FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 6 No. 3, June, 2022, pp 15 - 35
DOI: https://doi.org/10.33003/fjs-2022-0603-967

WATER QUALITY PARAMETERS FOR POLLUTION SOURCE IDENTIFICATION IN WARWADE DAM WATER, DUTSE, JIGAWA STATE, NIGERIA: STATISTICAL AND SPATIOTEMPORAL EVALUATION

***[1]Uduma, Maria B., [2]Abubakar, M. M., [2]Muhammed S. I.**

[1]Department of Biology, Federal College of Education, Kano
[2]Department of Biological Sciences, Federal University Dutse, Jigawa State, Nigeria

*Corresponding authors' email: maria_uduma@yahoo.com

ABSTRACT

This study looked at the diminishing water quality of Warwade Dam, which is one of Dutse's largest dams. Statistical methodologies such as descriptive statistics, principal component analysis (PCA), regression analysis, correlation analysis, and cluster analysis were used to examine selected water quality characteristics. pH, conductivity, TDS, temperature, and DO captured the highest variability (98.94%) in the data set, according to the PCA scree plot. Ca and Mg were the most variable from the PCA scree plot among the metals studied (99.96 percent). The correlational analysis revealed that the characteristics differed in terms of spatio-temporal and/or limnological aspects. Variables within a cluster are relatively similar, whereas variables outside a cluster are very dissimilar, according to cluster analysis employing hierarchical dendrogram to establish linkages, relationships, and differences among parameters. The validation test was passed by 67 percent of the parameters in the general linear regression model (very strong fit at P value 0.05). The PCA scree plot of 99.39 percent variability revealed that the dam concentrations of nitrate, nitrite, and chloride were identical.

Keywords: Principal component, Cluster, Correlation, Descriptive statistics, Regression analyses, Warwade Dam, Spatio-Temporal

INTRODUCTION

Dam water is one of the most important available water supplies for human consumption, its quality is a key environmental problem (Simeonov *et al.,* 2004). Dams have always been extensively polluted throughout the history of human civilization, owing to their easy accessibility to garbage disposal and, in many cases, the lack of a regulatory framework. However, following the industrial revolution, the carrying capacity of water sources to treat wastes was drastically reduced (Mahvi *et al.,* 2005). Dam water quality is influenced by anthropogenic activities such as urbanization, industry, and agriculture, as well as natural processes such as precipitation inputs, erosion, and weathering of crustal elements (Jarvie *et al.,* 1998). Precipitation, surface runoff, interflow, ground water flow, and pumped in and out flows all have a significant impact on pollution concentrations in dams (Altaf *et al.,* 2013). The protection of these water resources has been given top priority in the twenty-first century, given the limited stock of dam water worldwide and the role that anthropogenic activities play in the deterioration of water quality (USEPA, 2007). According to Don-Pedro and Colleagues (2004) that the spatiotemporal study of water is a crucial stage in the protection and conservation of water resources. Because the nonlinear structure of environmental data makes it difficult to comprehend spatiotemporal variations in water quality, statistical methodologies are utilized to provide representative and reliable water quality analysis (Dixon *et al*., 1996).

In the examination of water quality data, multivariate statistical approaches such as cluster analysis (CA) and principal component analysis (PCA) have been widely employed as unbiased tools for obtaining relevant results (Singh et al., 200). It's also been frequently utilized to define and evaluate water quality, as well as to investigate spatiotemporal fluctuations induced by natural and anthropogenic processes (Helena *et al.,* 2000). CA and PCA look for groups and sets of variables that have comparable properties, which could help us simplify our observations. Cluster analysis is a multivariate statistical technique that allows you to put things together based on how similar they are. CA groups items into clusters according on how similar they are to each other in terms of a predefined selection criterion.

The most frequent approach to CA is Bray-Curtis cluster analysis, which gives intuitive similarity correlations between any one sample and the entire dataset and is often represented by a dendrogram (a tree diagram). The dendrogram is a visual representation of the clustering process, displaying a picture of the groups and their proximity while dramatically reducing the original data's dimensionality. Principal Component Analysis is used to minimize the dimensionality of a data set with a large number of associated variables by transforming the data set into a new set of variables known as the principal components (PCs). The PCA is a data reduction approach that determines how many variations are required to explain the data's apparent variance.

PCs are calculated using eigenvalues and eigenvectors from covariance or other cross-product matrices, which characterize the dispersion of many observed parameters. Furthermore, the linear combinations of the original variables and the eigenvectors are included. PCA can be used to minimize the number of variables while still explaining the same amount of variance (principal components). PCA also tries to explain the association between data in terms of underlying characteristics that aren't readily visible. We provide an approach for using CA and PCA to examine the influence of all sources of contamination in Warwade dam water and to determine the characteristics responsible for spatiotemporal variability in water quality.

The Warwade village relies on the dam for its survival, as it provides water for both home and agricultural uses. As a result, the current study is a step forward in addressing the dam's worsening circumstances and recommending specific solutions for its long-term maintenance.

MATERIALS AND METHODS

Study Area

The Warwade Dam is located near Dutse in the Nigerian State of Jigawa (Figure 1). The dam, which is located at 11045'N and 9013'E, is 1.4 kilometres long, 6 meters wide, and 7 meters deep, with a total storage capacity of 300 million cubic meters. The dam was inaugurated in May 1977 by late Audu Bako, the military sole administrator of old Kano State (FMWR, 2018).

The dam offers water for a variety of domestic purposes, including irrigation, aquaculture (fish production), recreation, and livestock, as well as enhanced fishing in the vicinity.

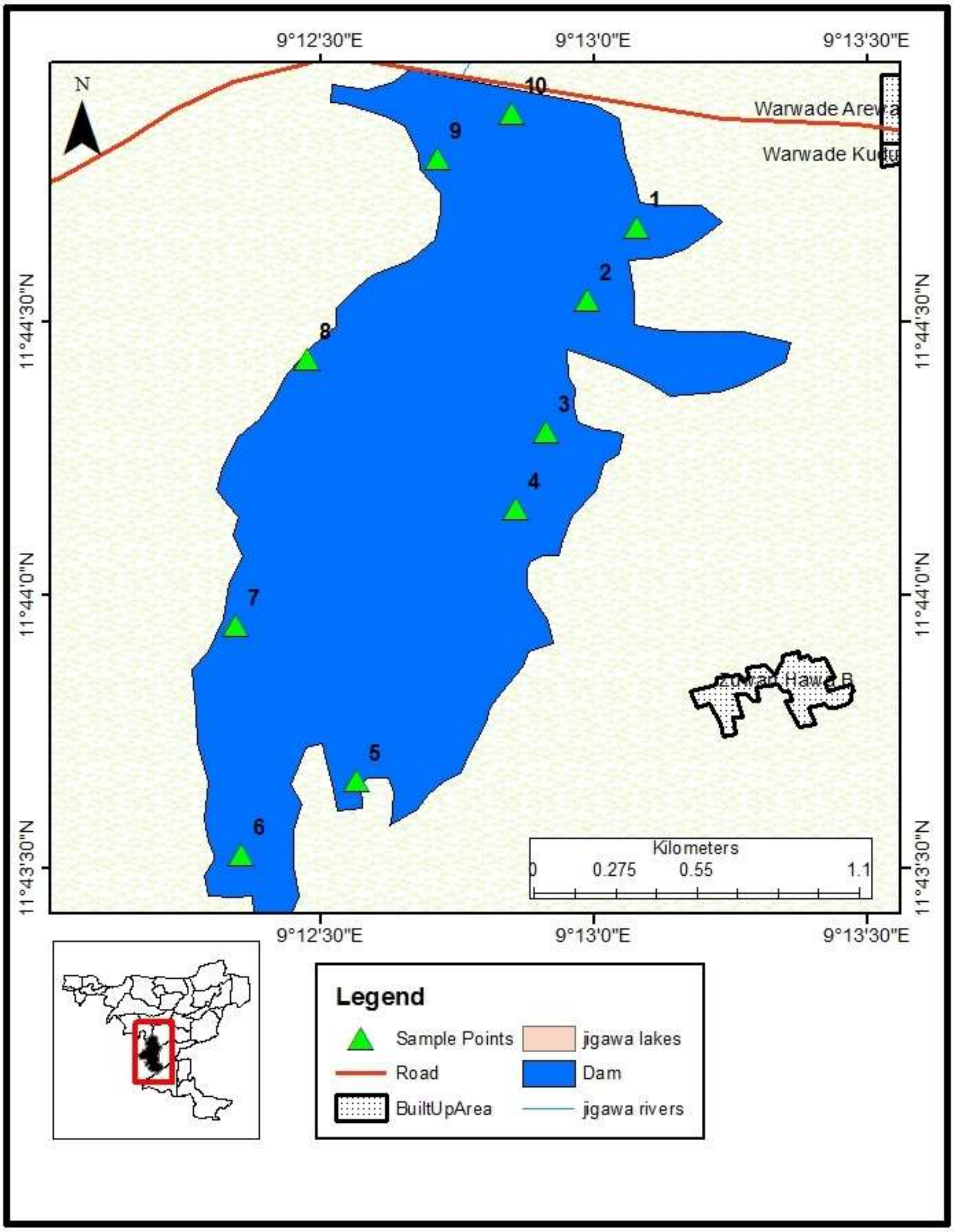

Figure 1: Map Of Warwade Dam Showing The Study Area

Methods

Sampling and Analysis.

Warwade Dam Sampling Sites

These sampling locations' coordinates were all determined using the GPS 12 model (GARMIN USA). Graduated lines were used to take depth measurements (Welcomme, 1985). The anthropogenic conditions of the ten sampling places chosen for this study were different.

Table 1: Warwade Dam Sample Locations

Location	Depth	Latitude	Longitude
1	6.5m	11.74454N	9.21796E
2	6.7m	11.74234N	9.21647E
3	1.4m	11.74343N	9.21507E
4	5.4m	11.74459N	9.214225E
5	6.4m	11.74588N	9.21336E
6	1.6m	11.74706N	9.21280E
7	1.9m	11.74825N	9.21446E
8	6.6m	11.74827N	9.21546E
9	1.7m	11.74770N	9.21655E
10	2.0m	11.74721N	9.21712E

Sampling and Analysis

Samples were collected once a month from each location from March 2020, to February, 2021. Surface water samples were collected from each sampling site and stored in preservative-free polyethylene and acid-washed vials. Turbidity, temperature, pH, and conductivity were measured at the sampling site, while the other parameters were measured in the laboratory. Total phosphorus, ammoniacal nitrogen, nitrite-nitrogen, nitrate-nitrogen, chloride, total hardness, and calcium are all examples of these elements. They were determined within 24 hours of sampling using APHA-approved techniques (APHA, 1998).

Statistical Analysis

Data regarding physicochemical properties of water samples were provided as mean values, and descriptive analysis was used to assess the data. In summarizing the temporal fluctuations of the measured water quality metrics, coefficient of correlation (CV) was employed. The observation period were separated into four fixed seasons: dry season (March, April, and May), rainy season (June, July, and August), pre-winter season (September, October, and November), and winter season (December, January, and February) (December, January, and February). To determine the nature and size of the relationship between various physicochemical parameters regression analy.sis (RA) was conducted To assess whether there was any significant association between water quality metrics and to validate the final conclusion, a best-fit model was determined (highest R^2, i.e. coefficient of determination).

Multivariate Statistical Methods

Data was examined using one-way analysis of variance (ANOVA) at a 0.05 percent level of significance to see if there were any significant differences between the locations for all water quality indicators (Zar, 2009). Cluster analysis (CA) and principal component analysis (PCA) were used to analyze the quality of the water in the dam (Yang et al., 2010). CA and PCA look for groups and sets of variables that have similar features, perhaps allowing us to simplify our observations by finding structure or patterns in the midst of chaotic or perplexing data (Ragno et al., 2007). The SPSS (v. 16) and PAST (v. 1.93) software packages were used to conduct all statistical analyses.

PCA also tries to explain the association between data in terms of underlying characteristics that aren't readily visible (Yu et al., 2003). All of the nutrient concentrations were log-transformed prior to modeling to bring the distribution closer to the normal. A multiparametric model was used to arrive at statistical results and tests. To assess the impact of anthropogenic activities and spatiotemporal fluctuations on physicochemical properties of Warwade dam water, we employed DS, ANOVA, RA, CA, and PCA.

