APPLICATION OF MULTIVARIATE STATISTICAL ANALYSIS IN ASSESSING SURFACE WATER QUALITY IN DEEPOR BEEL AREA, GUWAHATI, ASSAM

DISSERTATION IS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENT FOR THE DEGREE OF

Master of Technology

in

Civil Engineering
(With specialization in Water Resources Engineering)
under

ASSAM SCIENCE AND TECHNOLOGY UNIVERSITY

SESSION: 2021-2023

Submitted by:

LAKSHANA KATAKI

M.Tech 4th Semester

Roll No.: 210620061010

Guided by:

Dr. TRIPTIMONI BORAH

Professor

Assam Engineering College

DEPARTMENT OF CIVIL ENGINEERING
ASSAM ENGINEERING COLLEGE
JALUKBARI, GUWAHATI-781013

CANDIDATE DECLARATION

(Session: 2021-2023)

I hereby declare that the work presented in the dissertation entitled "APPLICATION OF MULTIVARIATE STATISTICAL ANALYSIS IN ASSESSING SURFACE WATER QUALITY IN DEEPOR BEEL AREA, GUWAHATI, ASSAM" in the partial fulfillment of the requirement for the award of the degree of Master of Technology in Civil Engineering with specialization in water Resources Engineering submitted in the Department of Civil Engineering, Assam Engineering College, Jalukbari, Guwahati-13, under Assam Science and Technology University, has been carried out by me under the supervision of Dr. (Mrs.) Triptimoni Borah, Associate Professor, Department of Civil Engineering, Assam Engineering College, Guwahati. Whatever I have presented in this report has not been submitted by me for the award of any other degree or diploma.

Dated:

Place: Guwahati

LAKSHANA KATAKI

M. Tech 4th Semester

Roll No: 210620061010

Department of Civil Engineering

Assam Engineering College

Guwahati-781013

CERTIFICATE OF SUPERVISION

(Session: 2021-2023)

This is to certify that the work contained in the report entitled "APPLICATION OF MULTIVARIATE STATISTICAL ANALYSIS IN ASSESSING SURFACE WATER QUALITY IN DEEPOR BEEL AREA, GUWAHATI, ASSAM" has been carried out by Lakshana Kataki, Roll No:210620061010 a student of M. Tech 4th semester in the Department of Civil Engineering, Assam Engineering College, Guwahati, under my guidance and supervision and submitted in the partial fulfillment of the requirement for the award of degree of Masters of Technology in Civil Engineering with specialization in Water Resources engineering under Assam Science and Technology University.

Dated:

Place: Guwahati

Dr. (Mrs.) Triptimoni Borah

Professor,

Department of Civil Engineering

Assam Engineering College

CERTIFICATE OF APPROVAL

(Session:2021-2023)

This is to certify that Lakshana Kataki, Roll No : 210620061010, M.Tech 4th Semester, Civil Engineering Department (Water Resources Department), Assam Engineering College, has submitted her dissertation on "APPLICATION OF MULTIVARIATE STATISTICAL ANALYSIS IN ASSESSING SURFACE WATER QUALITY IN DEEPOR BEEL AREA, GUWAHATI, ASSAM " in partial fulfillment of the requirements for the award of the Masters of Technology in Civil Engineering with specialization in Water Resource Engineering under Assam Science and Technology University.

Dated:

Place: Guwahati

DR. JAYANTA PATHAK

Professor & Head of the Department

Department of Civil Engineering

Assam Engineering College

Guwahati-781013

ACKNOWLEDGEMENT

I wish to express my deep sense of gratitude to my guide Dr. (Mrs.) Triptimoni Borah, Associate Professor, Department of Civil Engineering, Assam Engineering College for her guidance, suggestion and encouragement throughout the course of my work. The confidence and firm belief that she had bestowed upon me and the amount of patience and the endurance she had, cannot be adequately expressed. The various values I have learned from her shall remain a source of inspiration for me forever.

I am indebted to all the faculties of Department of Civil Engineering for the deep insights and discernment given through the various courses they had taught.

I also owe to all my friends who have supported me during the course of my study**.**

LAKSHANA KATAKI

Roll No: 210620061010

Department of Civil Engineering

Assam Engineering College,

Guwahati - 781013

NOTATIONS

Symbols	Description
BIS	Bureau of Indian standards
IS	Indian Standards
KM	Kilometer
WHO	World Health Organization
$CaCO_3$	Calcium Carbonate
pH	Hydrogen-ion concentration
Cl	Chloride
Fe	Iron
F	Floride
D.O	Dissolved Oxygen
B.O. D	Biological Oxygen Demand
NO_3	Nitrate
GPS	Global Positioning System
Mg/l	Milligram per litre
ppt	Parts per trillion
ppm	Parts per million

TABLE OF CONTENTS

CHAPTER 1:	**INTRODUCTION**	**1-3**
1.1:	General	1-3
1.2:	Objective of the Study	3
CHAPTER 2:	**LITERATURE REVIEW**	**4-10**
CHAPTER 3:	**WATER QUALITY PARAMETERS**	**11-21**
3.1:	Introduction	11-12
3.2:	Water Quality Standards	12
3.3:	Water Quality Parameters	12
3.3.1:	Physical Parameters	12-15
3.3.2:	Chemical Parameters	15-19
3.3.3:	Biological Parameters	20-21
CHAPTER 4:	**DEEPOR BEEL, GUWAHATI – THE STUDY AREA**	**22-25**
4.1:	Introduction	22
4.2:	Location and Boundaries	23
4.3:	Physiography	23
4.4:	Climate	23-24
4.5:	Hydrology	24
4.6:	Resources of Deepor Beel	24
4.7:	Environmental Problems of Deepor Beel	24-25
CHAPTER 5:	**METHODOLOGY**	**26-31**
5.1:	Selection of Sites	26
5.2:	Collection and Pre-Treatment of Samples	27
5.3:	Analytical Methods of Testing	27
5.3.1:	Determination of Temperature	27
5.3.2:	Determination of Colour	27
5.3.3:	Determination of Odour	28
5.3.4:	Determination of Total Hardness by Titration	28-29
5.3.5:	Determination of Chloride by Titration (Argentometric Method)	29
5.3.6:	Determination of B.O.D.	30-31

CHAPTER 6:	**RESULTS: WATER QUALITY ANALYSIS**	**32-41**
6.1:	Result	32
6.1.1:	Tabular Output of Results	32-35
6.1.2:	Graphical Output of Results	35-41
CHAPTER 7:	**MVDA**	**42-84**
7.1:	Multivariate Data Analysis	42
7.1.1:	Advantages and Disadvantages of MVDA	42
7.1.2:	Dependence Methods	42-43
7.1.3:	Interdependence Methods	43-44
7.2:	Analysis of Variance and Correlation	44
7.2.1:	One-way ANOVA	44-46
7.3:	Hierarchical Cluster Analysis (HCA)	47
7.3.1:	Considerations for HCA	47-48
7.3.2:	Output for Hierarchical Cluster Analysis for Autumn Season	48-49
7.3.3:	Ward Linkage for Autumn Season	49-50
7.3.4:	Dendrogram Plot for Autumn Season for Nine Sites	50
7.3.5:	Output for Hierarchical Cluster Analysis for Winter Season	51
7.3.6:	Ward Linkage for Winter Season	52
7.3.7:	Dendrogram Plot for Winter Season for Nine Sites	53
7.3.8:	Output for Hierarchical Cluster Analysis for Spring Season	53
7.3.9:	Ward Linkage for Spring Season	54
7.3.10:	Dendrogram Plot for Spring Season for Nine Sites	55
7.3.11:	Output for Hierarchical Cluster Analysis for Summer Season	55
7.3.12:	Ward Linkage for Summer Season	56
7.3.13:	Dendrogram Plot for Summer Season for Nine Sites	57
7.4:	Principal Component Analysis (PCA)	57-59
7.4.1:	PCA for Autumn Season	59-62
7.4.2:	PCA for Winter Season	62-64
7.4.3:	PCA for Spring Season	64-66
7.4.4:	PCA for Summer Season	67-69
7.5:	K-means Cluster Analysis	69-72

7.6:	Correlation Matrix (Pearson Correlation Method)	72-75
7.7:	Water Quality Index	75
7.7.1:	Calculation of WQI by Weighted Arithmetic Water Quality Index Method	75-77
7.8:	Discussion	77-82
CHAPTER 8:	**CONCLUSION**	**83-84**
	References	85-86

