16.1009	11.9072	9.7825	9.0538
Cumulative % of variance	19.7129	35.8139	47.7112	57.5037	66.5575

Table 5: Rotated Component Matrix of Physico-Chemical Data (Eleyele reservoir)

Parameter	PC1	PC2	PC3	PC4	PC5
Tur	-0.3331	0.0532	-0.3732	0.3129	-0.1198
Col	-0.0151	0.3386	-0.2232	0.5262	-0.0298
PH	0.2382	-0.3966	0.0495	-0.2315	0.1291
DO	0.3421	-0.0807	-0.4188	-0.1203	-0.3991
Alk	0.2555	-0.4556	-0.2217	-0.0091	-0.2733
TH	-0.3505	-0.3958	0.2310	0.0947	-0.0853
CaH	-0.3971	-0.2137	0.3358	0.0136	-0.1016
Cl	-0.3972	-0.0067	-0.3004	-0.1377	-0.2913
Fe	-0.0537	-0.0650	0.2297	0.1149	-0.6947
Si	0.2552	-0.2032	-0.0964	0.3415	0.0721
Sol	-0.3162	-0.1871	-0.3896	-0.1766	0.2667
DS	-0.1044	-0.4592	-0.2138	0.3268	0.2674
SS	-0.1801	0.1314	-0.2604	-0.5109	-0.0328
Eigenvalue	2.6342	2.0285	1.9485	1.3846	1.0845
% of variance component	20.2631	15.6035	14.9888	10.6508	8.3423
Cumulative % of variance	20.2631	35.8667	50.8555	61.5063	69.8486

Due to standardization, all principal components (PC) have mean zero, the standard deviation is also given for each of the components and it is the square root of eigenvalue. The purpose is to find the correlation between the principal components and the original variables. The first five PC's in each reservoir were retained for subsequent analysis (Table 3). While this results to neglection of some important information, the objective was to get insight for the majority of variation. According Deluzio, Wyss, Costigan, Sorbie, and Zee (1999) the smaller variance PC's are harder to interpret.

Water Pollution Level using PCA in Eleyele Reservoir.

Variables correlating with each of the five PCs are indicated in Table 6.

Table 6: Variables correlating with each of the five PCs in Eleyele dataset .

PC1	Cor	PC2	Cor	PC3	Cor	PC4	Cor	PC5	Cor
Cl	0.6447	DS	0.6541	DO	0.5846	Col	0.6191	Fe	0.7235
CaH	0.6445	Alk	0.6488	Sol	0.5438	Si	0.4019	DO	0.4157
TH	0.5889	PH	0.5649	Tur	0.5210	DS	0.3846	Cl	0.3033
Tur	0.5407	TH	0.5638	Cl	0.4194	Tur	0.3681	Alk	0.2846
Sol	0.5131	CaH	0.3045	SS	0.3635	PH	-0.2724	Sol	-0.2778
SS	0.2923	Si	0.2894	Col	0.3116	SS	-0.6012	DS	-0.2785
PH	-0.3866	Sol	0.2665	Alk	0.3095				
Si	-0.4142	Col	-0.4822	DS	0.2985				
Alk	-0.4148			Fe	-0.3207				
DO	-0.5553			TH	-0.3225				
				CaH	-0.4688				

Table 6 show that Cl, CaH, TH, Tur and Sol were positively highly correlated with PC1 while DO was negatively highly correlated with PC1. DS, Alk, PH and TH were positively highly correlated with PC2 while Col was negatively highly correlated with PC2. One can observe that DO, Sol and Tur were positively highly correlated with PC3 while CaH was negatively correlating with PC3. We can observe that Col was highly positively correlated with PC4, and SS was highly negatively correlated with PC4. Furthemore one can observe that Fe was highly positively correlated with PC5, and Sol was high negatively correlated with PC5.

Water Pollution Level using PCA in Asejire Reservoir.

Variables correlating with each of the five PCs are indicated in Table 7.

Table 7: Variables correlating with each of the five PCs in Asejire dataset.