RESULTS AND DISCUSSION

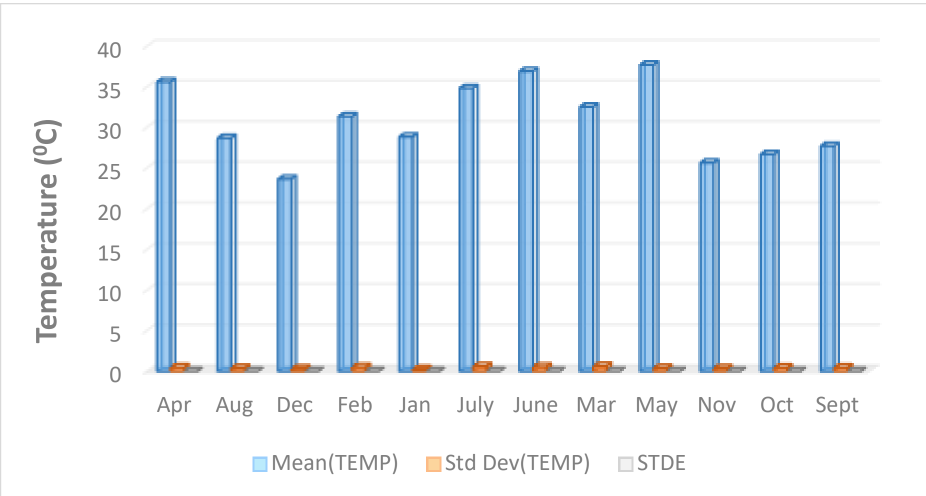

Figure 2: Monthly Temperature Variations (^{0}C)

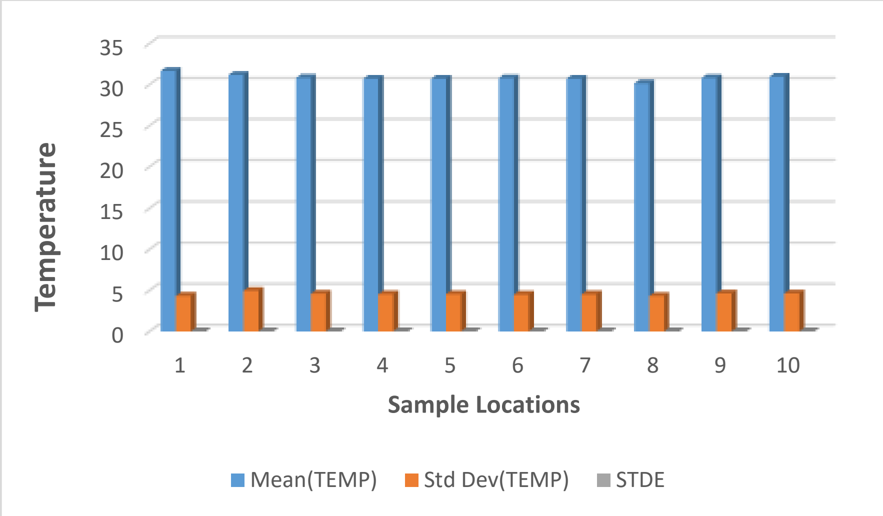

Figure 3: Temperature Variations at Different Locations (^{0}C)

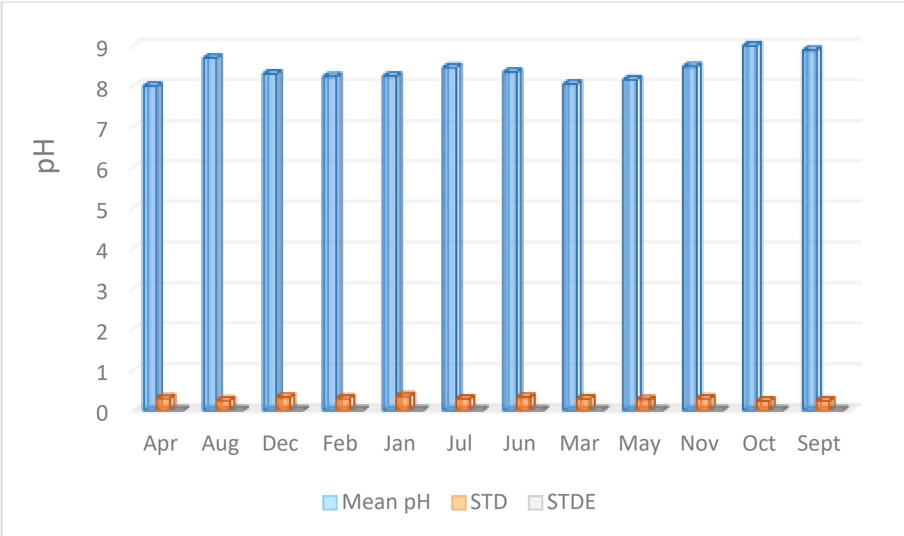

Figure 4: Monthly pH Variations

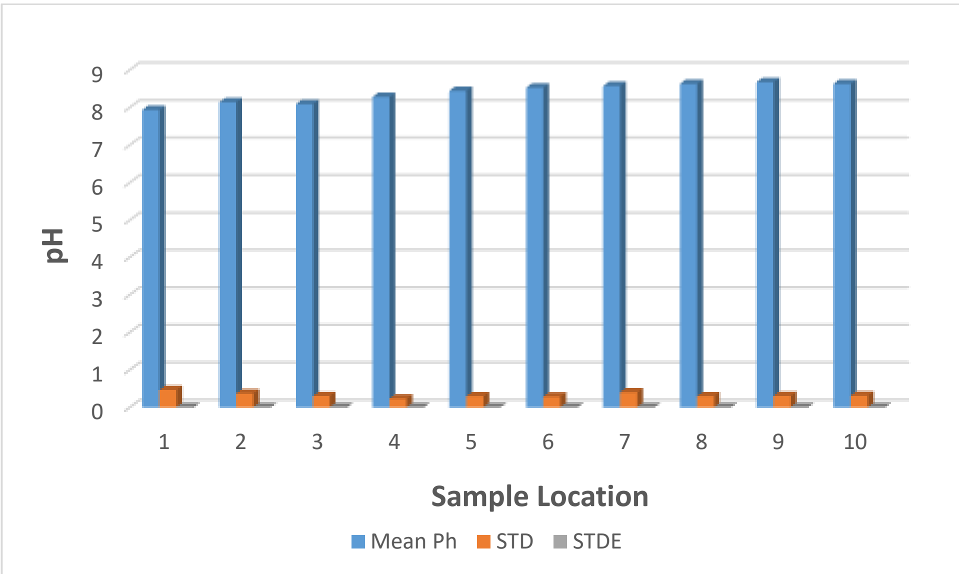

Figure 5: pH Variations at Different Locations

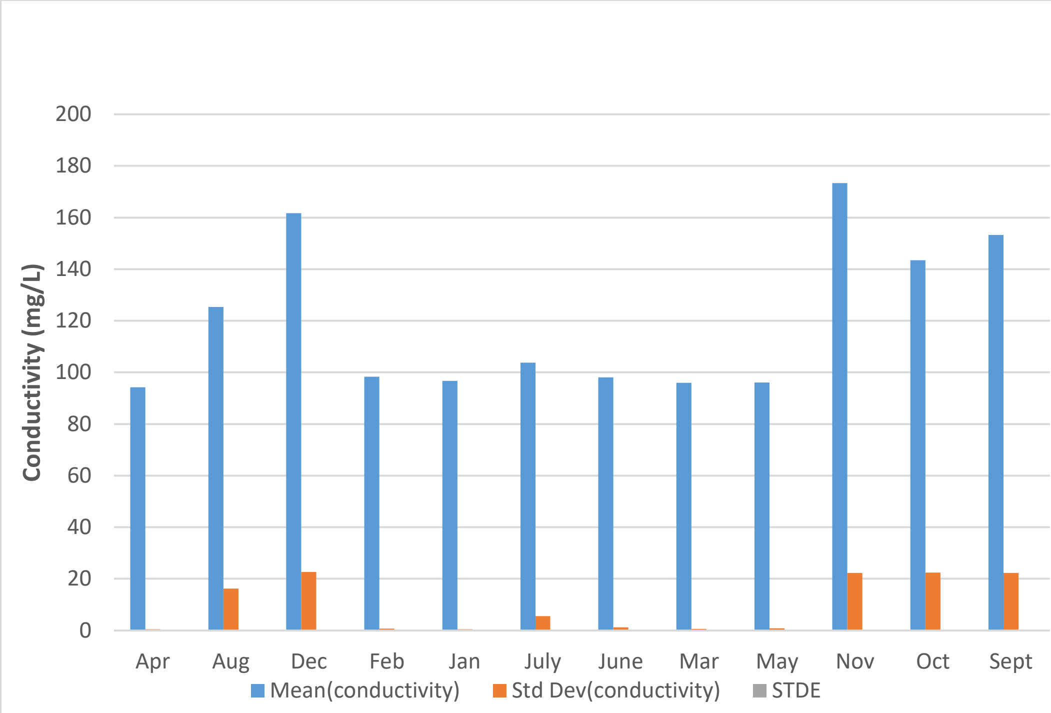

Figure 6: Monthly Conductivity Variations (mg/L)

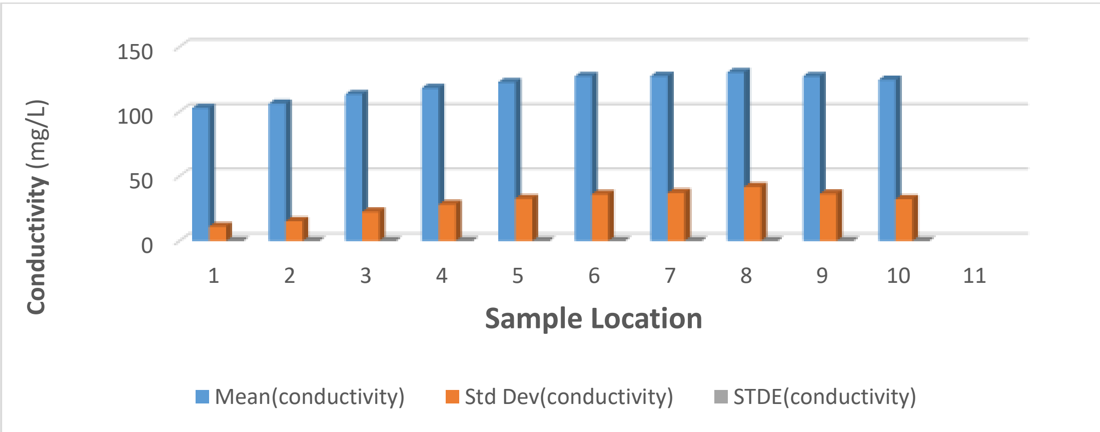

Figure 7: Conductivity Variation at Different Locations (mg/L)

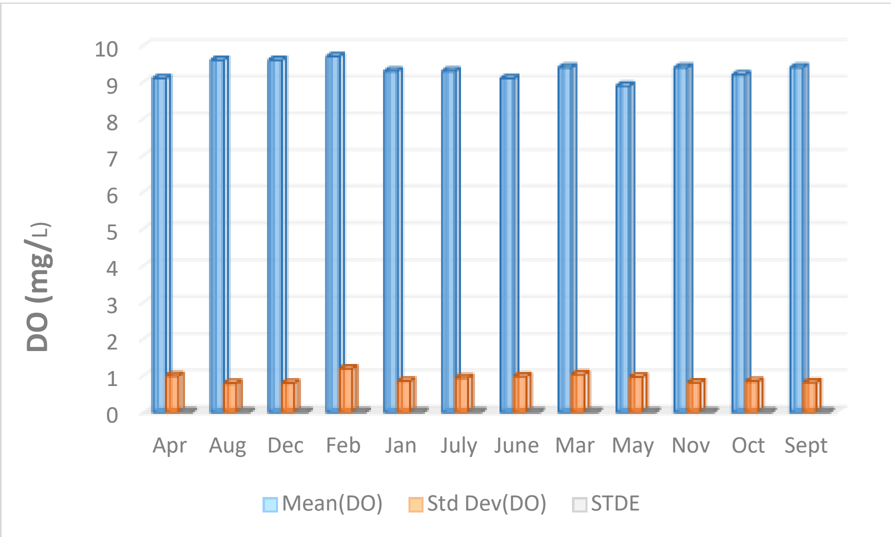

Figure 8: Monthly Variation of DO (mg/L)