LIST OF FIGURES

Fig No.	Description	Page No.
1.1	Image of Deepor Beel	2
4.1	Map of Assam Showing Deepor Beel	22
5.1	GPS Location of Sampling Sites of Deepor Beel	26
6.1	Frequency Distribution of Temperature Variation in the Sampling Sites	35
6.2	Frequency Distribution of the pH Concentration in the Sampling Sites	36
6.3	Frequency Distribution of the Total Hardness Concentration in the Sampling Sites	36
6.4	Frequency Distribution of the Total Dissolved Solids Concentration in the Sampling Sites	37
6.5	Frequency Distribution of Dissolved Oxygen Concentration in the Sampling Sites	37
6.6	Frequency Distribution of B.O.D. Concentration in the Sampling Sites	38
6.7	Frequency Distribution of Chloride Content in the Sampling Sites	38
6.8	Frequency Distribution of Salinity Concentration in the Sampling Sites	39
6.9	Frequency Distribution of Electric Conductivity in the Sampling Sites	39
6.10	Frequency Distribution of Iron Concentration in the Sampling Sites	40
6.11	Frequency Distribution of Nitrate Concentration in the Sampling Sites	40
6.12	Frequency Distribution of Lead Concentration in the Sampling Sites	41
6.13	Frequency Distribution of Turbidity Concentration in the Sampling Sites	41
7.1	Proximity Matrix for Autumn Season	49
7.2	Agglomeration Schedule for Autumn Season	50
7.3	Dendrogram Plot for Autumn Season	50
7.4	Proximity Matrix for Winter Season	51
7.5	Agglomeration Schedule for Winter Season	52
7.6	Dendrogram Plot for Winter Season	53
7.7	Proximity Matrix for Spring Season	53
7.8	Agglomeration Schedule for Spring Season	54
7.9	Dendrogram Plot for Spring Season	55
7.10	Proximity Matrix for Summer Season	55
7.11	Agglomeration Schedule for Summer Season	56
7.12	Dendrogram Plot for Summer Season	57
7.13	Scree Plot for Autumn Season	60
7.14	Component Plot for Autumn Season	61
7.15	Scree Plot for Winter Season	63
7.16	Component Plot for Winter Season	63
7.17	Scree Plot for Spring Season	65

7.18	Component Plot for Spring Season	66
7.19	Scree Plot for Summer Season	68
7.20	Component Plot for Summer Season	68
7.21	Graphical Comparison between means of K-means Clusters	72

LIST OF TABLES

Table No.	Description	Page No.
3.1	Standards for Water Quality Parameters as per BIS (IS-10500:2012)	20
6.1	Concentrations of Water Quality Parameters of the Sampling Sites for Autumn Season	32
6.2	Concentrations of Water Quality Parameters of the Sampling Sites for Winter Season	33
6.3	Concentrations of Water Quality Parameters of the Sampling Sites for Spring Season	33
6.4	Concentrations of Water Quality Parameters of the Sampling Sites for Summer Season	34
7.1	One-way ANOVA Datasheet for Autumn Season using SPSS	44
7.2	One-way ANOVA Datasheet for Winter Season using SPSS	45
7.3	One-way ANOVA Datasheet for Spring Season using SPSS	45
7.4	One-way ANOVA Datasheet for Summer Season using SPSS	46
7.5	PCA Analysis for Autumn Season	59
7.6	PCA Analysis for Winter Season	62
7.7	PCA Analysis for Spring Season	64
7.8	PCA Analysis for Summer Season	67
7.9	Statistics of the Four Seasons Detected from the K-means Clustering Analysis	71
7.10	Correlation Matrix of the Water Quality Parameters in Autumn Season	73
7.11	Correlation Matrix of the Water Quality Parameters in Winter Season	74
7.12	Correlation Matrix of the Water Quality Parameters in Spring Season	74
7.13	Correlation Matrix of the Water Quality Parameters in Summer Season	75
7.14	Water Quality Rating as per Weighted Arithmetic Water Quality Index Method	76
7.15	BIS Standards for various Water Quality Parameters for Drinking Purpose	76
7.16	Weighted Arithmetic Water Quality Index Values	77

ABSTRACT

Deepor Beel is a prominent and popular wetland located in the city of Guwahati in the state of Assam. Notably, it is the only Ramsar site present in the state. Deepor Beel is known for its rich biodiversity that it withholds along with its scenic beauty. But, increasing pollution and other anthropogenic activities has led to the deterioration of the water quality of this very wetland. The present study was carried out to analyze the water quality parameters of the Beel and check the correlation that existed among various parameters. The present study was carried out to analyze the 13 physico-chemical water quality parameters of the Beel and to run a multivariate data analysis in IBM SPSS software in order to check any relationship among each other. The multivariate analysis carried out for our study area includes one way ANOVA, Hierarchical Cluster Analysis (HCA), Principal Component Analysis (PCA) and K-Means Cluster Analysis available in the IBM SPSS software package. A total of thirteen parameters viz. Water temperature, Dissolved Oxygen, Biological Oxygen Demand, pH, Turbidity, Total Hardness, Chloride Content, Total Dissolved Solids, Salinity, Conductivity, Iron content, Nitrate content and Lead content were tested for nine sampling locations in and around the areas of Deepor Beel. The samples for testing were collected at 9 locations around the Beel in order to note the difference in results among the different sites. Standard methods were followed for the collection, sampling and analysis of the water quality parameters. Moreover, Water Quality Index for all 9 sampling sites throughout all season were calculated to classify the water quality according to the degree of purity with the help of the measured water quality parameters.

Key words: *Physico-chemical parameters, water quality analysis, Deepor Beel, ANOVA, HCA, PCA, K-means Cluster Analysis, Pearson Correlation Matrix, WQI*

CHAPTER 1

INTRODUCTION

1.1. GENERAL

Deepor Beel is a permanent freshwater lake, situated at the south western fringe of Guwahati, the state capital of Assam a North Eastern State in India. This lake was once a channel of Brahmaputra River (MoEF, 2018). Deepor Beel is categorized as Burma Monsoon Forest Biogeographic Region Wetland (Roy & Kalita, 2011). It is a large (589 ha) and shallow (depth 1 m–4 m) lake with a 12 km buffer area (NWA, 2013). The lake plays an important role in the local economy. Approximately 1200 local families from 14 surrounding villages are dependent upon it for maintaining livelihood by fishing, cultivation and the collection of medicinal plants, orchids and Nymphaea nuts and flowers (Islam et al., 2014; NWA, 2013). The lake is extremely important to the environment as it serves crucial ecological and hydrological services as the sole major storage basin of storm water for Guwahati city (Deka & Goswami, 1992). It is an important temporary shelter for many migratory birds including some globally threatened species (NWA, 2013). As Deepor Beel is home to large congregations of birds (about 150 species), a portion of 414 ha of this lake have been declared as a bird sanctuary (Barman & Saikia, 1995). Due to its importance for the bird population, the lake is considered as an Important Bird and Biodiversity Area (IBA) by the Bird Life International with a high priority for conservation (BLI, 2001). Beside birds, 50 species of fishes, 18 genera of phytoplankton, 21 genera of zooplankton, at least 20 amphibian species, 12 species of lizards, 18 species of snakes and 6 species of turtle and tortoise, along with a large variety of tropical aquatic fora (NWA, 2013; Saikia & Saikia, 2011) thrive in and around the lake. The lake also provides a habitat to many species of amphibians, reptiles, insects, microphytes, macrophytes, weeds, etc., which are both ecologically and economically important (Bera et al., 2008). For being habitat for so many species, including some threatened ecological communities.

Fig 1.1: Image of Deepor Beel

Deepor Beel was declared as a Ramsar Site in 2002 (NWA, 2013). Unfortunately, various anthropogenic activities are degrading the environment of this lake. Heavy fishing and poaching of water birds poses a threat to the biodiversity (RIS, 2002). Pesticides and fertilizers, used in the adjacent agricultural areas, are carried to the lake by run off and pollute the water. Water hyacinth infestation occurs due to eutrophication of the lake as a result of excessive fertilizer run off (RIS, 2002). Deposition of Municipal Solid Waste in the adjacent dumping ground contaminates the lake with toxic substances and heavy metals (Roy & Kalita, 2011). Discharge of sewerage through Bahini and Bharalu Rivers (Roy & Kalita, 2011), as well as increase in human settlement and industrial activities in the periphery of the lake, also contribute significantly to the pollution (Mozumder & Tripathi, 2014) of this lake. As Deepor Beel is quite important in both ecological and economical aspects, the quality of lake water plays the pivotal role in proper sustenance of the biological resources and livelihood of the local people. Developing an understanding of changes in the water quality of the lake is essential for the proper management of the environment and economy of this region. Previous studies on Deepor Beel were mostly on the issues of biodiversity (Bera et al., 2008; Islam et al., 2014), impact of urbanization (Bhattacharyya & Kapil, 2010), cause and effect of pollution from different sources (Choudhury & Gupta, 2017; Mozumder & Tripathi, 2014; Roy & Kalita, 2011; Sayed et al., 2015) and short-term assessment of water quality (Dutta et al., 2016; Islam et al., 2014). However, no study on long-term assessment of the lake was found, neither any with water quality trend analysis. The water quality of the lake was, therefore, assessed on long-term basis, along with trend analysis. This was a longitudinal study involving the testing of water samples from the lake over a period of one years to assess temporal water quality trends (both qualitative and quantitative). Physicochemical parameters of the lake were assessed for determining the quality of water (Roy & Majumder, 2019). The parameters were selected in such a way that they may be utilized in multiple WQI, used in this study. This study provides insight into changes in water quality of Deepor Beel. It is hoped that this information will help in the development of an improved management system. Assessment of water quality trends in Deepor Beel, Assam, to assist in ease of understanding WQI, ranges of WQI are given linguistic terms which express the overall quality of water. For example, NSF WQI has five classes, namely “Excellent”, “Good”, “Medium”, “Bad” and “Very Bad”. Such a classification system is often an outcome of the human preference. In this study, artificial neural network (ANN) - based unsupervised classification of the samples was performed for a non-preferential classification of different water quality ranges.