PC1	Cor	PC2	Cor	PC3	Cor	PC4	Cor	PC5	Cor
Sol	0.8163	SS	0.6271	Cl	0.6033	PH	0.6286	Col	0.8633
DS	0.6833	DO	0.4989	Tur	0.4709	Fe	0.4519	PH	0.4065
Alk	0.6545	Fe	0.4434	DS	0.3938	TH	0.4101	Fe	-0.3041
CaH	0.5317	Alk	0.4282	Fe	0.3313	CaH	0.3275		
TH	0.4165	Tur	0.4145	Alk	-0.2633	Sol	-0.2987		
SS	0.3133	Cl	0.3700	SS	-0.2676	SS	-0.4702		
DO	-0.5700	DS	-0.3879	CaH	-0.3818				
		Si	-0.7432	TH	-0.5396				

Table 7 show that Sol, DS and Alk are positively highly correlated with PC1 while DO was negatively moderate correlated with PC1. SS was positively high correlated with PC2 and Si was negatively high correlated with PC2.Also we observe that Cl was positively high correlated with PC3 while TH was negatively correlated with PC3. One can observe that PH was highly positively correlated with PC4, and SS was moderate negatively correlated with PC4. Furthermore one can observe that Col was high positively correlated with PC5, and Fe was low negatively correlated with PC5.

3.4. Water Turbidity level using the Generalized Linear Model Model: Eleyele reservoir

To get insight on the distribution function of turbidity, the Cullen and Frey graph were used. The kurtosis and squared skewness of Turbidity sample were plotted. From Cullen and Frey graph the possible distribution were determined. Based on goodness of fit plots and AIC the best distribution was selected. To find out the best covariates of turbidity variation in both reservoirs, a step wise regression was used. The best predictors were selected and used to fit the model. The datasets were analysed using both GLM and randomizations.

Water Turbidity level using GLM for Eleyele reservoir

From Cullen and Frey graph the possible distribution for Turbidity in Eleyele reservoir includes Weibull, Log normal or gamma. Based on goodness of fit plots and AIC the gamma distribution was selected since its fit better. To find out the best covariates of turbidity variation in Asejire reservoir, a step wise regression was used. The variables Cl, Col, Si and Sol were selected as the best predictors and used to fit the model. The data set was analysed using both GLM and randomizations. The gamma distribution with the inverse link function were used to run the analysis. The results obtained are shown in Table 8.

Table 8: The GLM and Randomisation results for Eleyele reservoir

Coeff	Estimates	95%CI	std.Error	t-value	GPV	RPV	SBT
Intercept	0.5095	(0.3421, 0.6796)	0.0857	5.948	0.0000		
Cl	-0.0036	(-0.0053, -0.0015)	0.0009	-3.826	0.0003	0.0022	-0.003770928
Col	-0.0044	(-0.0068, -0.0014)	0.0014	-3.174	0.0025	0.0115	-0.003247818
Si	-0.0078	(-0.0148, -0.0013)	0.0034	-2.258	0.0279	0.0277	-0.003609548
Sol	-0.0004	(-0.0008, 0.0000)	0.0002	-1.991	0.0514	0.0855	-0.002212920

Coeff=Coefficient, GAPV= Gamma p-value, RPV= Randomisation p-value, SBT=Standardized beta coefficient.

After running the analysis, the residual plots were used to examine the homogeneity, normality and independence assumptions respectively and show that all were satisfied. The slope of the line was 1.1 which indicate that the link function was appropriate, the dispersion (0.028) indicates under-dispersion. The randomisation p-values were close to the model p-values. The fitted regression equation for turbidity level was given by

$$\frac{1}{\widehat{\text{Tur}}} = 0.5095 - 0.0036 \times \text{Cl} - 0.0044 \times \text{Col} - 0.0078 \times \text{Si} - 0.0004 \times \text{Sol}. \tag{5}$$

Using Equation 5, it is seen from the unstandardised coefficient that for every unit increase in Cl, 0.004 unit decrease in the inverse of Turbidity is predicted, holding all other variable constant. Also for every unit increase in Col, 0.004 units decrease in the inverse of turbidity level is predicted, holding all other variables constants. Similarly, for every unit increase in Si, 0.008 units decrease in the inverse of turbidity level is predicted, holding all other variables constants. Further for every unit increase in CaH, 0.0004 unit decrease in the inverse of turbidity level is predicted, holding all other variables constants.

However, the standardization of variables before running regression gives the actual interpretation where all variables are on the same scale, and it compares the magnitude of the coefficient to show which variable has more effect. The standardized beta coefficient revealed that Cl (0.0037) was the strongest unique contribution in turbidity. The beta coefficient value for Si (0.0036) was the second highest, followed by Col (0.0032). The least contributor was Sol (-0.0022).

From Table 8 the parameter Sol is not statistically significant at 95% confidence level since its p-values is greater than 0.05. However, Cl, Col and Si are significant since their p-values are less than 0.05. The intercept is significant at 95% confidence level. This means that Sol do not produce significant turbidity effects while Cl, Col and Si produce significant turbidity effects.