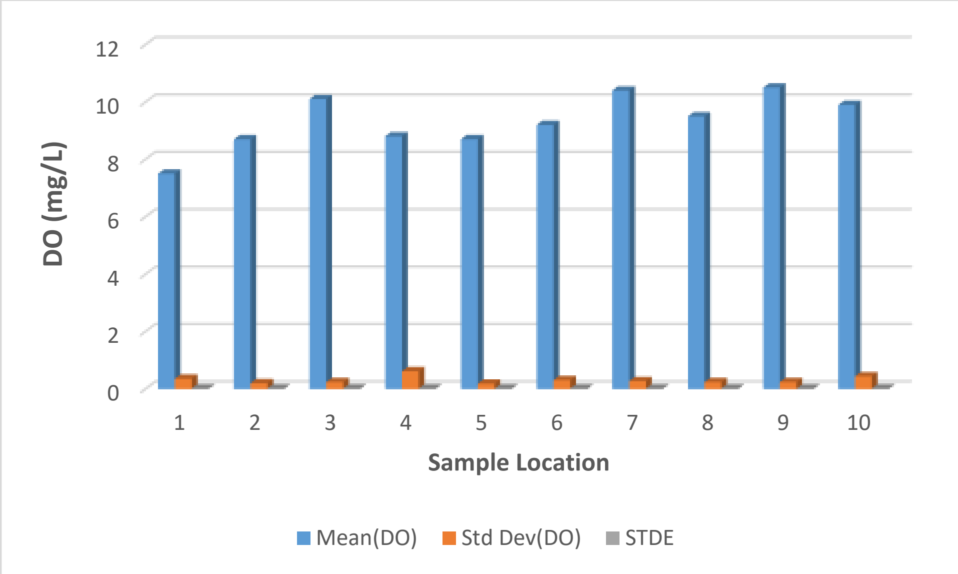

Figure 9: DO Variation at Different Locations (mg/L)

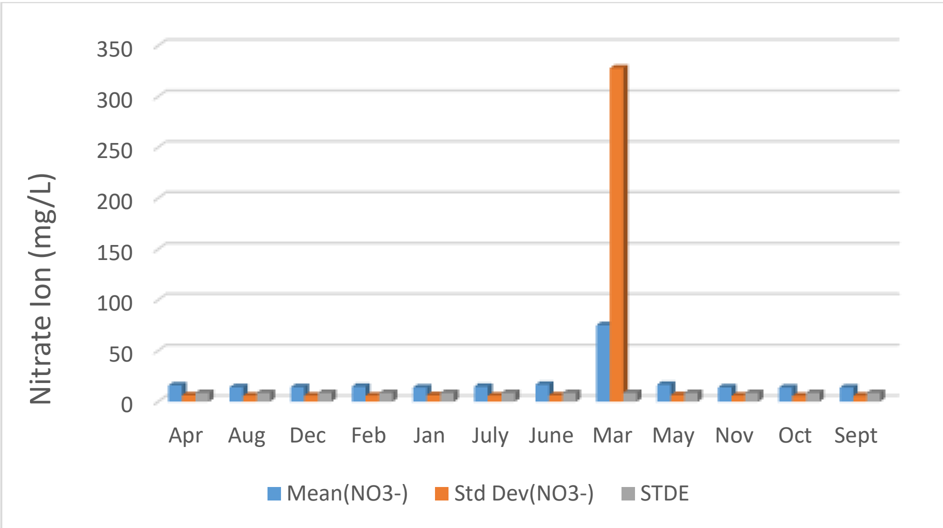

Figure 10: Monthly Nitrate Ion (NO_3^-) Variations (mg/L)

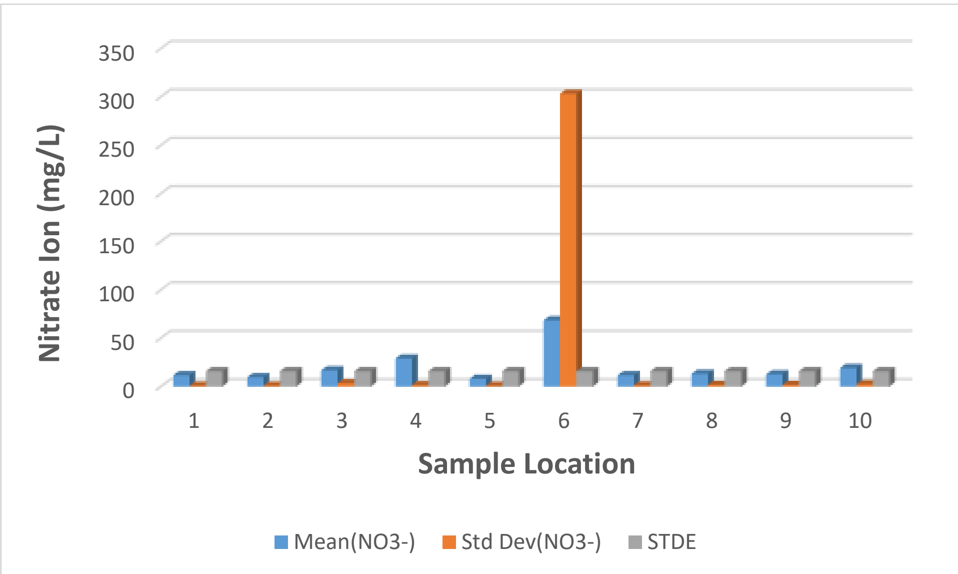

Figure 11: Nitrate Ion (NO_3^-) Variation at Different Locations (mg/L)

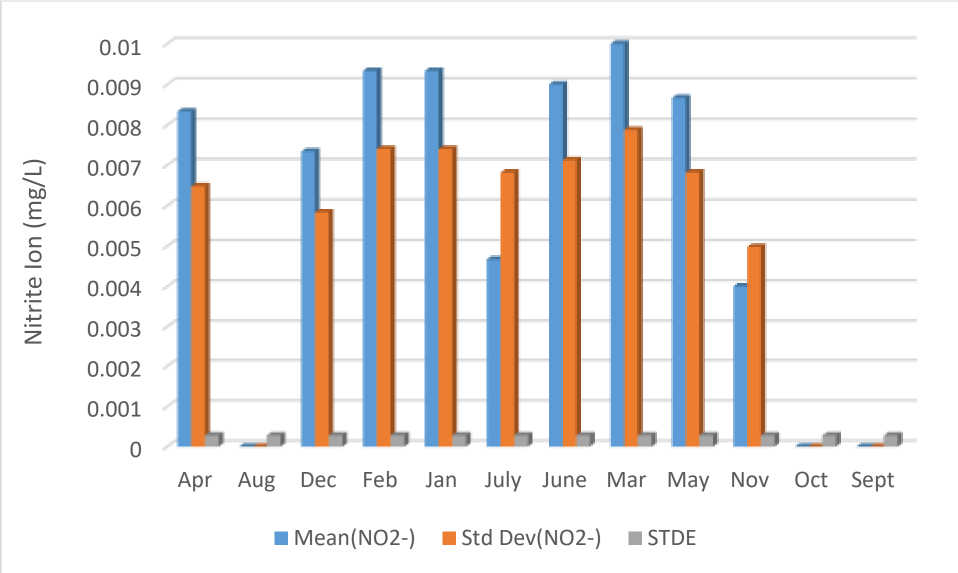

Figure 12: Monthly Nitrite Ion (NO_2^-) Variations (mg/L)

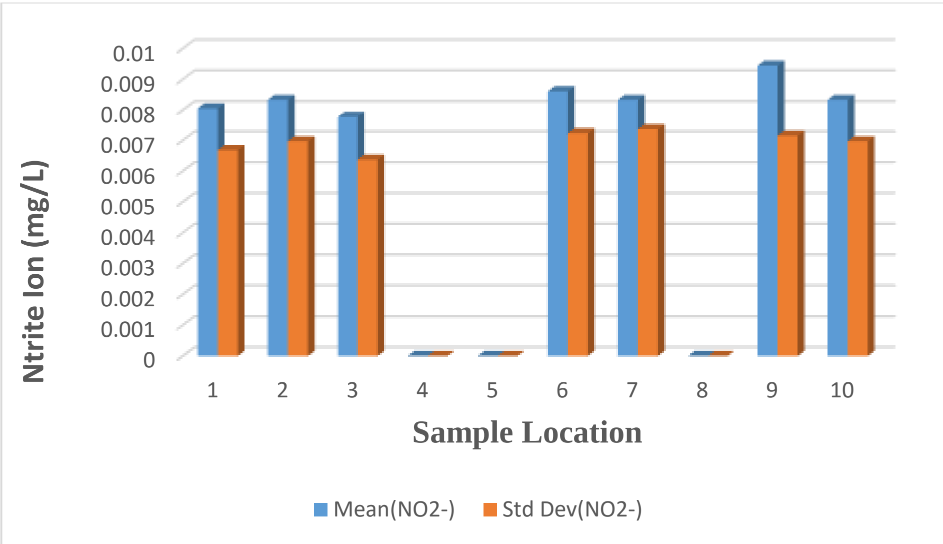

Figure 13: Nitrite Ion (NO_2^-) Variation at Different Locations (mg/L)

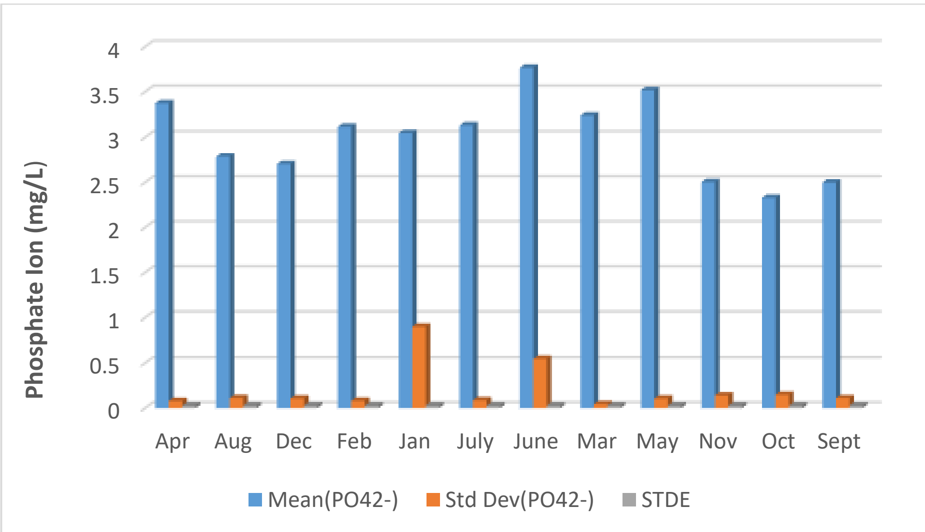

Figure 14: Monthly Phosphate Ion (PO42-) Variations (mg/L)

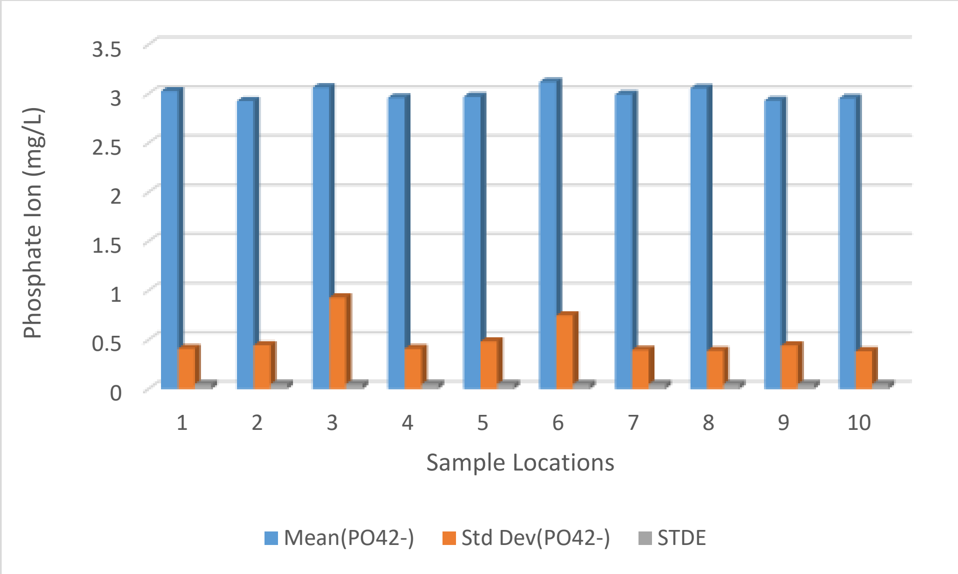

Figure 15: Phosphate Ion (PO42-) Variation at Various Locations (mg/L)