1.2. OBJECTIVE OF THE STUDY

The objectives of this study are:

1. To study the cause behind the contamination of sites in and around the Deepor Beel.

2. To carry out a systematic water quality analysis for a year (Winter, Spring, Summer, Autumn) in order to know the spatio-temporal changes in water quality of the water body.

3. To analyze the concentration of water quality parameters by some statistical analysis in order to co-relate these parameters and know their effect in the area under study.

4. To determine the most potential parameter/factor responsible for the aforesaid effect of water quality deterioration.

CHAPTER 2

LITERATURE REVIEW

- **Ritabrata Roy. et al (2019)**, their study was mainly focused on the assessment of water quality trends in Deepor Beel, which is located in the south western fringe of Guwahati, the state capital of Assam which is a north eastern state of India. The lake was once a channel of river Brahmaputra. This lake is a productive habitat of lots of fish species as well as flora and fauna, also, it is considered as a home of many birds. So, the study was carried out to know the prevailing conditions of water quality parameters of the Beel. In their study, the different physical and chemical parameters that were checked are water temperature, P^H, DO, Total hardness, EC, BOD, COD, TSS, TDS, nitrate, phosphate and turbidity. The water samples were taken from Deepor Beel in every month for three years of period. Ten fixed sampling location were used consistently throughout the study. The sampling points were selected in such a way that they may spread throughout the mid-range of the entire surface of the lake. After the three years of long observation of various sample tested, the result showed that there was distinct variation in selected water quality parameters of the Beel water. They also put a statistical analysis of the results obtained by 10^{th} order polynomial regression and PNN. This study assessed the water quality and seasonal water quality variations of Deepor Beel. Qualitative trends of water quality were assessed by Mann–Kendall Test and quantitative trends by Polynomial Regression and PNN Regression. Their study suggested that the water quality of Deepor Beel was fair during the study period, with some deterioration during summer. However, lake water was little turbid and eutrophic throughout the study period. The water quality deteriorated during summer and improved during spring and post monsoon in the study period. The water of Deepor Beel also exhibited a gradual yearly deterioration during summer and post monsoon. So, it can be concluded that Deepor Beel requires the implementation of a better management plan, as soon as possible, for its improvement and sustenance.

- **Asari N.A. (2017)**, in his journal paper "Seasonal variations in physicochemical characteristics of water samples of Surajpur Wetland, National Capital Region, India" analyzed the quality of water and the correlations among the various water quality parameters in the wetland.

The study was carried out for 2-year period ad sample collection was doe from 5 different locations on monthly basis. All total 12 parameters were tested for each sample. The most changeable and sensitive water quality parameters such as Water Temperature, Water depth, pH, Turbidity and Dissolved Oxygen (DO) were measured in the field using field test kit. Remaining parameters like Total Hardness, Total Alkalinity, Chloride, Nitrate, Fluoride, Phosphate and Iron were measured according to standard laboratory methods. The data obtained were analyzed using a software PAST (version 2.15). One-way analysis of variance (ANOVA) was used to check the significance of difference among the results of the parameters in different months. Moreover, in order to check the relationship among various physico-chemical parameters, Pearson linear correlation was used in the investigation. The results obtained from various tests showed that the values of pH and turbidity in water remained more or less uniform throughout the study in both the years. Thus, from the results obtained it was concluded that the water quality status of the wetland was in favorable condition with minerals concentration within permissible limits and was good enough to support rich biodiversity to form a complete food web in the Surajpur wetland ecosystem.

- **Bhat et. al (2014)**, their study was mainly focused on statistical assessment of water quality parameters for pollution source identification in sukhnag stream, which is a major inflow stream of lake Wular, Kashmir Himalaya. Wular lake is the largest fresh water lake in India. Their study statically analyses the deteriorating water quality of the Sukhnag stream. It was carried out for a year. The parameters such as depth, transparency, temperature, pH, and conductivity were determined on the spot by water quality testing machine kit. Other parameters like Ortho phosphorus, total phosphorus, ammoniacal nitrogen, nitrite-nitrogen, nitrate-nitrogen, organic nitrogen (Kjeldahl nitrogen minus ammonical nitrogen), alkalinity, free CO2, conductivity, chloride, total hardness, calcium hardness, magnesium hardness, sodium, and potassium were determined in the laboratory within 24 hours of sampling by adopting standard methods of Golterman and Clymo (1969) and APHA (1998) [23–25]. To evaluate the significant differences among the sites for all water quality variables they used one-way analysis of variance (ANOVA) at 0.05% level of significance [28]. They performed all statistical analysis by using the software SPSS also used CV, *t*- test, ANOVA, RA, CA, and PCA to evaluate the impact of anthropogenic activities and spatio-temporal variations on physicochemical characteristics of Sukhnag stream. According to their journal stream water quality was subjected to two multivariate techniques:

cluster analysis (CA) and principal component analysis (PCA) [29]. In this study, CA showed strong spatial and temporal association on the basis of variations of principal pollution factors and indicated that the effects of human activities on water quality vary spatially as well as temporally. The dendrogram indicates pollution status as well as the effect of contamination at the sampling site. Principal component analysis was carried out to extract the most important factors and physicochemical parameters affecting the water quality. PCA correlates with total P, NO_3–N, NO_2–N, organic-N, Ca^{2+}, Na^+, TS, TSS, TDS, and free CO_2 were the 11 major factors affecting the water quality of Sukhnag stream. overloading of total-P, NO_3– N, Ca^{2+}, Na^+, TS, TSS, and TDS are responsible for the heaviest pollution problem in the stream. From multivariate analysis it could be concluded that the stream water quality is primarily influenced by agricultural runoff and wastewater discharge. The results from PCA suggested that most of the variations in water quality are explained by the natural soluble salts, nonpoint source nutrients, and anthropogenic organic pollutants. Results of regression analysis clearly showed that, in peak flow season, runoff raises the concentration of most of the inorganic and organic parameters.

- **Thukral A K. et. al (2014)**, in his journal paper, studied the characterization of change in Harike wetlands using Landsat satellite data. Their study was focused on reduction in the area of the wetland due to the pressure implemented because increasing human interventions such as urbanization, agricultural land expansion, etc. The study was carried out through data collected from Landsat satellite from 1989 to 2000 and another period extending from 2001 to 2010. The satellite images of Landsat TM and ETM of the study area were obtained from the USGS (source: http://glovis.usgs.gov/). In their study, the classification of the images was divided under five land cover types: Waterbody, Wetland I, Wetland II, Barren land and Agricultural areas. The classified images obtained were compared with GPS, topographical sheets and available wetland maps of the study area, to determine how each site represented on the ground as observed during ground surveying for verification. All image processing was performed using ERDAS Imagine 9.1, and ArcView GIS 3.2. Wetland maps were prepared in ERDAS imagine and Surfer 8. An e-Trex, Garmin Global Positioning System (GPS) receiver was used to determine the geo-coordinates of a given area in terms of its latitude and longitude. From detailed review of their image comparisons, it was noted that there was significant change in area of wetland since 1989 as well as changes were observed in the water levels during the period from 1989 to 2010. More specified results showed that the wetland classes decreased from 82% (7154 ha) in 1989 to 71% (6195 ha) in 2000 and gradually to 69% (6020 ha) in 2010 indicating towards shrinking of wetland area. A total of thirteen percent of the wetland area was lost between 1989 and 2010. Barren land and Agricultural land were the non-wetland classes, which area increased significantly from 18% (1585 ha) in 1989, increasing to 29% (2544 ha) in 2000, and subsequently to 31% (2718 ha) in 2010. This study indicated towards the decreasing of the wetland area, which will have a significant effect on the water quality status as well as the lifeforms presented in the habitat.

- **Chandra et al. (2012)** have described, lake water is a source of drinking and domestic use water for rural and urban population of India. The main goal of their study was to assess drinking water quality of various lakes i.e., Porur lake Chennai, Hussain Sager Hydrabad Vihar lake Mumbai in India. For this, lakes water samples were collected from six different sites and composite sample prepared were analyzed for pH, turbidity, electrical conductivity (EC), total dissolved solids (TDS), total alkalinity (TA), total hardness (TH)and calcium hardness (Ca-H), chemical oxygen demand(COD), biochemical oxygen demand (BOD), dissolved oxygen (D.O.), sulphate (as SO4 2-), nitrate (as NO3) and chloride (Cl-) levels .Some heavy metals like Iron, Zinc, Cadmium, Mercury, Nickel and Chromium were also analyzed in these samples. There were variations for EC (141-1041 "#$% &), turbidity (2-9 NTU), TDS (107.1–935.8 mg/L), SO4 2- (4–8 mg/L), TA (42–410 mg/L), TH (41-280 mg/L), Ca–H (14- 10 mg/L), BOD (5-9mg/L), COD (4–32 mg/L) NO3(1.1-3.6 mg/L) and Cl- (49-167 mg/L) levels at different sites. Water pollution indicates that these parameters were manifold higher than the prescribed limit by the WHO & BIS standard. Wu-Seng Lung, A. M. Asce [04] has studied, a two-layer time-variable model is developed to quantify seasonal variations of pH and alkalinity levels in acidic lakes. The model incorporates the CO2/ HCOJ/ CO5 equilibrium with internal sources and sinks of alkalinity and acidity in the water column. External alkalinity and $C0_2$ acidity loadings are also incorporated. The modeling framework is applied to the Bickford Reservoir in Massachusetts and to Woods Lake and Panther Lake in Adirondack Park, New York. In general, in-lake alkalinity generation by reduction processes in the Bickford Reservoir during the summer months is simulated by the model. The observed response to snowpack release in Woods Lake and Panther Lake during the spring months is also reproduced by the model. All three model applications are efficiently run on a personal computer system.