Water Turbidity level using the GLM for Asejire reservoir

The kurtosis and squared skewness of exponential transformed turbidity were obtained. From Cullen and Frey graph, the possible distribution includes normal, log normal and gamma. The goodness of fit plots were used to compare the empirical distribution and multiple parametric distributions fitted on the Asejire dataset. Based on AIC the normal distribution was selected since fittted better than others. To find out the best covariates of Turbidity variation, a step wise regression was used. The variables Col, Alk, Cl and Fe were selected as the best covariates and used to fit the model. The dataset was analysed using both GLM and randomizations. The Gaussian distribution with the identity link function were used to run the analysis. The results obtained are shown in Table 9.

Table 9: The GLM and Randomization results of Asejire reservoir

Coeff	Estimates	95%CI	Standard Error	t-value	GAPV	RPV	SBT
Intercept	-54.5069	(-113.1119, 4.0980)	29.9011	-1.823	0.0738		
Col	4.6187	(-2.2769, 11.5143)	3.5183	1.313	0.1947	0.3553	0.1656
Alk	0.0603	(-0.2406, 0. 3607)	0.1533	0.393	0.6956	0.8691	0.0493
Cl	0.4239	(-0.0906, 0.9384)	0.2625	1.615	0.1120	0.6453	0.2070
Fe	162.9664	(-16.2059, 342.1386)	91.4161	1.783	0.0802	0.1410	0.2302

Coeff=Coefficient, GAPV= Gaussian p-value, RPV= Randomisation p-value, SBT=Standardized beta coefficient.

After running the analysis the residual plots were used to examine the homogeneity, normality and independence assumptions. All assumptions were satisfied. The slope of the line was 1 which indicate that the link function was appropriate. The randomisation p-values disagree with the Gaussian p-values and the data has over dispersion. The fitted regression equation for turbidity level was given by

$$\widehat{\exp(\mathrm{Tur})} = -54.5069 + 4.6187 \times \mathrm{Col} + 0.0603 \times \mathrm{Alk} + 0.4239 \times \mathrm{Cl} + 162.9664 \times \mathrm{Fe}. \quad (6)$$

Using Equation 6, it is seen from the unstandardised coefficient that for every unit increase in Col, 4.6 unit increase in the exp(Tur) is predicted, holding all other variable constant. Also for every unit increase in Alk, 0.06 units increase in the exp(Tur) level is predicted, holding all other variables constants. Similarly for every unit increase in CaH and Cl, 0.4 and 163 units increase respectively in the exp(Tur) level is predicted, holding all other variables constants. Furthermore, for every unit increase in Fe, 163 unit increase in the exp(Tur) level is predicted, holding all other variables constant.

However, the standardization of variables before running regression gives the actual interpretation where all variables are on the same scale, and it compare the magnitude of the coefficient to show which variable has more effect. The standardized beta coefficient revealed that Fe (0.2302) was the strongest unique contribution in exp(turbidity). The beta coefficient value for Cl (0.2070) was the second highest, followed by Cl (0.1656). The least contributor was Alk (0.0493).

4. Conclusions and Recommendation

4.1. Conclusion

The primary purpose of this study was to investigate the water pollution level and whether PCA and GLM techniques could be useful to identify the water pollution level and turbidity level respectively. According to established criteria, the PCs that results from PCA were considered for analysis when the eigenvalues were greater than 1. Results show that, out of the 13 PCs selected, only 5 were finally considered to efficiently explain the majority of variance in data by 66.6% and 69.8% for Asejire and Eleyele reservoir respectively. From the PCA results, its clear that Cl, CaH, DS, DO, Col and Fe were found to be the most abundance parameters responsible for water pollution in Eleyele reservoir. And Sol, SS, Cl, PH and Col were found to be the important parameters responsible for water pollution in Asejire reservoir.

GLM results show that eight physico chemical parameters (DO, Alk, TH, SS, Fe, PH, CaH and DS) are less important in explaining turbidity variation in Eleyele reservoir. Also water parameters (PH, DS, Si, Sol, TH, CaH, DS and SS) are less important in explaining turbidity variation in Asejire reservoir. The results from the concentration of water sample in Table 2, Figure 1 and Figure 2 show that Eleyele reservoir is more polluted than Asejire reservoir. Generally all reservoirs are polluted and this may be due to anthropogenic and industrial activities such as intensive agricultural activities, livestock wastes, domestic wastes, organic wastes, inorganic wastes and industrial area near the river channel.