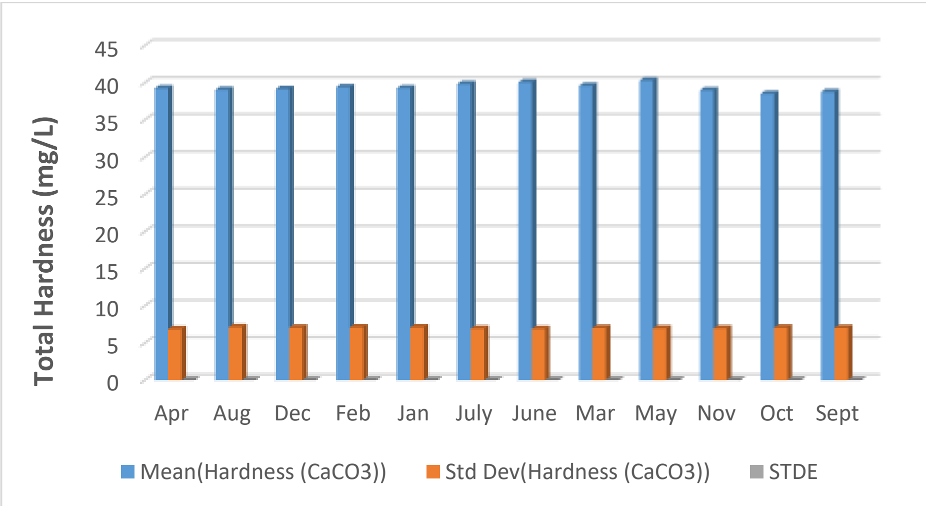

Figure 16: Monthly Total Hardness Variations

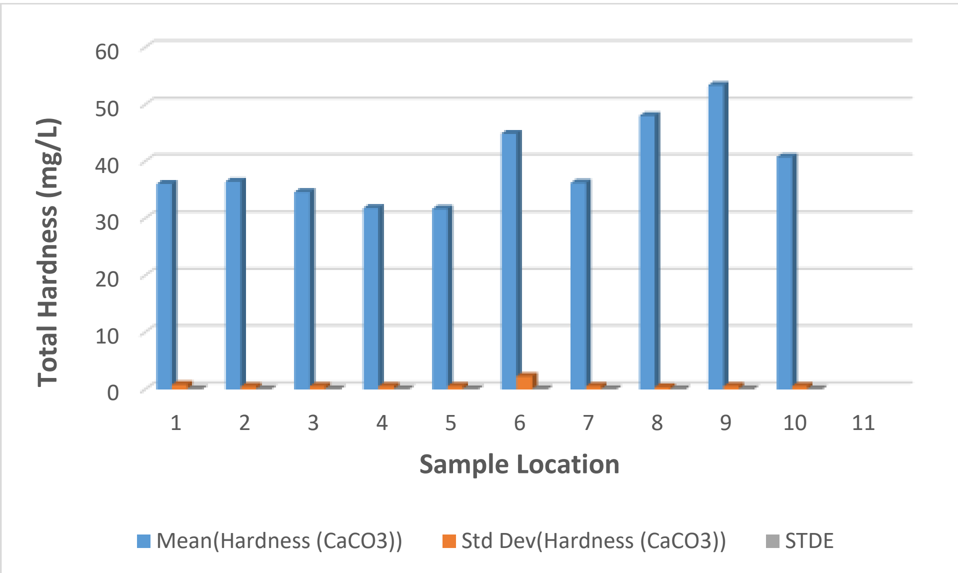

Figure 17: Total Hardness Variation at Different Locations

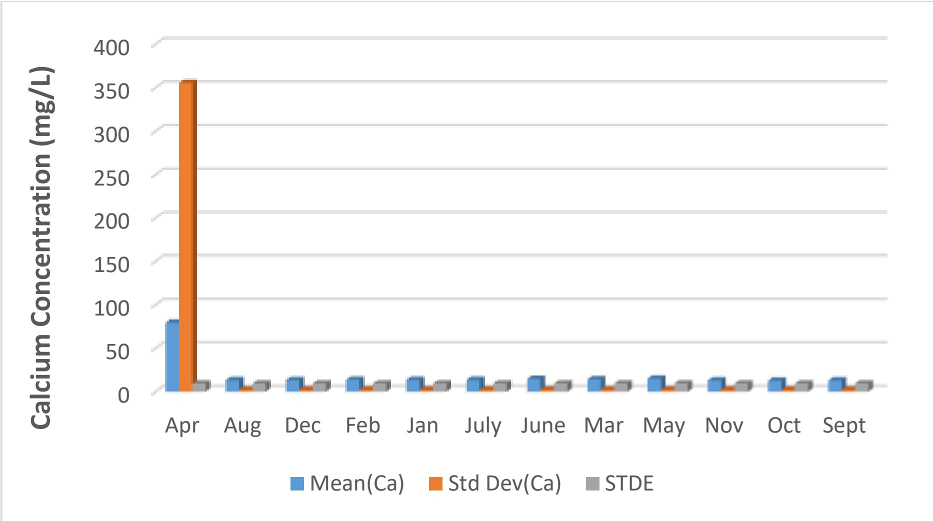

Figure 18: Monthly Calcium Variation (mg/L)

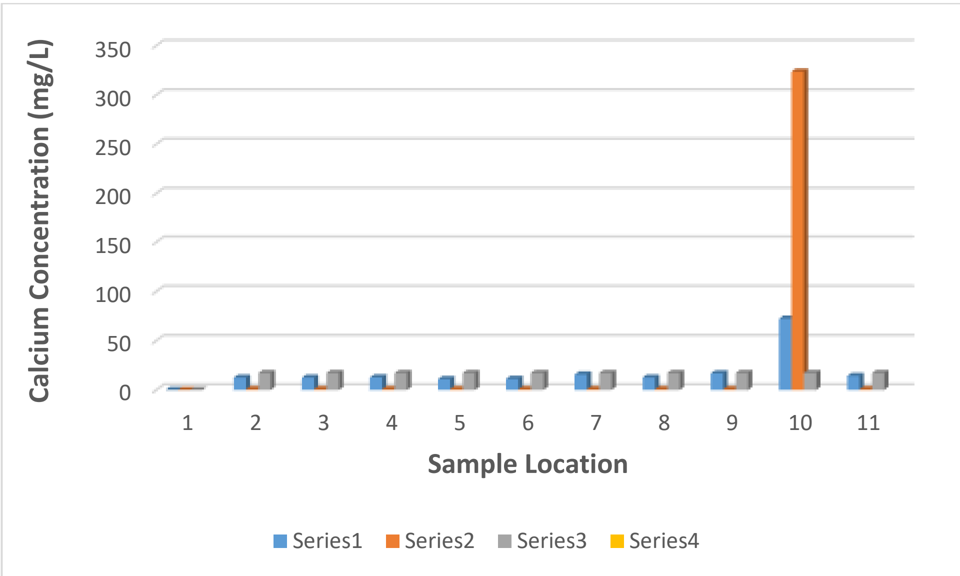

Figure 19: Calcium Variation at Different Location (mg/L)

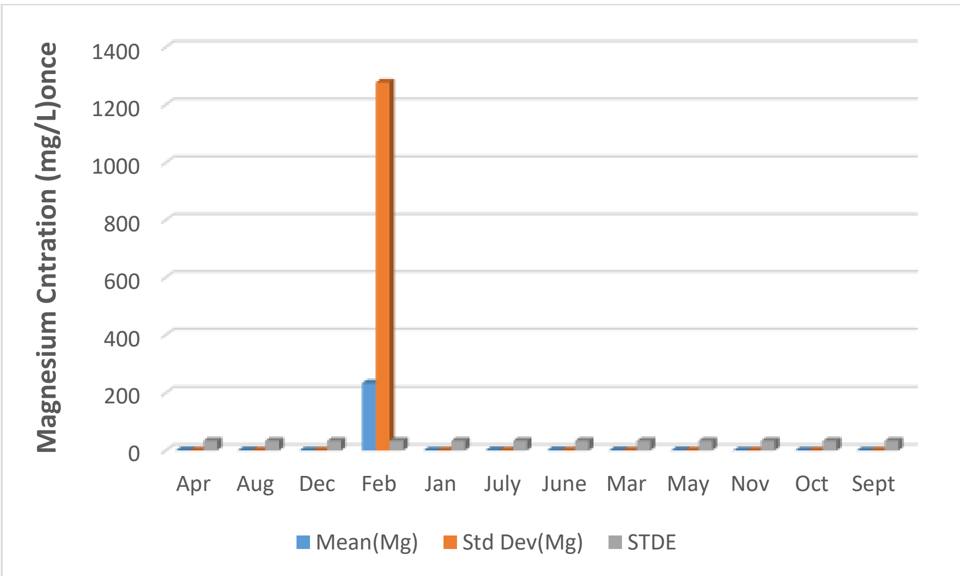

Figure 20: Monthly Magnesium Variation (mg/L)

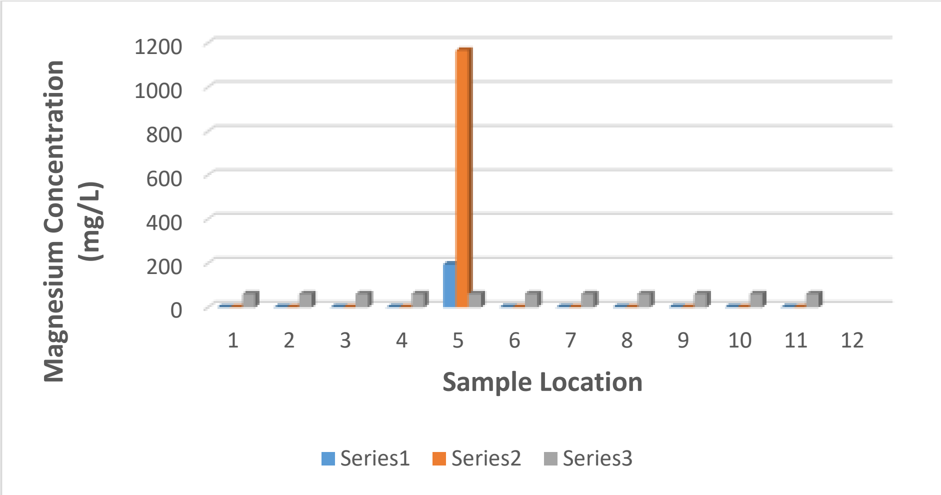

Figure 21: Magnesium Variation at Different Location (mg/L)

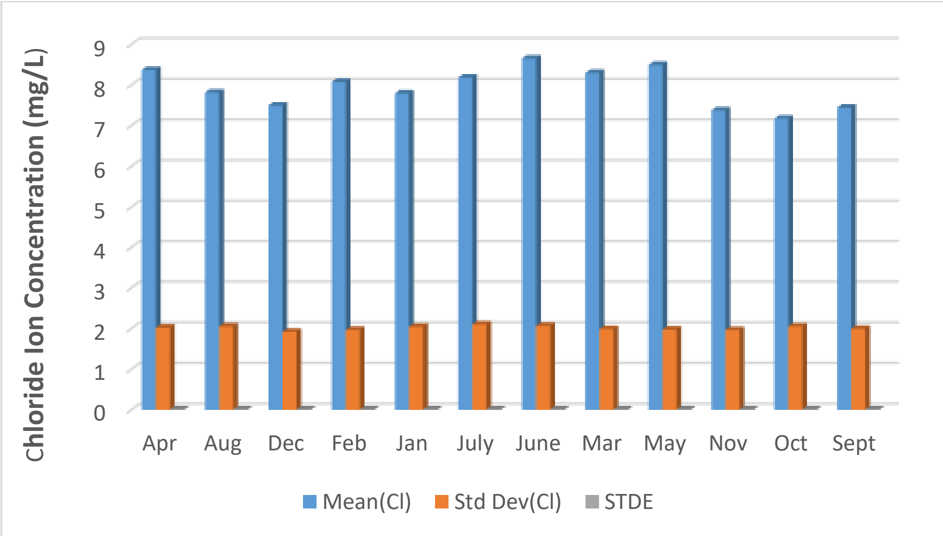

Figure 22: Monthly Chloride Ion Variation (mg/L)