➢ **Akumtoshi Lkr et. al (2020)**, in their article "assessment of water quality status of Doyang river, Nagaland, using water quality index", investigated water quality index of Doyang river in three different seasons throughout a year. Sampling was done during the first week of each month from eight selected stations of the river. They categorized the months into three different seasons namely pre-monsoon (PRM), monsoon (MON) and post-monsoon (POM) for data interpretation. In this study twelve physicochemical parameters like P^H, EC, TDS, total alkalinity, total hardness, Ca^{2+}, Mg^{2+}, Cl^-, NO_3^-, SO_4^{2-}, DO, BOD were tested. After a year of long observation of various sample in different season, there were distinct variations in selected water quality parameters of different seasons. The calculation of WQI using Weighted Arithmetic Index involves the estimation of 'unit weight' assigned to each physicochemical parameter selected. Different units and dimensions of the selected parameters are transformed into a common scale using the assigning units shows the drinking water quality standards and the unit weights assigned to each parameter used for the calculation of WQI.As a result they found 25<WQI<50 for the three different seasons. WQI directly relates to the water quality status. So, their quality of water falls under the class of good water sample which is suitable for drinking, irrigation and industrial purpose.

➢ **S. N. Thitame et al.(2010)** have studied, in present investigation an attempt was made for assessment of Seasonal Variation in Physicochemical Characteristics and Quality of Pravara River Water for Irrigation during year 2008. The study reveals that most of the physicochemical parameters of river water at five selected sites show moderate variation in their concentration for all seasons. However, site 3 and 4 stands evidence of discharge of waste water from the city in the river. This intern indicated the quality of water for irrigation in the study area. The Sodium absorption ratio and Residual sodium carbonate values show good water quality for irrigation. However, at site 3 and 4 the values of Kelly's index and Soluble Sodium Percentage exceed their standards in monsoon season indicating doubtful quality of water for irrigation.

➢ **Sonzogni et al. (1986)** have studied, results from a study of water quality planning and management alternatives for the Great Lakes are used to identify cost-effective pollution control strategies. Mathematical models and other systems analysis techniques are applied to estimate pollutional loadings, specific water quality problem areas, costs and pollutant reductions offered through alternative management strategies. A determination of how these alternatives may be expected to achieve water quality objectives for the Great Lakes is made. Data from a diversity of Great Lakes research efforts are compiled, integrated, and used to project local and lake wide water quality conditions over the next twenty years. A set of management tools, including a near shore water quality index and a series of environmental quality maps, are developed to promote communication and interpretation of Great Lakes water quality data among technical and nontechnical interests. Findings from the study support a staged approach to pollution control, whereby the most cost-effective programs are implemented and their results assessed before more expensive control measures are undertaken.

- **Jadhav et al. (2012)** have explained, in order to understand the water quality of Triveni Lake, Physico-chemical parameters were studied and analyzed for the period of one year i.e., December 2010 to November 2011. Various physicochemical parameters, such as water temperature, air temperature, pH, humidity, conductivity, free Co2, total solid, dissolved oxygen, Total alkalinity, Total hardness, $CaCo_3$, Ca^{2+}, Mg^{2+} were studied. The results revealed that there was significant seasonal variation in some physicochemical parameters and most of the parameters were in normal range and indicated better quality of lake water. It has been found that the water is best for drinking purpose in winter and summer seasons.

- **Aydin Uncumusaoğlu A. (2018),** in his article "Statistical assessment of water quality parameters for pollution source identification in Bektaş Pond (Sinop, Turkey)", has researched on 21 physico-chemical and 7 heavy metal parameters obtained from four different sampling points for one year in water of Bektaş Pond by using multivariate statistical methods such as ANOVA, Pearson correlation (PC), Hierarchical cluster analysis (HCA) and Principal Component Analysis (PCA) to determine water quality as well as suitability of water for aquatic life. From his study, it was revealed that main pollution source of the pond water is non-point pollution i.e., agricultural pollution and soil leaching for this region. Result of HCA shows no significant difference between the stations but has a significant difference between seasons. A suggested solution to the problems is "best environmental practice" principle may be applied to minimize the out-of-source pollution and to efficiently use and control stocks of freshwater resources.

- **Tiri et. al (2015),** in their article “Assessment of the quality of water by hierarchical cluster and variance analyses of the Koudiat Medouar Watershed, East Algeria”, have studied about the spatial and temporal variation of water surface quality of the Koudiat Medouar Watershed, Eastern Algeria, to which end, multivariable method such as HCA and ANOVA was used. The overall evaluation during the study period showed alkaline nature of water in the area. Higher EC value was observed in water collected from sampling station 2. The ANOVA results indicate that all of the water quality parameters are significant except for Na, K and HCO_3 in the station 1 and EC in the station 2 and pH and NO_3 in the last station.

- **Nnorom et. al (2019),** in their article “Multivariate statistical approach and water quality assessment of natural springs and other drinking water sources in Southeastern Nigeria”, have investigated the physico-chemical and trace elements contents of ground and surface water sources used for domestic purposes. In this connection, multivariate statistical analysis such as ANOVA, HCA etc., were used to characterize potentially toxic elements (PTEs) in natural springs and other drinking water sources based on likely origin, inter relationships, extent of involvement in water contamination and PTE’s ability to cause harm to unsuspecting consumers. The result showed Fe levels were above its permissible limit in about 92% of samples from streams. Overall, Fe, Al, Mn, Se and Zn were the dominant elements. Moreover, Water Quality Index approach indicated that all drinking water sources had either excellent or good water quality with the exception of a borehole, which had poor water quality.

CHAPTER 3

WATER QUALITY PARAMETERS

3.1 INTRODUCTION:

Water quality refers to the physical, chemical and biological characteristics of water in relationship to a set of standards (Wikipedia). For utilization of water for various purposes, the used water should meet certain requirements in its standard. Thus, the suitability of water to support and sustain the requirements as well as various processes is termed as water quality. Quality analysis of water is a very important tool or monitoring and updating of the limits of water parameters for proper maintenance of environmental balance, for example limits on the concentrations of toxic substances for drinking water use, or restrictions on temperature and pH ranges for water supporting invertebrate communities. Generally, surface water and ground water sum up the available water resources. The water quality of different location available on earth may not be same. The deterioration or variation in limits of the quality of water may occur due to natural factor as well as human influences. The most important of the natural influences are geological, hydrological and climatic, since these affect the quantity and the quality of water available. The purest form of water is considered to be rain water. But, during precipitation, rain water acquires dissolved gases from the atmosphere and gets polluted. Rain water after precipitation comes in contact with the impurities present on the ground and it percolates underground resulting in contaminating the ground water quality. Moreover, during volcanic eruptions under water some poisonous gases may come out and thus causing variation in the parameters of water thereby affecting the water quality. In arid and coastal regions, high salinity of water is a natural problem.

Human activities are the worst factor affecting the water quality. Human activities contaminate both surface water and ground water through widespread interference in all aspects of nature without taking proper preventive measures to maintain the quality of available water resources. Examples of human activities worsening the water quality involves dumping of industrial and municipal wastes frequently, agricultural and residential activities which involves fertilizers, pesticides and animal wastes, leaking of fuel storage tanks, landfills road salt and more.

Deterioration of water quality will have equal impact on both aquatic ecosystem as well as the surrounding ecosystems including human beings. So, complete and detail assessment of the water quality parameters must be done to check the standard limits and preservation of the water quality through measures.

3.2 WATER QUALITY STANDARDS:

In the setting of standards, agencies make political and technical/scientific decisions about how water will be used. In case of natural water bodies, they also make some reasonable estimate of original and fresh conditions. Water quality in this project is used with reference to a set of guidelines values and standards set by World Health Organization (WHO) and Bureau of Indian Standards (BIS).

3.3 WATER QUALITY PARAMETERS:

The parameters defining, he quality of water can be sectioned into different categories or types. A brief explanation about their characteristics and measurement techniques is put forwarded as follows: -

1) Physical parameters.

2) Chemical parameters.

3) Biological parameters.

3.3.1 PHYSICAL PARAMETERS:

The important physical parameters of water include turbidity, colour, odour, temperature, electrical conductivity (EC), total dissolved solids (TDS), viscosity, specific weight and vapour pressure.