4.2. Recommendation

It is recommended to have a coordinate effort which involves both community and government stakeholders in preventing and controlling water pollution. From the results it is important to improve turbidity monitoring efficiency network in Eleyele and Asejire reservoirs by reducing the number of physico-chemical monitoring parameters of water equality from 12 to 4. This reduction reduces monitoring cost without losing important water quality parameters which explain turbidity level variation in water. We state some interesting topics which are worthy of investigation for further studies:

1. The influence of agricultural chemicals, oil spills, organic wastes and poor sewage disposal on water bodies pollution should be examined.
2. Investigation of temporal variation and seasonal effect on water quality parameters: A case study of Eleyele and Asejire reservoir.
3. To study more water chemistry in gaining deeper understanding of the microbiological impacts as related to public health.

References

Awoyemi OM, Achudume AC, Okoya AA (2014). "The Physicochemical Quality of Groundwater in Relation to Surface Water Pollution in Majidun Area of Ikorodu, Lagos State, Nigeria." *American Journal of Water Resources*, **2**(5), 126–133.

Basu S, Lokesh K (2014). "Application of Multiple Linear Regression and Manova to Evaluate Health Impacts Due to Changing River Water Quality." *Applied Mathematics*, **5**(05), 799.

Damanik-Ambarita MN, Everaert G, Forio MAE, Nguyen THT, Lock K, Musonge PLS, Suhareva N, Dominguez-Granda L, Bennetsen E, Boets P, *et al.* (2016). "Generalized Linear Models to Identify Key Hydromorphological and Chemical Variables Determining the Occurrence of Macroinvertebrates in the Guayas River Basin (Ecuador)." *Water*, **8**(7), 297.

Deluzio KJ, Wyss UP, Costigan PA, Sorbie C, Zee B (1999). "Gait assessment in unicompartmental knee arthroplasty patients: Principal component modelling of gait waveforms and clinical status." *Human Movement Science*, **18**(5), 701–711.

Everitt BS, Dunn G (2001). *Applied multivariate data analysis.2nd ed.* Oxford University Press.

Gajbhiye S, Sharma S, Awasthi M (2015). "Application of principal components analysis for interpretation and grouping of water quality parameters." *International Journal of Hybrid Information Technology*, **8**(4), 89–96.

Gupta P, Choudhary R, Vishwakarma M (2009). "Assessment of water quality of Kerwa and Kaliasote rivers at Bhopal district for irrigation purpose." *International Journal of Theoretical & Applied Sciences*, **1**(2), 27–30.

Koklu R, Sengorur B, Topal B (2010). "Water quality assessment using multivariate statistical methodsâĂŤa case study: Melen River System (Turkey)." *Water resources management*, **24**(5), 959–978.

Kolawole OM, Ajayi KT, Olayemi AB, Okoh AI (2011). "Assessment of water quality in Asa River (Nigeria) and its indigenous Clarias gariepinus fish." *International journal of environmental research and public health*, **8**(11), 4332–4352.

McCue T, Carruthers E, Dawe J, Liu S, Robar A, Johnson K (2008). "Evaluation of generalized linear model assumptions using randomization." *Unpublished manuscript. Retrieved fro m http://www. mun. ca/biology/dschneider/b7932/B7932Final10Dec2008. pdf.*

Mueller BA, Newton K, Holly EA, Preston-Martin S (2001). "Residential water source and the risk of childhood brain tumors." *Environmental Health Perspectives*, **109**(6), 551.

Mustapha A, Abdu A (2012). "Application of principal component analysis & multiple regression models in surface water quality assessment." *Journal of environment and earth science*, **2**(2), 16–23.

Obisesan O, Lawal BT, Adelakun A (2013). "Data Visualization and Change-point detection in Environmental Data:The Case of Water Pollution in Oyo State Nigeria." *Journal of Science Research*, **12**, 181–190.

Olanrewaju A, Ajani E, Kareem O (2017). "Physico-Chemical Status of Eleyele Reservoir, Ibadan, Nigeria." *Journal of Aquaculture Research & Development*, **8**(9).

Tiwari S (2015). "Water Quality Parameters–A Review." *International Journal of Engineering Science Invention Research and Development*, **1**, 319–324.

Zeman CL, Kross B, Vlad M (2002). "A nested case-control study of methemoglobinemia risk factors in children of Transylvania, Romania." *Environmental health perspectives*, **110**(8), 817.

Affiliation:

K.O.Obisesan
Department of Statistics
Faculty of Science
University of Ibadan
Nigeria
Telephone:+2348052416373 E-mail:ko.obisesan@ui.edu.ng

Privatus Christopher
African Institute for Mathematical Sciences
Bagamoyo, Pwani
Tanzania
Telephone:+255 752901052 E-mail:privatus@aims.ac.tz

Journal of Environmental Statistics http://www.jenvstat.org
Volume 8, Issue 5 Submitted: 2017-08-24
September 2018 Accepted: 2018-04-18