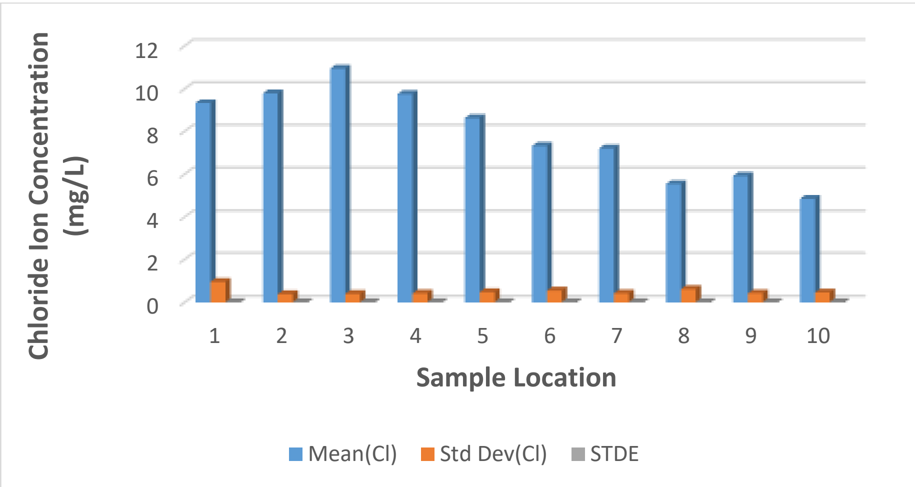

Figure 23: Chloride Ion Variation at Different Location (mg/L)

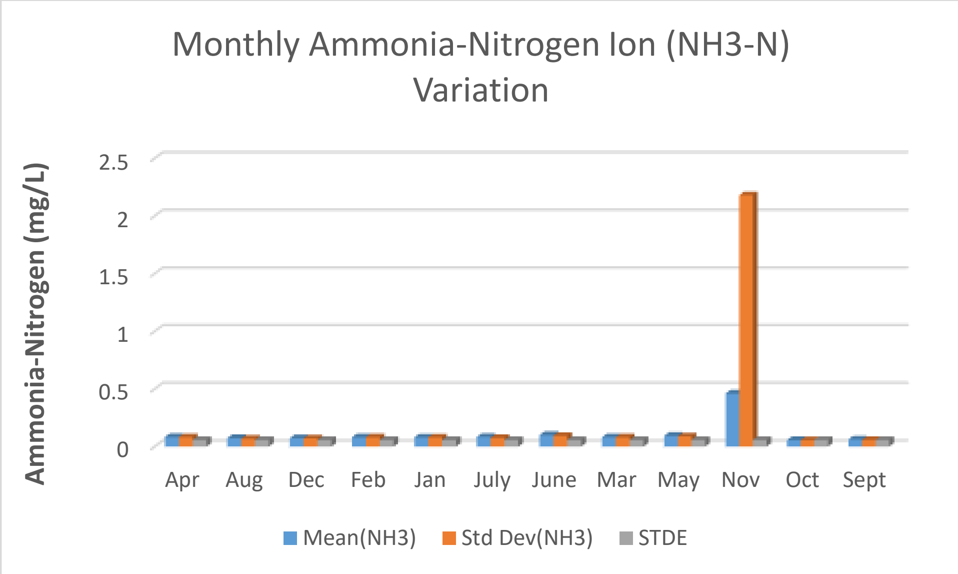

Figure 24: Monthly Ammonia-Nitrogen Ion Variation (mg/L)

Table 2: Correlation Analysis of the Parameters

	pH	EC	Temp.	DO	NO3-	NO2-	PO42-	NH3	T.Hard	Cl^-	Turbd	Ca
pH	1.00											
EC	0.54	1.00										
Temp.	-0.41	0.76	1.00									
DO	0.43	0.24	0.19	1.00								
NO3-	-0.04	-0.05	0.02	0	1.00							
NO2-	-0.41	-0.36	0.3	0.16	0.1	1.00						
PO42-	-0.42	-0.6	0.72	-0.1	0.03	0.01	1.00					
NH3	-0.01	0.07	-0.04	0.01	0.03	0.01	-0.04	1.00				
T. Had	0.33	0.11	0.03	0.24	-0.36	0.18	0.07	0.07	1.00			
Cl^-	-0.64	-0.37	0.24	-0.45	0.02	0.05	0.19	-0.06	-0.68	1.00		
Turbd	0.03	-0.01	-0.05	0.07	0	0.02	-0.03	0	0.11	-0.06	1.00	
Ca	-0.01	-0.04	0.06	0.08	0	0.04	0.04	0	0.13	-0.06	0	1.00

Table 3: Summary of Regression Analysis

S/N	**Parameter**	**RMSE(D)**	**R^2**	**P –value** ≤ 0.05
1	pH	0.077	0.98	0.0001
2	EC	0.095	0.99	0.0001
3	Temperature	0.132	0.99	0.0001
4	DO	0.097	0.99	0.0.0001
5	NO3	0.643	0.34	0.4474
6	NO2	0.003	0.87	0.0001
7	PO42-	0.130	0.80	0.0.0001
8	NH3-N	0.266	0.34	0.3404
9	T. Hardness	0.536	1.0	0.0001
10	Cl-	0.138	1.0	0.0001
11	Turbidity	0.217	0.34	0.4737
12	Ca	0.487	0.34	0.4231

Cluster Analysis

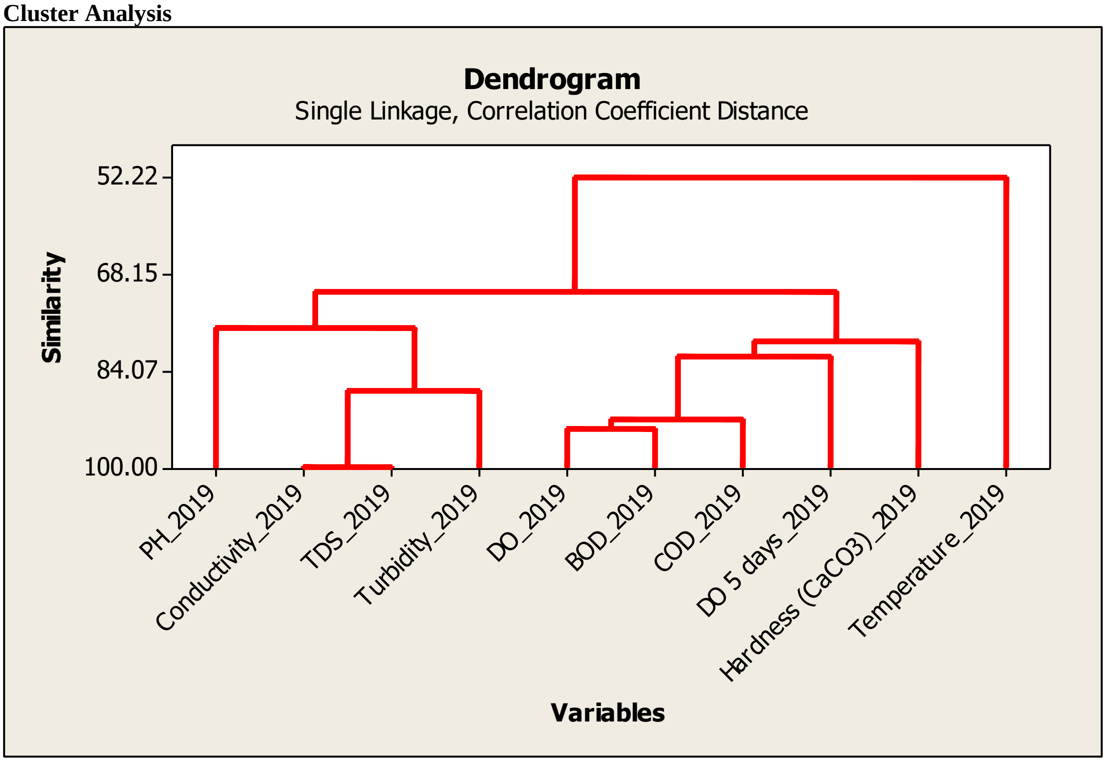

Figure 25: A Dendrogram of Cluster Analysis for Physico-chemical Parameters

A dendrogram is a visual representation of the hierarchical relationship between items. It's most often produced as a result of hierarchical clustering. This dendrogram was generated using the final piece of four clusters that exist at various levels of similarity. The first cluster is made up entirely of PH. Conductivity, TDS, and turbidity are the three variables that make up the second cluster. The third cluster includes five variables: Do, BOD, COD, Do 5 days, and Hardness (CaCo3), while the fourth cluster only includes temperature. Within a cluster, variables are relatively similar, while variables outside of a cluster are very distinct.

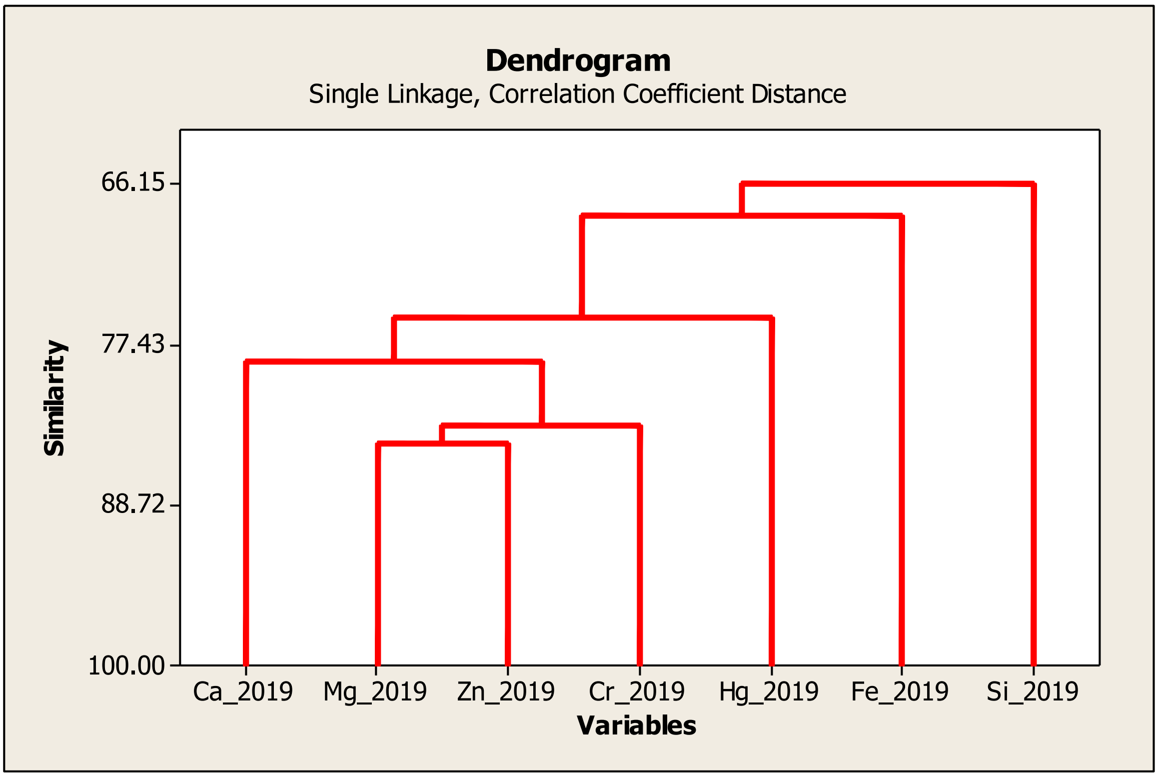

Figure 26: A Dendrogram of Cluster Analysis for selected Metals

A dendrogram is a visual representation of the hierarchical relationship between items. It's most often produced as a result of hierarchical clustering. This dendrogram was generated using the final piece of five clusters that exist at various levels of similarity. The first cluster is entirely made up of Ca. The second cluster is made up of three metals: magnesium, zinc, and chromium. The third cluster is made up of mercury. The fourth cluster is made up entirely of Fe, while the fifth cluster is entirely made up of Si. Within a cluster, variables are relatively similar, while variables outside of a cluster are very distinct.