Temperature:

Temperature measurement in water does not directly imply to whether water is polluted or not. But, temperature of water affects some of the important properties and characteristics of water such as density, specific weight, viscosity, surface tension, solubility of dissolved gases and etc. Chemical and biological reaction rates increase with increasing temperature. Reaction rates usually assumed to double for an increase in temperature of 10 °C. Dissolved oxygen is indirectly related to temperature, as temperature increases D.O of water decreases. Increase in temperature increases the growth rate of aquatic microorganism which leads to higher consumption of consume dissolved

O_2 and level of dissolved O_2 decreases. Moreover, temperature also affects disinfection process because efficiency of disinfection is lower at lower temperature.

Colour:

Pure water is colourless. Appearance of any colour in water points towards presence of polluted materials in water. Natural water system is often colored by foreign material. Colour in water is imparted mainly due to dissolved materials and suspended materials. If the colour is due to suspended material, it is called as apparent colour. Colour given by dissolved material that remains even after removal of suspended material is called true colour or real colour. The maximum acceptable level of colour in water is 15 TCU (True Colour Unit). Objections to high colour are generally on aesthetic grounds rather than on the basis of a health hazard.

Odour and Taste:

Drinking water should be tasteless and odourless. Odour in water indicates water pollution. Odour and taste may develop in water due to natural and anthropogenic effects. Disinfection process of water might give out some taste and odour in water supply. Some natural impurities dissolved in water can also give taste and odor. Inorganic salts such as NaCl, KCl, etc. dissolve in water give taste whereas compounds like H_2S can give both taste and odour.

Odour is measured in T.O.N (Threshold Odour Unit). For, drinking standards odour should be less than 3 T.O.N.

Turbidity:

Turbidity is opaqueness of water. It is the measure of resistance offered by the suspended particles to the passage of light through water. the degree of which water losses its transparency due to presence of suspended particulates. Water will be rate as highly turbid if the concentration of suspended particles is more. Turbidity is often used as an indicator to check quality of water based on clarity and concentration of suspended solids in water. Presence various factors increases the turbidity of water such as- suspended sediments like clay or silt, re-suspended sediments from bottom of a water body, dumped waste and other discharges, phytoplankton, algal growth etc.

Turbidity is a surface phenomenon. The sunlight is reflected and cannot penetrate deep due to the presence of suspended matters as a result it reduces the photosynthetic activities of certain plant species underneath. Moreover, the suspended solid particles absorb the heat from the sunlight, which makes the turbid water warmer, and so reducing the concentration of oxygen in the water, which directly affects the survival of the aquatic organisms. Turbid water appears to be murky, cloudy and dirty from esthetic view. For proper disinfection removal of turbidity of water is important.

Earlier, turbidity is measured on silica scale and expressed in terms of Turbidity Units (T.U). Jackson Turbidity Unit was used which work on the principle of light absorption. Nowadays, Instrument which is used to measure turbidity is called 'Nephelometer'. Nephelometer is a digital electronic device which works on the principle of light scattering and measure very low turbidity in water with high degree of precision instantly. It measures turbidity in terms of Nephelometric Turbidity Units (N.T.U). The instrument always takes into account the amount of light scattered perpendicular to the incident ray (90 degrees to the light path) to measures turbidity.

1 N.T.U = 1mg of formazin polymer mixed in 1 litre of distilled water, then the turbidity produced is 1 N.T.U

Total Dissolved Solids (TDS):

Total Dissolved Solids is the measure of the mass of solid material dissolved in a given volume of water. TDS is measured in grams per litre. Total dissolved solids include inorganic salts (mainly salts of calcium, magnesium, sodium, bicarbonates, chlorides ad sulfates) ad dome small amounts of organic matter that are dissolved in water. The sources of TDS comprise of natural sources, urban run-off, industrial wastewater, sewage, chemicals in water treatment process etc. TDS test provides a qualitative measure of the amount of dissolved solids present in the water sample. The presence of TDS is not a health hazard but concerned with the aesthetics of water. The limit for drinking water standards of TDS is set for not greater than 500 mg per litre. But a higher TDS level indicates the presence of certain cations and anions in relation with the water quality parameters:

(i) Cations combined with carbonates of $CaCO_3$, $MgCO_3$ etc. indicates hardness, scale formation, bitter taste in water.

(ii) Cations combined with Chlorides like NaCl, KCl results in salty or brackish taste and increase corrosivity of water.

TDS test is done generally through two methods: Gravimetric analysis and Electrical conductivity.

Gravimetric analysis is the most accurate methods and involve evaporating the liquid solvent ad measuring the mass of residues left. Although, this method is very time consuming, but it provides more reliable results. This method is applicable for measurement of total dissolved solids I all natural waters, in raw, process and treated agricultural, municipal and industrial wastewaters and in treated drinking water. Electrical conductivity is a measure of the capacity of water to conduct electrical current. Conductivity is directly related to the concentration of salts dissolved in water and thereby directly linking to the Total Dissolved Solids. The measure of TDS I the field directly is difficult, so its measurement is done through the conductivity method. Electrical conductivity is a fast method which ca be measured using a conventional conductivity meter or TDS meter.

Electrical conductivity ca be related to TDS with the help of the following equation:

$$TDS = k_e.EC$$

where, TDS is expressed in mg/litre.

EC is measured I terms of micro-siemens per cm.

k_e is the correlation factor which ranges from .55 to .80.

3.3.2 CHEMICAL PARAMETRS:

Water quality is most affected by the chemical matters preset in it. Chemical parameters may include organic and inorganic matters. Some of the concerned chemical parameters include water pH, Total Hardness, Calcium Hardness, Magnesium Hardness, Alkalinity, presence of Chloride, Fluoride, Iron, Arsenic, Lead, Nitrate etc. The chemical parameters which are investigated in this study are listed below:

Hydrogen-Ion Concentration (pH):

The pH is a quantitative measure of the hydrogen ion concertation of water indicating the measurement of the acidity or alkalinity of a solution. The pH scale generally rages from 0 to 14.

The pH scale indicates:

water is acidic if pH is less than 7

water is neutral if pH is equal to 7

water is alkaline if pH is greater than 7

pH is calculated as the negative logarithm of the hydrogen ion concertation, i.e., $pH=-log_{10}[H]$. It is measured in units of moles per litre, of hydrogen ions. The normal rage for pH in drinking water is between 6.5 to 8.5 (as per IS 10500:2012). The pH of pure water is considered to be 7. The effect of pH is not direct o to our health. When pH level is less than 6.5 it increases acidity resulting in metallic or sour taste of drinking water, blue-green staining of sinks and other household fixtures. Moreover, with increase of pH indicates alkaline water which results in scale buildup in household plumbing.

pH can be measured by two techniques: colorimetric and potentiometric. The colorimetric method involves adding a suitable indicator to a solution ad matching the colour of the solution to a standard solution containing the same indicator. But potentiometric method is more accurate. This method uses a pH meter to determine hydrogen ion concentration having two electrodes. The electrodes potential of the indicator electrode is linearly related to changes in hydrogen ion concentration and thus pH is known.

Total Hardness (TH):

The characteristics of water that prevents formation of lather or foam with soap is termed as the harness of water. Water which has high dissolved minerals in it, generally calcium and magnesium is considered hard. As water moves through soil and rock, it dissolves very small amounts of minerals and holds them in solution. The degree of hardness becomes greater as the calcium and magnesium content increases in the water. Hardness of water is generally of two types:

(i) Carbonate or temporary hardness: The bicarbonates and carbonates of calcium and magnesium usually causes this type of hardness. Temporary hardness can be removed to some extent by boiling or removed fully by addition of lime.

(ii) Non-carbonate or permanent hardness: Permanent hardness is primarily caused by presence of calcium chloride, calcium sulphate, magnesium chloride or magnesium sulphate. It cannot be removed by boiling, and some special treatment is required for its removal. Non- carbonaceous hardness can be removed by water softening methods such as lime soda process, demineralization process and zeolite process.

Total hardness is the sum of the carbonate harness and non-carbonate hardness. It is measured in terms of parts per million(ppm) or mg/litre of $CaCO_3$.Water is considered soft when the concentration of $CaCO_3$ is below 60 mg/l; moderately hard when between 60-120 mg/l; hard when between 120-180 mg/l and very hard when more than 180 mg/l. Hard water is not a health hazard but it has serious impacts on household items. Water hardness causes damages to boilers, cooling towers and other equipment that handles water. Hard water can cause mineral build-up in water pipes and eventually clog them. Total hardness is measured by EDTA (Ethylene Diamine Tetra-acetic Acid test).

Alkalinity:

The alkalinity refers to the measure of the capacity of the water to neutralize the acids. Alkalinity of water may be due to the presence of one or more of a number of ions. These include hydroxides, carbonates and bicarbonates. Most alkalinity in surface water comes from calcium carbonate (CaCO3) that come from rocks and soil. Limestone contains high level of calcium carbonate. The process is enhanced if the rocks and soil have already been broken up before entering the water. The dissolved minerals get into the water through construction and other processes. In simple terms, the pH of a solution is a measure of how strong the bases are in a solution, whereas the alkalinity measures the amount of chemical bases present in the solution. Alkalinity is determined through titration. It is usually measured in unit of mEq/L (milliequivalent per litre).

Normal drinking water generally has a neutral pH of 7. Alkaline water typically has a pH of more than 8. But pH alone isn't enough to impart substantial alkalinity to water. Although alkaline drinking water is considered safe, it has some side effects including the lowering of natural stomach acidity, which helps kill bacteria and expel other undesirable pathogens from entering your bloodstream. Additionally, an overall excess of alkalinity in the body may cause gastrointestinal issues and skin irritations. Alkalinity is usually measured in unit of mEq/L (milliequivalent per litre).

Chloride (Cl):

Chlorides are salts resulting from the combination of the gas chlorine with a metal. Some common chlorides include sodium chloride (NaCl) and magnesium chloride ($MgCl_2$). Chloride exists in all natural waters, the concentrations varying very widely and reaching a maximum in sea water (up to 35,000 mg/l Cl). In fresh waters the sources include soil and rock formations, sea spray and waste discharges. Chloride contents are very high in sewage and industrial effluents. Chloride does not pose a health hazard to humans under standard limits. Public Drinking Water Standards require chloride levels not to exceed 250 mg/L. Water will begin to taste salty and will become increasingly objectionable as the concentration level rises further above 250 mg/L. Chlorine alone as Cl_2 is highly toxic and it is often used as a disinfectant. In combination with a metal such as sodium it becomes essential for life. Small amounts of chlorides are required for normal cell functions in plant and animal life. Criteria for protection of aquatic life require levels of less than 600 mg/L for chronic (long-term) exposure and 1200 mg/L for short-term exposure.

Concentration of chloride is measured by titration by Mohr’s method.

Dissolved Oxygen (D.O):

Dissolved oxygen is the amount of oxygen that is dissolved in water. D.O is one of the most important factors that determines the survivality of the aquatic organisms. Dissolved oxygen is different from the oxygen that is present in water molecules. Only about ten molecules of oxygen per millions of water is actually dissolved in water. This dissolved oxygen is breathed by fish and zooplankton and is needed by them to survive. Oxygen is dissolved in water from atmosphere through direct absorption, from areas where groundwater discharges into streams or as a waste product of plant photosynthesis. Temperature affects the formation of dissolved oxygen. Dissolved oxygen decreases in water with increases of temperature. Moreover, dissolved oxygen is also affected by movement of water. Rapidly moving water, such as in a mountain stream or large river, tends to contain a lot of dissolved oxygen, whereas stagnant water contains less. The standard level of dissolved oxygen in water is 4 mg/L. Dissolved oxygen level is important water quality indicator. Measurement of level of D.O in water samples in done through Wrinkler's Iodometric method (titration).

D.O levels in water is lowered mainly due to point source pollution, that results in decomposition by bacteria and nutrient pollution, that leads to excess plant and algal growth. Sufficient D.O levels in water is a primary factor for the all forms of living organism in water. If dissolved oxygen concentrations drop below a certain level, fish mortality rates will rise. In the ocean, coastal fish begin to avoid areas where DO is below 3.7 mg/L. Even benthic organisms show reduced growth and survival rates on falling of D.O levels to 1 mg/L.

Biological Oxygen Demand (B.O.D):

Biochemical oxygen demand (BOD) represents the amount of oxygen consumed by bacteria and other microorganisms while they decompose organic matter under aerobic conditions at a specified temperature. Biological oxygen demand (BOD) generally represents how much oxygen is needed to break down organic matter in water. Measurement of BOD is used as an index of the degree of organic pollution in water. BOD directly affects the amount of dissolved oxygen in rivers and streams. The greater the BOD, the more rapidly oxygen is depleted in the stream. This means less oxygen is available to higher forms of aquatic life. The consequences of high BOD are the same as those for low dissolved oxygen: aquatic organisms become stressed, suffocate, and die. Sources of BOD include leaves and woody debris; dead plants and animals; animal manure; effluents from pulp and paper mills, wastewater treatment plants, feedlots, and food-processing plants; failing septic systems; and urban storm water runoff.

B.O.D is measured by a test in which amount of oxygen consumption is determined from a sample at 5 days period at a temperature of 20^0C. Light must be excluded from the incubator where the sample will be placed for 5 days, to prevent algal growth that may produce oxygen in the bottle.

B.O. D in mg/L = $(D.O_i - D.O_f) \times D.F$

Where, DO_i and DO_f are the initial and final concentrations of dissolved oxygen in mg/l.

D.F is the dilution factor.

3.3.3 BIOLOGICAL PARAMETERS:

Biological parameters of water are important water quality testing factors. Biological parameters of water indicate the presence of microbiological organisms and pathogens in water. From, health prospective, biological parameters are more important to test for, then physical and chemical parameters. Presence of these micro-organisms in water can cause deterioration of health when consume directly. These organisms that affect the quality status of water generally includes bacteria, protozoa, virus and algae.

DRINKING WATER QUALITY STANDARDS

Table 3.1: Standards for water quality parameters as per BIS (IS- 10500:2012)

Sl. No.	**Water quality parameter**	**Desirable limit**	**Maximum permissible limit**
1	Hydrogen-ion concentration (pH)	6.5 – 8.5	6.5 – 8.5
2	Turbidity (NTU)	1.0	5.0
3	Alkalinity (mg/l)	200	600
4	Total Hardness (mg/l)	200	600
5	Iron (mg/l)	0.3	1.0
6	Nitrate (mg/l)	<4.5	4.5
7	Chloride(mg/l)	250	1000
8	Fluoride (mg/l)	1.0	1.5
9	Sulphate (mg/l)	200	400
10	Arsenic (mg/l)	0.01	0.05
11	Lead (mg/l)	< 0.01	0.01

12	Zinc (mg/l)	5	15
13	Nickel (mg/l)	0.02	No relaxation
14	Mercury (mg/l)	0.001	No relaxation

CHAPTER-4

DEEPOR BEEL, GUWAHATI- THE STUDY AREA

4.1 INTRODUCTION

Deepor Beel is a permanent fresh water lake and one of the largest Beel in the Brahmaputra valley. It is located about 10 km south-west of Guwahati city, in Kamrup district of Assam, India. Deepor Beel is the only Ramsar site in Assam and among the third Ramsar site of the north eastern region of India. It was included in the list of Ramsar Site in November 2002, undertaking conservation measures on the basis of its biological and environmental importance. Its basin is drained by a system of rivulets and hill streams that connect the neighboring hills and the forests to the river Brahmaputra through an outlet called the Khanajan.

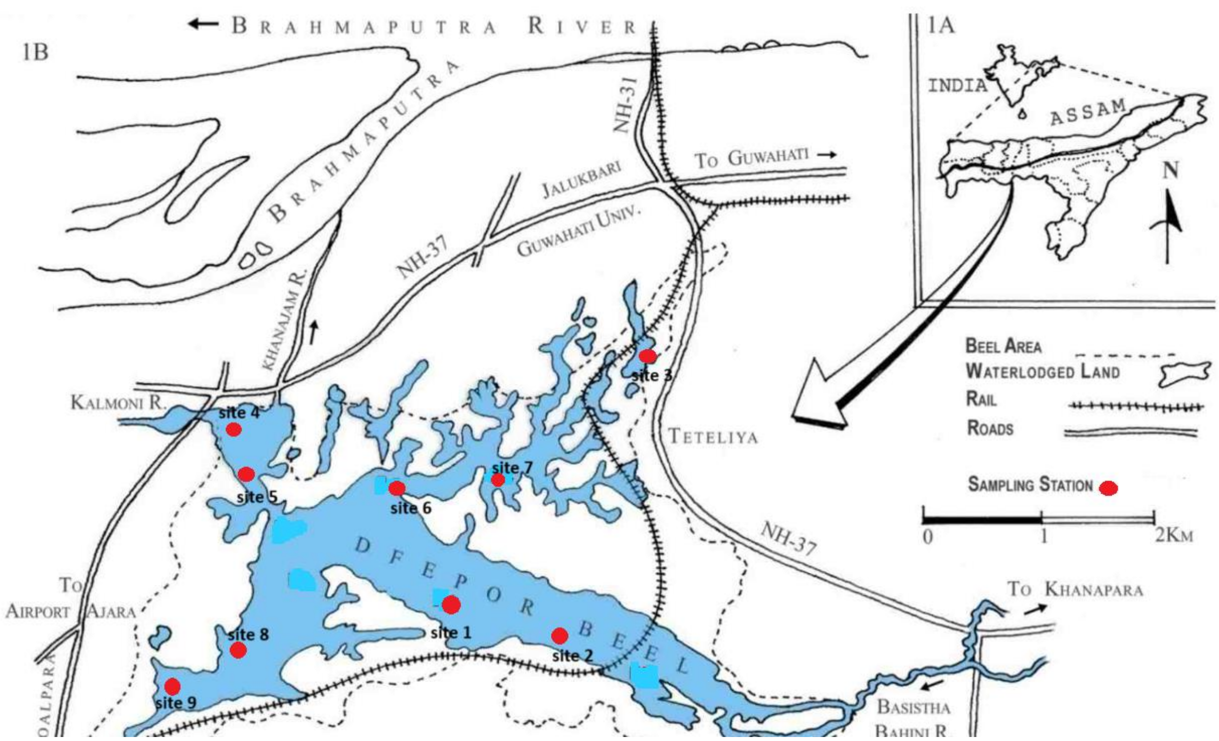


Fig 4.1: Map of Assam showing Deepor Beel

4.2 LOCATION AND BOUNDARIES

The Beel is located between latitude: 26° 05'- 26°11' N and longitude: 91°35' - 91°43' E, it covers an area of 40.14 sq.km. The northern-eastern side of the Beel is thickly populated and is encircled by various government institutions like Guwahati University, Assam Engineering College, Assam Ayurvedic College, and Forest School. The national highway 37 (NH-37) is located in the northern and north-western side of the Beel and touches its periphery at different places like Dharapur, Azara etc. It is bounded by the PWD road, northern fringe of the Rani and Garbhanga Reserve Forests on the south.