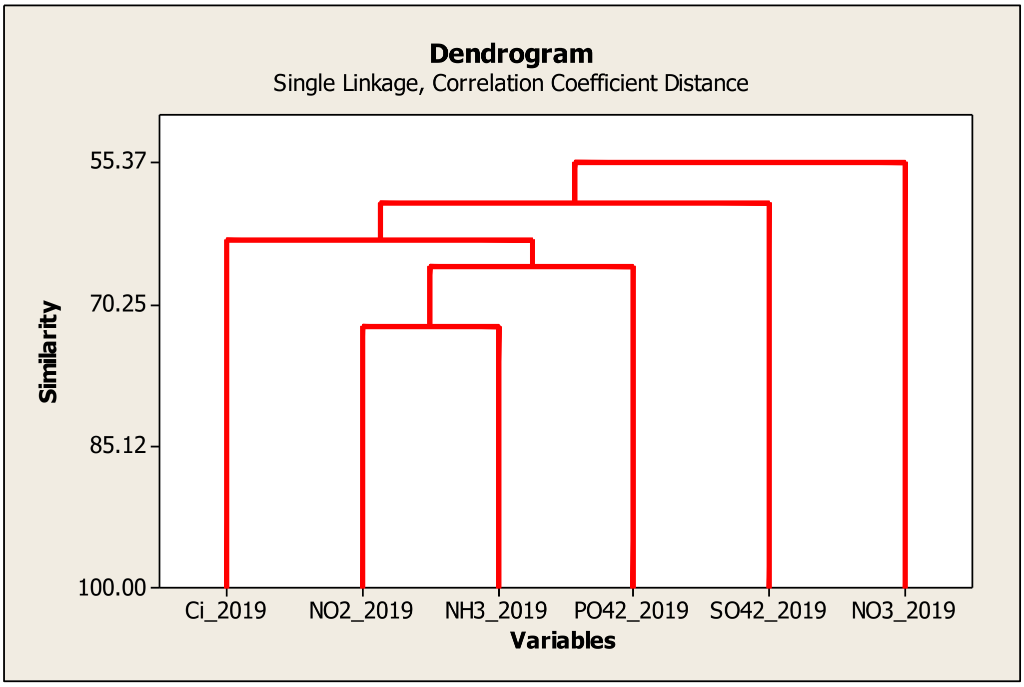

Figure 27: A Dendrogram of Cluster analysis for Non- Metals

A dendrogram is a visual representation of the hierarchical relationship between items. It's most often produced as a result of hierarchical clustering. This dendrogram was generated using the final piece of five clusters that exist at various levels of similarity. The first cluster is made up entirely of Ci. The second cluster is made up of two non-metallic elements: NO2 and NH3. The third cluster is made up entirely of PO42. The fourth cluster is made up entirely of SO42, while the fifth cluster is entirely made up of NO3. Within a cluster, variables are relatively similar, while variables outside of a cluster are very distinct.

Principal Component analysis

Table 4: Principal Component Analysis for Physicochemical Parameters (Communalities)

	Raw		Rescaled	
	Initial	Extraction	Initial	Extraction
PH	.172	.172	1.000	1.000
Conductivity	1008.820	1008.820	1.000	1.000
TDS	434.639	434.639	1.000	1.000
Temperature	20.366	20.366	1.000	1.000
DO	1.113	1.113	1.000	1.000
BOD	281.855	281.855	1.000	1.000
DO__5_days	.377	.377	1.000	1.000
COD	1421.972	1421.972	1.000	1.000
Hardness_CaCO3	48.044	48.044	1.000	1.000
Turbidity	737.956	737.956	1.000	1.000

Extraction Method: Principal Component Analysis.

The Principal Component's Communality The entire influence of all factors on a single observable variable is referred to as analysis. It's the total of all the squared factor loadings for all the factors that affect the observed variable, and it's the same as R2 in multiple regression. The value ranges from 0 to 1, with 1 indicating that the variable is fully defined by the components and does not have any uniqueness. A value of 0 on the other hand, implies that the variable cannot be predicted by any of the components. Because their rescaled communalities are all equal one, the variables in the preceding table are fully defined by the factors connected with them.

These numbers represent the percent of variability assigned to the model in the same way that R squared values in multiple regression do. If you look at the total variance explained in the preceding studies, you'll note that this is how the percent of variance column is generated. We want this value to be as high as possible, as close to one as feasible, since we want the observed dataset to be mirrored in the model.

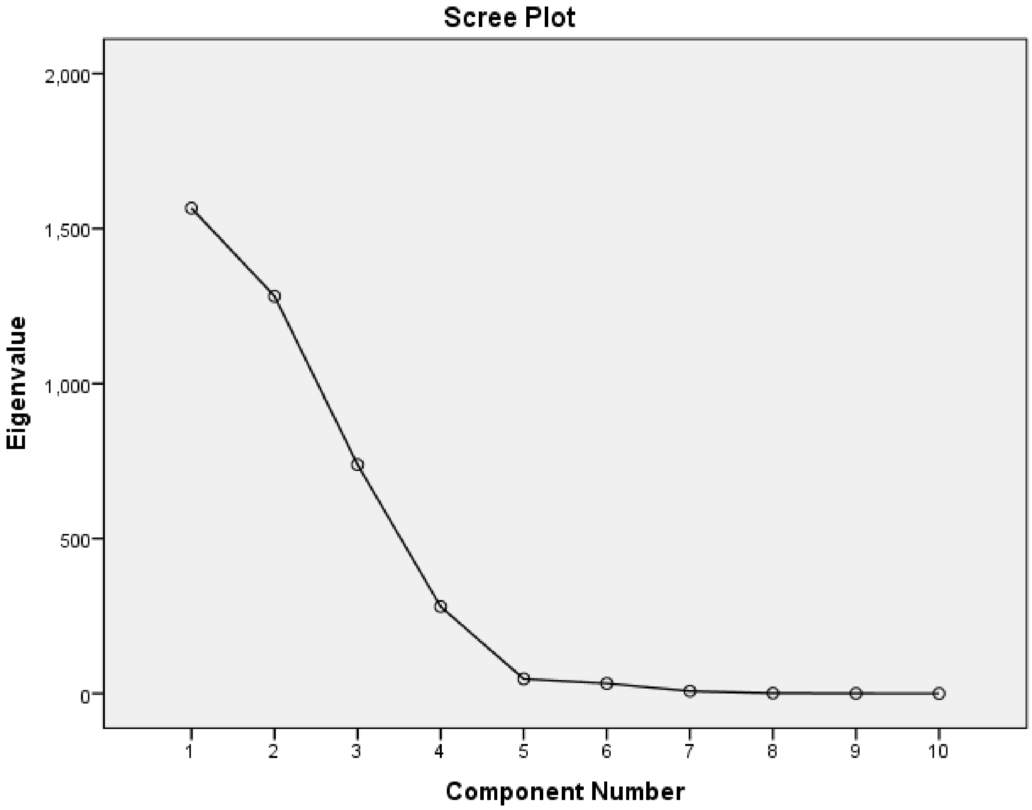

Figure 28: A Scree Plot of Principal Component Analysis

A scree plot can be used to show the cumulative proportion of total variance explained by each PC. A scree plot shows how much variation each principal component catches from the data. PC1 captures the greatest, PC2 the second, and so on. Each one adds to the data's information, and there are as many principal components as there are features in a PCA. The first five main components in the following diagram may be kept because they capture the most variability in the data. This suggests that the first five PCs can sufficiently represent the original ten variables. The first five PCs explained 98.938 percent of the data set variability (see table below)

Table 5: Total Variance Explained

	Component	Initial Eigenvalues[a]			Extraction Sums of Squared Loadings		
		Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
Raw	1	1565.628	39.583	39.583	1565.628	39.583	39.583
	2	1281.889	32.409	71.992	1281.889	32.409	71.992
	3	738.109	18.661	90.653	738.109	18.661	90.653
	4	280.590	7.094	97.747	280.590	7.094	97.747
	5	47.109	1.191	98.938	47.109	1.191	98.938
	6	32.716	.827	99.766	32.716	.827	99.766
	7	7.938	.201	99.966	7.938	.201	99.966
	8	1.111	.028	99.994	1.111	.028	99.994
	9	.123	.003	99.997	.123	.003	99.997
	10	.100	.003	100.000	.100	.003	100.000
Rescaled	1	1565.628	39.583	39.583	2.202	22.023	22.023
	2	1281.889	32.409	71.992	1.749	17.492	39.515
	3	738.109	18.661	90.653	1.025	10.252	49.767
	4	280.590	7.094	97.747	1.022	10.220	59.987
	5	47.109	1.191	98.938	1.149	11.485	71.472
	6	32.716	.827	99.766	.092	.924	72.397
	7	7.938	.201	99.966	.454	4.540	76.937
	8	1.111	.028	99.994	1.453	14.535	91.472
	9	.123	.003	99.997	.283	2.831	94.303
	10	.100	.003	100.000	.570	5.697	100.000

Extraction Method: Principal Component Analysis.
a. When analyzing a covariance matrix, the initial eigenvalues are the same across the raw and rescaled solution.

Table 6: Component Matrix[a]

	Raw									
	Component									
	1	2	3	4	5	6	7	8	9	10
PH	.162	.156	.020	.030	.108	.021	-.022	.094	-.032	.313
Conductivity	23.536	21.087	.483	-.614	.577	-3.027	.288	.003	.001	-.001
TDS	15.133	13.484	.312	-.368	-1.027	4.747	.040	-.006	.000	-.001
Temperature	-2.590	-2.298	-.319	.116	.635	.343	2.781	.070	.006	.001
DO	.213	.126	.077	.076	.360	.066	-.157	.924	-.160	-.039
BOD	1.142	1.090	.141	16.709	-.379	-.062	.008	-.001	.001	.000
DO__5_days	.134	.055	.004	-.027	.042	.017	-.134	.489	.309	.013
COD	27.825	-25.444	-.553	.025	.007	-.006	.001	-.002	.000	.000
Hardness_CaCO3	.558	.515	.804	.923	6.708	.946	-.271	-.062	.007	-.003
Turbidity	-.079	-1.097	27.142	-.098	-.189	-.025	.036	.000	.000	.000

Extraction Method: Principal Component Analysis.

a. 10 components extracted

Table 7: Component Matrix[a]

	Rescaled									
	Component									
	1	2	3	4	5	6	7	8	9	10
PH	.391	.376	.049	.072	.260	.050	-.053	.225	-.078	.754
Conductivity	.741	.664	.015	-.019	.018	-.095	.009	.000	.000	.000
TDS	.726	.647	.015	-.018	-.049	.228	.002	.000	.000	.000
Temperature	-.574	-.509	-.071	.026	.141	.076	.616	.016	.001	.000
DO	.202	.119	.073	.072	.341	.062	-.149	.876	-.152	-.037
BOD	.068	.065	.008	.995	-.023	-.004	.000	.000	.000	.000
DO__5_days	.219	.090	.007	-.044	.068	.027	-.218	.797	.504	.021
COD	.738	-.675	-.015	.001	.000	.000	.000	.000	.000	.000
Hardness_CaCO3	.081	.074	.116	.133	.968	.137	-.039	-.009	.001	.000
Turbidity	-.003	-.040	.999	-.004	-.007	-.001	.001	.000	.000	.000

Extraction Method: Principal Component Analysis.

a. 10 components extracted

Table 8: Communalities

	Raw		Rescaled	
	Initial	Extraction	Initial	Extraction
Ca	10530.338	10530.338	1.000	1.000
Mg	.082	.082	1.000	1.000
Cr	2.342	2.342	1.000	1.000
Fe	85.488	85.488	1.000	1.000
Zn	1.964	1.964	1.000	1.000
Hg	.001	.001	1.000	1.000
Si	.793	.793	1.000	1.000

Extraction Method: Principal Component Analysis.

The Principal Component's Communality The entire influence of all factors on a single observable variable is referred to as analysis. It's the total of all the squared factor loadings for all the factors that affect the observed variable, and it's the same as R2 in multiple regression. The value ranges from 0 to 1, with 1 indicating that the variable is fully defined by the components and does not have any uniqueness. A value of 0 on the other hand, implies that the variable cannot be predicted by any of the components. Because their rescaled communalities are all equal one, the variables in the preceding table are fully defined by the factors connected with them.