4.3 PHYSIOGRAPHY

The Deepor Beel is located in a U-shaped valley bounded by steep highlands in the north and south side of the Beel. The Deepor Beel and its fringe areas are made up of recent alluvium consisting of clay, silt, sand and pebbles whereas the hills in the north and south side of the Beel's are of Archaean age (envfor.nic.in/divisions/csurv/WetlandInventory.pdf). The wetland receives most of the surface runoff from the nearby hills which is one of the reasons of sedimentation of the wetland. Deposited soil in the bed of the wetland is the cause of lowering the depth of the Beel. It is commonly stated that the Beel together with those adjoining it are an abandoned channel of the Brahmaputra system.

4.4 CLIMATE

Deepor Beel has a meso-thermal climate, characterized by high humidity and moderate temperature (Singh & Dutta 1960). The temperature ranges between 10.6°C to 30°C. The annual average precipitation is 3000 to 4000 mm. Most of the rainfall occurs during monsoon period (May-September). The monsoon season (May -September) has a maximum temperature of 32°C and minimum of 27.3°C. The pre-monsoon season (March-May) has a maximum temperature of 27° C and minimum of 24° C, and relative humidity between 50.5-76.8%. The relative humidity is 82.5%. Warm humid and cloudy weather is characteristics for this season. The retreating monsoon covers the period from September to October with maximum and minimum temperatures of 27° and 25° C respectively. The relative humidity is 82% and the rainfall gradually decreases to average as the season advances, when the morning mist and fogs start appearing. The winter season begins in November and continues until January. The average field temperature during this period remains at 20 ± 2°C and the relative humidity measures about 77.5%. This season also experiences occasional rainfall due to the west monsoon.

4.5 HYDROLOGY

Basistha and Kalmani rivers and monsoon run off are the major sources of water for the wetland. In the rainy season the depth of the Beel increases up to four meter while in the dry season the depth drops to one meter. Khonajan channel drains the Beel into the Brahmaputra River, 5 km to the north. It acts as a natural stormwater reservoir during the monsoon season for the Guwahati city. Although the Beel is perennial, the watery area of the Beel decreases to 10.1 sq.km during dry season and increases to 40.14 sq.km during monsoon season.

4.6 RESOURCES OF DEEPOR BEEL

Deepor Beel is one of the richest biodiversity areas within the wetland ecosystem of Assam. Its partially deep water and partially shallow water as well as the presence of high land support large numbers of plant and animal species. Again, the presence of hilly terrain and natural forest adjoining the Beel area support large numbers of endangered and threatened vertebrate species. Deepor Beel with its rich ecological system, provides habitat to a large number of bird species. 219 species of birds including more than 70 migratory species are reported in the Beel area. During winter season, various migratory bird species are seen to be habituating in the Deepor Beel. Some of the globally threatened species of birds like spot-billed pelican, lesser adjutant stork, Baer's pochard, Pallas' sea eagle, greater adjutant stork. Among the large number of migratory water fowl, the Siberian crane regularly migrates to this habitat during its annual journey (Wikipedia). During the summer, large parts of the Beel are covered by aquatic vegetation like water hyacinth; aquatic grasses, water lilies and others sub merged, emergent and floating vegetation. Moreover, a large number of fishes are also found in this wetland. The scattered forest present within the Beel area supports a large variety of lizard species.

4.7 ENVIRONMENTAL PROBLEMS OF DEEPOR BEEL

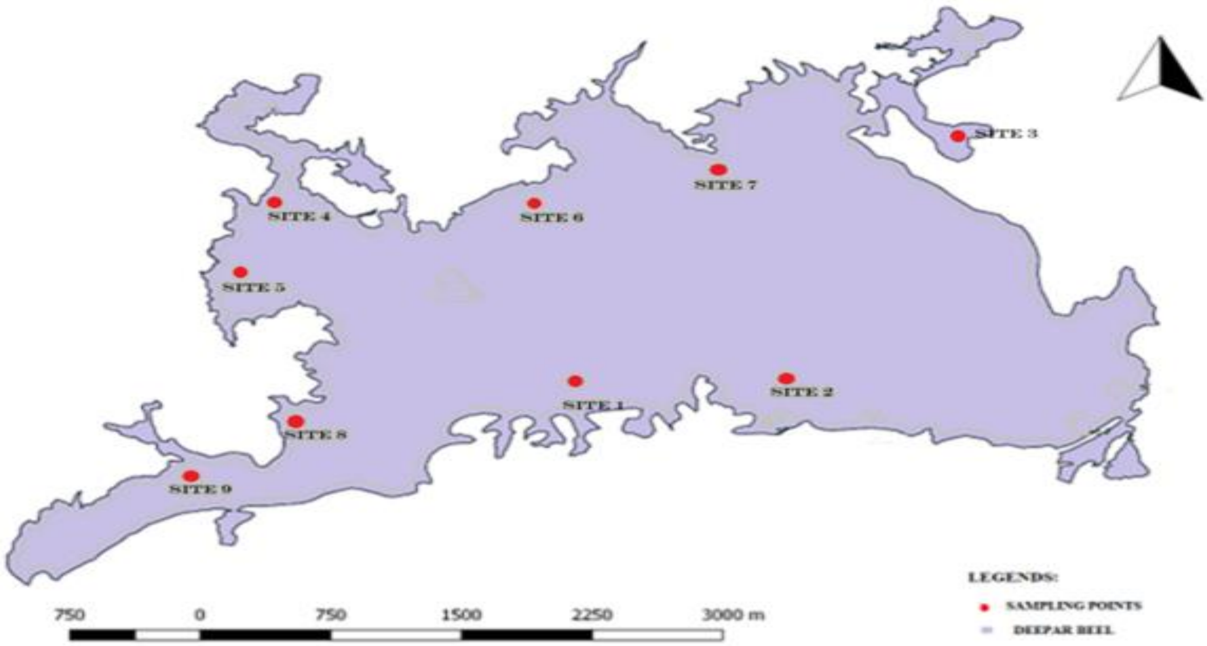
The Northeast Frontier Railway (NFR) constructed a railroad along the southern boundary of the Deepor Beel in 2001. The rail-line across the wetland with high embankment has affected the entire ecosystem. Due to the Construction of the rail-line, there has been blockage in water flow system. The rail traffic will definitely increase in future for the growth achieved by Guwahati city and for the large quantity of resources availability in the region such as, mineral resources (coal, and oil), agricultural products (rice, tea, fruits, and spices) and forest products (timber and plywood) etc. which will have a significant impact on the wetland system. Another notable cause for the wide spread deterioration of the quality of water of Deepor Beel is the Guwahati Municipality dump yard (24 Ha) located in Boragaon, which lies in the eastern corner of Deepor Beel. The dump yard becomes functional in 2005, three years after Deepor Beel's status as Ramsar Site. Women and children from the adjacent slum areas of Boragaon, are seen in the dump yard for collection of plastics and other products.

CHAPTER-5

METHODOLOGY

5.1 SELECTION OF SITES

Study area Deepor Beel is in the Kamrup district, situated at the South Western part of Assam, a North Eastern State in India. The lake covers an area of 40 km^2 in the South Western outskirts of the city of Guwahati—the state capital of Assam. The lake is located between 26°06′02″N to 26°08′34″N and 91°36′29″E to 91°42′24″E (NWA, 2013).



Site No. (Name)	GPS Locations (Latitude, Longitude)	Site No. (Name)	GPS Locations (Latitude, Longitude)	Site No. (Name)	GPS Locations (Latitude, Longitude)
Site 1 (Near ASTU)	(26°08′25.43″ N,91°39′01″ E)	Site 4 (Chakardeo village)	(26°06′46.29″ N,91°39′16.41″ E)	Site 7 (Boragaon dumping site (ii))	(26°06′36.05″ N,91°38′39″ E)
Site 2 (Near MCA building, AEC)	(26°08′30.48″ N, 91°39′44.58″ E)	Site 5 (Mikir Para, Rani)	(26°08′8.24″ N, 91°39′16.41″ E)	Site 8 (Near GIMT)	(26°07′21.62″ N, 91°37′44.97″ E)
Site 3 (Near Tetelia)	(26°06′44.04″ N, 91°39′12.82″ E)	Site 6 (Boragaon dumping site (i))	(26°06′57.56″ N, 91°40′35.48″ E)	Site 9 (Azara)	(26°08′11.48″ N, 91°40′5.63″ E)

Fig 5.1: GPS location of Sampling sites of Deepor Beel.