These numbers represent the percent of variability assigned to the model in the same way that R squared values in multiple regression do. If you look at the total variance explained in the preceding studies, you'll note that this is how the percent of variance column is generated. We want this value to be as high as possible, as close to one as feasible, since we want the observed dataset to be mirrored in the model

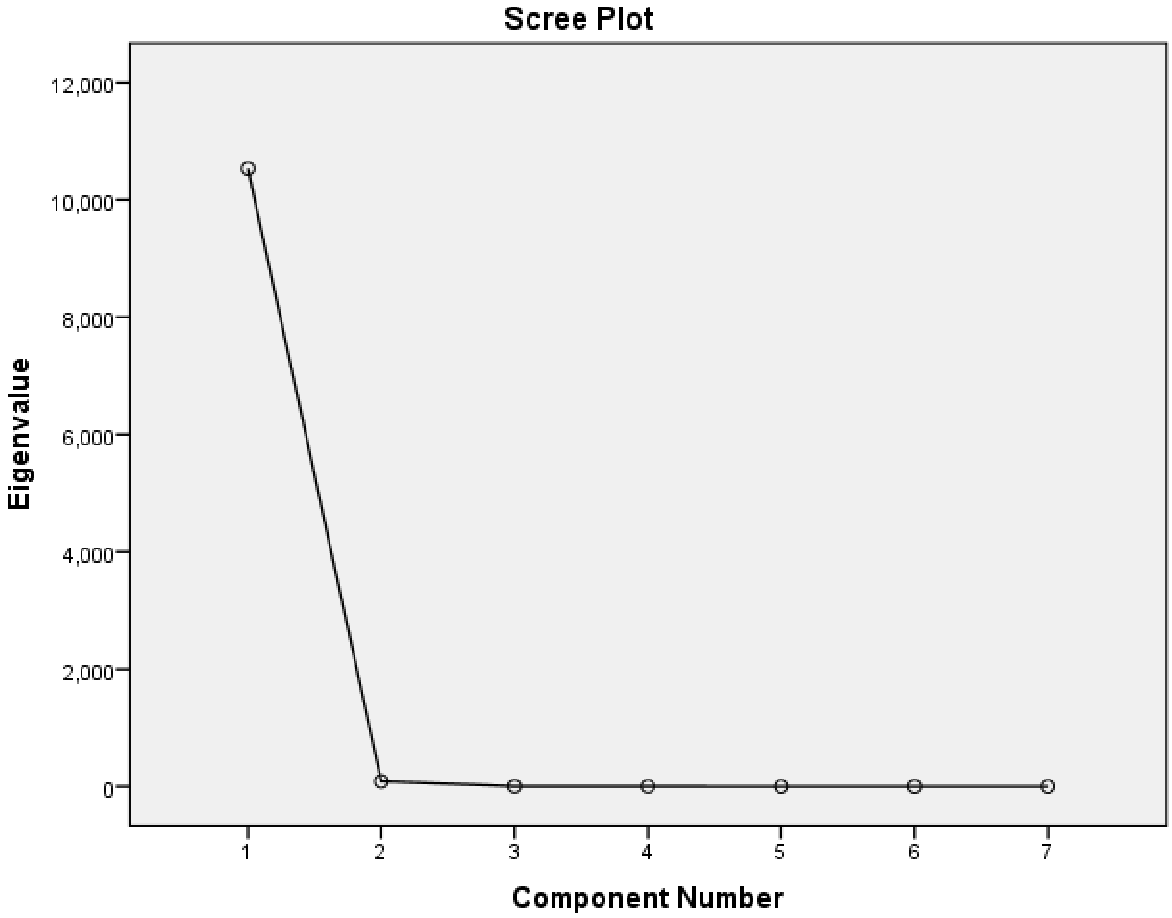

Figure 29: A Scree Plot of Principal Component analysis

A scree plot can be used to show the cumulative proportion of total variance explained by each PC. A scree plot shows how much variation each principal component captures from the data. PC1 captures the most, PC2 the second, and so on. Each one adds to the data's information, and there are as many principal components as there are features in a PCA. The first two main components in the preceding diagram may be kept because they capture the most variability in the data. This suggests that the first two PCs can sufficiently represent the original seven variables. The first two PCs accounted for 99.957% of the variance in the data sets (see table below).

Table 9: Total Variance Explained

	Component	Initial Eigenvalues[a]			Extraction Sums of Squared Loadings		
		Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
Raw	1	10530.342	99.146	99.146	10530.342	99.146	99.146
	2	86.047	.810	99.957	86.047	.810	99.957
	3	3.175	.030	99.986	3.175	.030	99.986
	4	.843	.008	99.994	.843	.008	99.994
	5	.560	.005	100.000	.560	.005	100.000
	6	.038	.000	100.000	.038	.000	100.000
	7	.001	6.819E-006	100.000	.001	6.819E-006	100.000
Rescaled	1	10530.342	99.146	99.146	1.010	14.427	14.427
	2	86.047	.810	99.957	1.382	19.744	34.171
	3	3.175	.030	99.986	2.008	28.692	62.863
	4	.843	.008	99.994	.933	13.332	76.196
	5	.560	.005	100.000	.620	8.863	85.059
	6	.038	.000	100.000	.464	6.626	91.685
	7	.001	6.819E-006	100.000	.582	8.315	100.000

Extraction Method: Principal Component Analysis.
a. When analyzing a covariance matrix, the initial eigenvalues are the same across the raw and rescaled solution.

Table 10: Component Matrix[a]

	Raw						
	Component						
	1	2	3	4	5	6	7
Ca	102.617	.001	-.001	.000	.000	.000	.000
Mg	.024	.095	.173	.003	-.065	.194	.000
Cr	-.002	.587	1.277	-.537	.278	-.007	-.001
Fe	-.013	9.245	-.141	.017	.009	.000	.000
Zn	.062	.473	1.172	.402	-.448	-.023	.000
Hg	-.001	.003	.011	-.017	.010	.002	.027
Si	-.019	-.049	.347	.626	.527	.008	.000

Extraction Method: Principal Component Analysis.
7 components extracted

Table 11: Component Matrix[a]

	Rescaled						
	Component						
	1	2	3	4	5	6	7
Ca	1.000	.000	.000	.000	.000	.000	.000
Mg	.085	.334	.607	.009	-.227	.679	-.001
Cr	-.001	.383	.835	-.351	.182	-.005	.000
Fe	-.001	1.000	-.015	.002	.001	.000	.000
Zn	.044	.338	.836	.287	-.320	-.017	.000
Hg	-.016	.084	.303	-.483	.290	.044	.763
Si	-.021	-.055	.390	.703	.591	.009	.000

Extraction Method: Principal Component Analysis.
7 components extracted

Table 12: Communalities

	Raw		Rescaled	
	Initial	Extraction	Initial	Extraction
Ci	4.141	4.141	1.000	1.000
NO3	36.596	36.596	1.000	1.000
NO2	4.823E-005	4.823E-005	1.000	1.000
SO42	.857	.857	1.000	1.000
PO42	.280	.280	1.000	1.000
NH3	.006	.006	1.000	1.000

Extraction Method: Principal Component Analysis.

The Principal Component's Communality The entire influence of all factors on a single observable variable is referred to as analysis. It's the total of all the squared factor loadings for all the factors that affect the observed variable, and it's the same as R2 in multiple regression. The value ranges from 0 to 1, with 1 indicating that the variable is fully defined by the components and does not have any uniqueness. A value of 0 on the other hand, implies that the variable cannot be predicted by any of the components. Because their rescaled communalities are all equal one, the variables in the preceding table are fully defined by the factors connected with them.

These numbers represent the percent of variability assigned to the model in the same way that R squared values in multiple regression do. If you look at the total variance explained table in the previous analysis, you'll note that this is how the percent of variance column is generated. We want this value to be as high as possible, as close to one as feasible, since we want the observed dataset to be mirrored in the model.

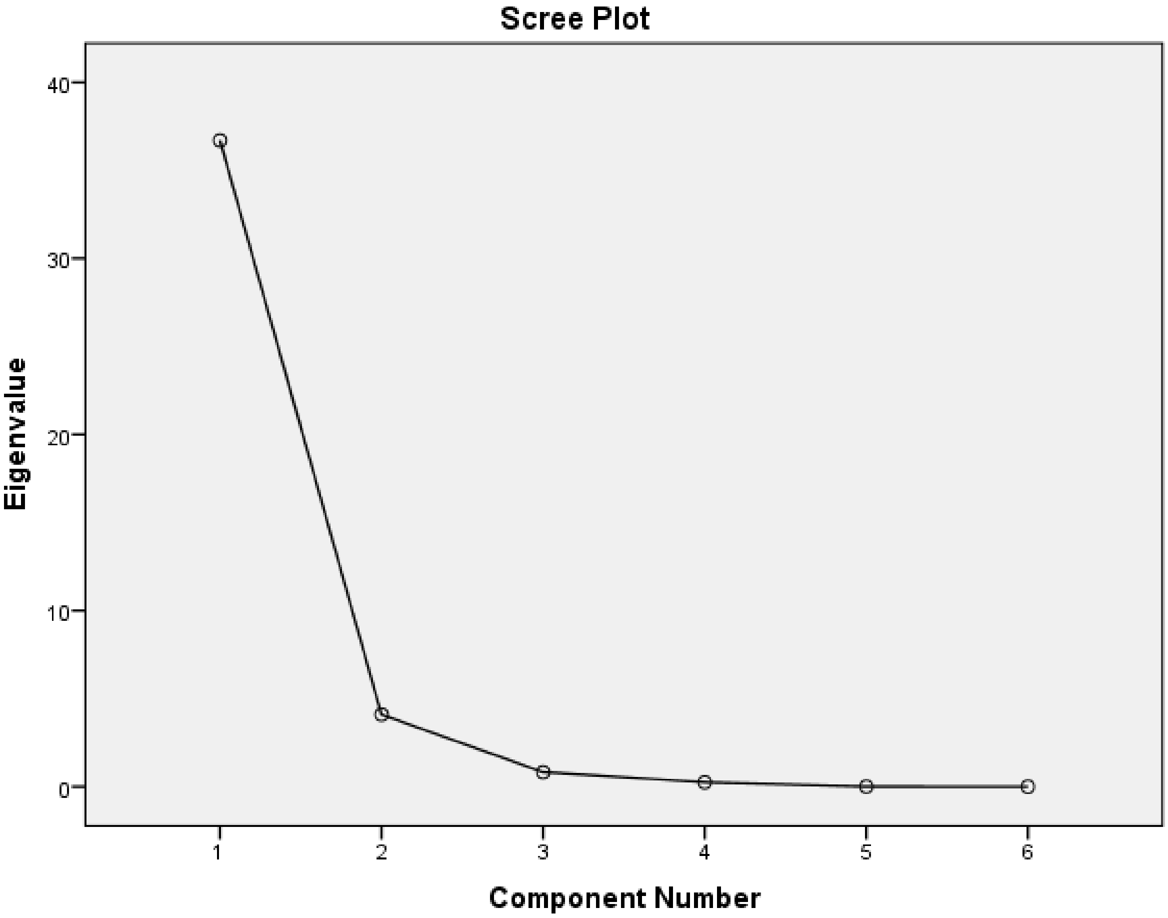

Figure 30: A Scree Plot of Principal Component Analysis

A scree plot can be used to show the cumulative proportion of total variance explained by each PC. A scree plot shows how much variation from the data each principal component captures. The most variety is captured by PC1, the second most by PC2, and so on. Each one adds to the data's information, and there are as many principal components as there are features in a PCA. The first three main components in the preceding diagram may be kept because they capture the most variability in the data. This suggests that the first three PCs can sufficiently represent the original six variables. The first three PCs explained 99.392 percent of the data set variability (see table below).

Table 13: Total Variance Explained

	Component	Initial Eigenvalues[a]			Extraction Sums of Squared Loadings		
		Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
Raw	1	36.702	87.639	87.639	36.702	87.639	87.639
	2	4.095	9.779	97.418	4.095	9.779	97.418
	3	.827	1.974	99.392	.827	1.974	99.392
	4	.250	.596	99.988	.250	.596	99.988
	5	.005	.012	100.000	.005	.012	100.000
	6	3.337E-005	7.969E-005	100.000	3.337E-005	7.969E-005	100.000
Rescaled	1	36.702	87.639	87.639	1.117	18.612	18.612
	2	4.095	9.779	97.418	1.114	18.564	37.175
	3	.827	1.974	99.392	1.079	17.981	55.157
	4	.250	.596	99.988	.949	15.817	70.974
	5	.005	.012	100.000	1.051	17.516	88.490
	6	3.337E-005	7.969E-005	100.000	.691	11.510	100.000

Extraction Method: Principal Component Analysis.
a. When analyzing a covariance matrix, the initial eigenvalues are the same across the raw and rescaled solution.

Table 14: Component Matrix[a]

	Raw						Rescaled					
	Component						Component					
	1	2	3	4	5	6	1	2	3	4	5	6
Ci	.250	2.019	-.010	-.025	-.001	.000	.123	.992	-.005	-.012	.000	.000
NO3	6.049	-.084	.032	-.006	.000	.000	1.000	-.014	.005	-.001	.000	.000
NO2	.000	.000	.001	.002	.003	.006	-.044	.056	.195	.280	.434	.831
S42	-.223	.006	.894	-.091	-.001	.000	-.241	.006	.966	-.098	-.001	.000
PO42	.040	.102	.165	.491	.000	.000	.076	.193	.311	.927	-.001	.000
NH3	-.014	.022	.008	.002	.070	.000	-.190	.298	.108	.028	.929	-.003

Extraction Method: Principal Component Analysis.
a. 6 components extracted.

Discussion

Physico- chemical Parameters

Figures 2–9 show the mean values of physicochemical parameters at several sampling sites in Warwade Dam over a 12-month period (March 2020 to February 2021). During the study period, water temperature, pH, and DO all showed a seasonal pattern. The temperature mean and standard deviation are (30.85 4.51C and 30.85 4.51C, respectively). Alkalinity is indicated by the pH mean and standard deviation of 8.39 0.42. DO levels varied across all sites and seasons, with a mean of 9.330.94 mg/L and a standard deviation of 9.330.94 mg/L. With a mean and standard deviation of 120.0 31.76 S/cm, EC variation was high between seasons and sampling sites. Anthropogenic activities such as garbage disposal and agricultural runoff are blamed for the moderate EC.

The mean and standard deviation of total phosphorus values are 3.0 0.53 mg/L.

With a mean and standard deviation of 0.08 mg/L, ammonical- N revealed seasonal and location fluctuations. The mean and standard deviation of total hardness are 39.35 and 6.93 mg/L, respectively.

TDS mean and standard deviation are 77.83 and 20.47 mg/L, respectively, with a CV of 26.3. The overall CV values revealed a considerable concentration difference. The mean and standard deviation of Ca2+ concentrations are 13.64 2.50 mg/L, with a coefficient of variation (CV) of 6.51. The mean and standard deviation of chloride ion concentrations are 7.93 2.03 mg/L, with a CV of 3.65. Using ANOVA (P 0.05), we found a considerable degree of regional and temporal fluctuation in the concentration of water quality indices.

Phosphate levels in dam water are influenced by domestic wastewaters, particularly those containing detergents and fertilizer runoff. The presence of anthropogenic contaminants is indicated by phosphorus concentrations (Filik et al., 2008). The spatial distributions of nitrate-N increased, owing primarily to contributions from agricultural runoff and sewage outflow (Wu et al., 2009).

Regression Analysis (RA).

To demonstrate the validity of the physico-chemical study, the least fit square plot was created by plotting the regression model of the actual value against the anticipated value. For all factors, the observed associations between real value variables and predicted variables were different and not significant. The validation test was passed by 67 percent of the parameters (very strong fit at P value 0.05), according to the findings of the statistical analysis using the general linear regression model (Table 3). The correlation analysis (Table 2), on the other hand, revealed a wide range of results, including positive, negative, and zero correlations or associations, indicating many sources of origin.

Cluster Analysis (CA).

Cluster analysis is effective for resolving classification problems in which the goal is to arrange elements or variables so that there is a strong degree of relationship between members of the same cluster and a weak degree of linkage between members of other clusters (Brogucira and Cabecadas, 2006). CA revealed a substantial geographical and temporal correlation based on differences in major pollution components in this study, indicating that the effects of human activities on water quality vary both regionally and temporally. The dendrogram depicts the level of pollution as well as the impact of contamination at the sampling locations. It presents a picture of the groups and their proximity, as well as a visual overview of the clustering operations. The researchers performed cluster analysis (CA) to find similarities between the ten sampling sites and four seasons.

CA created a dendrogram based on the percentage of similarity and dissimilarity of the dam water quality characteristics, which was used to group the sampling sites and months. Figure 12 depicts a dendrogram of % similarity of ten study locations based on physicochemical variables. The similarity of research sites was analyzed from 82 percent to 100 percent to determine the degree of relationships between sites as a cluster.

Principal Component Analysis (PCA).

The most relevant factors and physicochemical characteristics affecting water quality were extracted using principal component analysis. It was difficult to draw clear conclusions due to the complicated relationships. Principal component analysis, on the other hand, could not only extract information to some extent and explain the structure of the data in detail, on temporal characteristics by clustering the samples, but it could also describe their different characteristics and help elucidate the relationship between different variables by using the variable lines. Principal component analysis was performed using SPSS 16.0 and PAST software to identify the main principal components from the original variables (Ogino et al., 2001). The 26 physicochemical characteristics were reduced to two primary variables (factors 1 and 2) from the leveling off point(s) in the scree plot based on the eigenvalues scree plot (Figure 5). (Cattell and Jaspers, 1967).

The first factor (17.16), which corresponds to the highest eigenvalue, accounts for almost 66.00 percent of the overall variance. The second factor (7.96), which corresponds to the second eigenvalue, accounts for 30.63 percent of the overall variance. The eigenvalues of the remaining 24 components are smaller than unity. A major factor is one that has an eigenvalue greater than one (Aruga et al., 1993). Farmers employ excessive fertilizers and pesticides during these seasons, resulting in point and nonpoint source contamination from orchards and farmland regions. Negative pH and DO (0.9) loading

CONCLUSION

Using multivariate statistics, this study looked at the relationship between spatiotemporal variability and water quality in the Warwade dam water. All of the parameters that were sampled showed significant spatiotemporal variability. Agricultural runoff and wastewater discharge are the primary influences on dam water quality, according to this multivariate study. The results of PCA revealed that natural soluble salts, nonpoint source nutrients, and anthropogenic contaminants account for the majority of differences in water quality. Runoff increases the concentration of most inorganic and organic parameters during peak mixing season, according to this regression analysis. The findings of this study will aid in the development of a comprehensive watershed management strategy to restore the dam's deteriorating water quality. We advocate restricting the use of excessive fertilizers in agriculture and installing sewage treatment systems in residential areas to stop and reverse the dam's declining water quality.

REFERENCES

Altaf F., G.Meraj, and S. A. Romshoo. (2013).“Morphometric Analysis to Infer Hydrological Behaviour of Lidder Watershed, Western Himalaya, India,” *Geography Journal*, vol. 2013, Article ID 178021.

American Public Health Association (A.P.H.A.). (1998). *Standard Methods For Examination of Water and Waste Water*, American Public Health Association, Washington,DC, USA, 20th edition, 1998.

American Public Health Association (APHA) (1985). *Standard Methods for the Examination of Water and Waste Water* (16th ed.) New York: American Public Health Association Inc. Pg 392-448.

Aruga R., G. Negro, and G. Ostacoli. (1993). “Multivariate data analysis applied to the investigation of river pollution,” *Fresenius’ Journal of Analytical Chemistry*, vol. 346, no. 10-11, pp. 968–975.

Brogueira M. J., and G. Cabec¸adas. (2006). “Identification of similar environmental areas in Tagus estuary by using multivariate analysis,” *Ecological Indicators*, vol. 6, no. 3, pp. 508–515.

Cattell R. B., and J. A. Jaspers. (1967). “A general plasmode (No. 30-10-5-2) for factor analytic exercises and research,” *Multivariate Behavioral Research Monographs*, vol. 67, p. 211.

Davis, E. B. (2006). *Trends in Environmental Research.* New York: Nova Science Publishers, Inc.

Dixon W, and B. Chiswell. (1996). “Review of aquatic monitoring program design,” *Water Research*, vol. 30, no. 9, pp. 1935–1948.

Don-Pedro K. N., E. O. Oyewo, and A. A. Otitoloju. (2004). “Trend of heavy metal concentration in Lagos Lagoon ecosystem, Nigeria,”*West African Journal of Applied Ecology*, vol. 5, pp. 103– 114.

Federal Ministry of Water Resources (FMWR) (2018) MANDATE to develop and implement policies, projects and programmes that will enable sustainable access to safe water

Filik Iscen C., ¨O. Emiroglu, S. Ilhan, N. Arslan, V. Yilmaz, and S. Ahiska. (2008). “Application of multivariate statistical techniques in the assessment of surface water quality inUluabat Lake, Turkey,” *EnvironmentalMonitoring and Assessment*, vol. 144, no. 1–3, pp. 269–276.

Helena B., R. Pardo,M.Vega,E.Barrado, J.M.Fernandez, andL.Fernandez. (2000). “Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal com ponent analysis,” *Water Research*, vol. 34, no. 3, pp. 807–816.

Hornby, A. S. (2005). *Oxford Advanced Learners Dictionary of Current* English (7thed.) London: Oxford University Press, UK. Pp367.

Illah Z., Khan, H., Wascem, A., Mahmood, Q., and Farooq, U. (2013) Water Quality Assessment of the RIVER KABUL at Peshawar, Pakistan: Industrial and Urban Waste Water Impacts. *Journal of Water Chemistry and Technology*. 35 (4): 170-176.

Jarvie H. P., B. A. Whitton, and C. Neal. (1998). “Nitrogen and phosphorus in east coast British rivers: speciation, sources and biological significance,” *Science of the Total Environment*, vol. 210- 211, pp. 79–109.

Mahvi A. H.,, J. Nouri, A. A. Babaei, andR.Nabizadeh. (2005). “Agricultural activities impact on groundwater nitrate pollutio,” *International Journal of Environmental Science and Technology*, vol.2, no. 1, pp. 41–47.

Ogino A., H. Koshikawa, T. Nakahara, and H. Uchiyama. (2001). “Succession of microbial communities during a biostimulation process as evaluated by DGGE and clone library analyses,” *Journal of Applied Microbiology*, vol. 91, no. 4, pp. 625–635.

Papoulis, A., (1990). *Probability & Statistics.* Vol. 2,Prentice-Hall,512pp.

Ragno G., M. D. Luca, and G. Ioele. (2007). “An application of cluster analysis andmultivariate classificationmethods to spring water monitoring data,” *Microchemical Journal*, vol. 87, no. 2, pp. 119– 127.

Shariff, U. (2009). Politics of Rivers Basin Authorities. Daily Trust, Nigeria, Tuesday, 14th April, 2009. P.11.

Simeonov V., P. Simeonova, and R. Tsitouridou. (2004). “Chemometric quality assessment of surface waters two case studies,” *Chemical and Engineering Ecology*, vol. 11, no. 6, pp. 449–469.

Singh K. P., A. Malik, D. Mohan, and S. Sinha. (2004). “Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India): a case study,” *Water Research*, vol. 38, no. 18, pp. 3980–3992.

USEPA. (2007). *Recent Recommended Water Quality Criteria*, United States Environmental Protection Agency.

Welcomme, R.L. (2001). Some considerations of the effects of differences in flood patterns on fish populations. Ecohydrology and Hydrobiology, 1: 313-321.

World Health Organisation (WHO). (1999). Guidelines for Drinking Water Quality. Health Criteria and Other Supporting Information (2nd ed.). New Delhi: AITBS Publishers, Pg 119-328.

Wu M. L, Y.-S.Wang, C.-C. Sun, H.Wang, Z.-P. Lou, and J.-D.Dong. (2009). "Using chemometrics to identify water quality in Daya Bay, China," *Oceanologia*, vol. 51, no. 2, pp. 217–232.

Yang H. Y., F. Zhou, H.-C. Guo. (2010). "Analysis of spatial and temporal water pollution patterns in Lake Dianchi using multivariate statistical methods," *Environmental Monitoring and Assessment, vol. 170,n0. 1-4, pp.407-416.*

Yu S., J. Shang, J. Zhao, and H. Guo. (2003). "Factor analysis ang dynamics of water quality of the Songhua river, Northeast China,"*Water, Air, and Soil Pollution*, vol. 144, no. 1–4, pp. 159– 169.

Zar J. H. (2009). *Biostastical Analysis*, 5th edition, Printice Hall, Englewood Cliffs, NJ, USA.