The above figure (Fig 5.1) shows the output results obtained from georeferencing the map of Deepor Beel in the ArcGIS software. Points marked red in colour were selected for pointing the stations in their preferred coordinates within Deepor Beel area using necessary steps in ArcGIS software.

5.2 COLLECTION AND PRE-TREATMENT OF SAMPLES

Samples were taken from Deepor Beel in two months (October 2022, January 2023). Nine fixed sampling locations were used consistently throughout the study (Fig. 5.1). The sampling points were initially selected with eye estimation in such a way that they may spread throughout the mid-range of the entire surface of the lake. Samples were taken in the middle of 15th to 20th day of a month from a depth of 1 m to ensure a regular sampling pattern. Sampling procedures were performed from 09:00 AM to 11:00 AM for minimizing the influence of fluctuation in water quality throughout the day. Some physico-chemical parameters like Dissolved Oxygen (DO), Total Dissolved Solids (TDS), Electrical Conductivity (EC), pH and Temperature (T), salinity were estimated in situ with Multiparameter Water Quality Analyzing device. Other parameters like Total Hardness (TH), chloride content (Cl), Chemical Oxygen Demand (COD) and Biochemical Oxygen Demand (BOD) were estimated in laboratory. Samples were collected in 1-L Sample Bottles and kept in ice box immediately after collection. The bottles were brought in the laboratory within 10 h after collection and preserved in refrigerator to estimate the parameters on the next day. In this study World Health Organization (WHO) (BIS, 2012) and Indian Standard 10500: 2012 (IS 10500: 2012) Guidelines (Moharana et al., 2014; WHO, 2008) were followed for the permissible limits of the Water Quality Parameters (WQP).

5.3 ANALYTICAL METHODS OF TESTING

5.3.1 Determination of temperature:

Measuring of temperature is important because of its effect on other testing parameters**.**

Procedure:

Measured with the help of mercury thermometer, graduated between 0º - 100º C at the time of collection of the samples.

5.3.2 Determination of Colour:

Procedure:

Colour of the samples is noted through direct visual observation at the time of collection of samples.

5.3.3 Determination of Odour:

Procedure:

Noted through smelling of the samples immediately after collection.

5.3.4 Determination of Total Hardness by Titration:

Methodology as per IS:3025 (Part 21)- Reaffirmed 2007

Apparatus:

1. Measuring cylinder
2. Burette
3. Pipette
4. Glass rod
5. Conical flask

Reagents:

1. Ammonium buffer solution
2. Eri-chrome black-T indicator
3. Standard Ethylene Di-amine Tetra Acetic Acid (EDTA) titrant.

Procedure:

(i) 25 ml of sample is taken in the conical flask and another 25 ml of water is pipetted to the flask to make a solution of 50 ml.

(ii) 2 ml of Ammonium buffer solution is added to the solution.

(iii) 1 to 2 ml of eri-chrome Black-T indicator is added and the colour of the solution turns to light red/pink.

(iv) Now, the burette is filled with EDTA.

(v) EDTA is added to the solution in the flask, titrated until the pink colour changes to sky blue.

Calculation:

Total hardness (mg/L) = (Burette reading x 1000)/ ml of sample.

5.3.5 Determination of Chloride by titration (Argentometric method):

Methodology as per IS:3025 (Part 32)- Reaffirmed 2007

Apparatus:

1. Measuring cylinder
2. Burette
3. Conical flask
4. Pipette

Reagents:

1. Silver nitrate ($AgNO_3$), 0.02 N
2. Potassium chromate (K_2CrO_4), 5%

Procedure:

(i) 50 ml of sample is taken in a conical flask.

(ii) 2 ml of K_2CrO_4 solution is added the sample and the solution will turn to yellow colour.

(iii) Now, the burette is filled with $AgNO_3$ solution.

(iv) $AgNO_3$ solution from burette is allowed to fall in to the conical flask, titration is continued till the yellow colour of the solution turns to brick colour.

Calculation:

Chloride (mg/L) = ((ml x N) of $AgNO_3$ x 1000 x 35.5)/ ml of sample.

5.3.6 Determination of Biological Oxygen Demand (B.O.D):

Methodology as per IS: 3025 (Part 44)- 1993.

Apparatus:

1. Burette
2. Measuring cylinder
3. B.O.D bottle
4. Conical flask
5. Pipette

Reagents:

1. Phosphate buffer solution
2. Magnesium sulphate solution
3. Calcium chloride solution
4. Ferric Chloride solution

Procedure:

(i) One BOD bottle is filled with the sample to determine the initial D.O of the water sample.

(ii) Another BOD bottle of the sample is kept for incubation for 5-day period at a temperature of 20^0C.

(iii) Prepare four blanks by siphoning out dilution water directly into the bottles.

(iv) Initial D.O of two bottles is determined and remaining two bottles are kept for incubation at 20^0C for 5 days.

(v) After 5 days, final D.O of the incubated bottles is determined.

Calculation:

$$B.O.D\ (mg/L) = (D.O_i - D.O_f) \times D.F$$

CHAPTER – 6

RESULTS: WATER QUALITY ANALYSIS

6.1 RESULT

The different water quality parameters were tested as per the methodology that were described in the previous chapter. Testing of the water quality parameters were done season-wise i.e., Autumn (October, 2022 – December, 2022), winter season (January, 2023 – March, 2023), spring season (April, 2023 – June, 2023) and summer season (July, 2023 – September, 2023). The results obtained in this testing have been presented in both tabular and graphical form in this chapter. The variation in the parameters in different season has been noted and discussed in details. With the results obtained from the test, the water quality status of the study area will be known.

6.1.1 Tabular output of results

The different parameters that were analyzed along with their results for the specified months are presented in a tabular form below:

Table 6.1: Concentrations of Water Quality Parameters of the sampling sites for Autumn Season

Water quality parameters	*SITE 1*	*SITE 2*	*SITE 3*	*SITE 4*	*SITE 5*	*SITE 6*	*SITE 7*	*SITE 8*	*SITE 9*
Temperature (˚C)	24.8	25.5	25.2	23	22.5	24	23.4	24.2	23.5
pH	7.2	7.35	7.42	6.25	6.45	6.92	6.97	6.78	6.59
Dissolved oxygen (mg/l)	5.6	5.4	3.5	4.89	4.1	3	3.6	4.8	3.8
Total Dissolved Solids (mg/l)	190	187.5	196.2	180	108	155.6	156.2	70	122.7
Salinity (ppt)	0.271	0.295	0.289	0.152	0.130	0.232	0.231	0.140	0.180
Electrical Conductivity (ms/cm)	0.365	0.234	0.387	0.165	0.175	0.309	0.312	0.155	0.238
Biological Oxygen Demand (mg/l)	1.3	1.9	2.47	2.1	2.7	3.6	3.1	2.03	2.3
Total Hardness (mg/l)	115	90	120	125	130	110	105	112	85
Chloride (mg/l)	100	105	60	95	50	30	35	75	80
Iron (mg/l)	0.95	0.53	0.60	1.1	1.45	2.52	2.78	1.03	1.12
Nitrate (mg/l)	0.007	0.009	0.045	0.03	0.08	0.15	0.20	0.06	0.01
Lead (mg/l)	0.022	0.095	0.02	0.04	0.08	0.12	0.19	0.027	0.03

Turbidity **(**NTU)	10	11	10	9	7	12	13	6.5	6

The above table is a demonstration of the concentration of the 13 different water parameters for the 9 different sampling locations. The values listed are based on laboratory analysis of the water samples for the first period of our study.

Table 6.2: Concentrations of Water Quality Parameters of the sampling sites for Winter Season

Water quality parameters	***SITE 1***	***SITE 2***	***SITE 3***	***SITE 4***	***SITE 5***	***SITE 6***	***SITE 7***	***SITE 8***	***SITE 9***
Temperature (˚C)	21.5	23	24.2	22	22.5	23.3	23	23.8	22.5
pH	6.94	7.15	7.25	5.85	6.3	6.5	6.78	6.2	6.1
Dissolved oxygen (mg/l)	6.2	5.7	4.2	5.1	6	4.2	3.9	5.2	4.5
Total Dissolved Solids (mg/l)	235	205	214.8	195	145.6	169.5	141	110	179.2
Salinity (ppt)	0.294	0.350	0.275	0.220	0.142	0.200	0.365	0.155	0.128
Electrical Conductivity (ms/cm)	0.405	0.245	0.355	0.205	0.178	0.292	0.325	0.207	0.250
Biological Oxygen Demand (mg/l)	1.05	1.5	2.3	1.75	1.1	2.25	4.8	1.7	1.97
Total Hardness (mg/l)	125	95	130	135	140	120	115	122	90
Chloride (mg/l)	90	110	75	70	45	35	50	70	65
Iron (mg/l)	0.97	0.63	0.68	1.22	1.49	2.6	2.83	1.03	1.12
Nitrate (mg/l)	0.005	0.007	0.01	0.025	0.04	0.13	0.17	0.038	0.009
Lead (mg/l)	0.01	0.049	0.016	0.037	0.063	0.095	0.099	0.025	0.028
Turbidity(NTU)	10	10	9	8	10	15	14	6	5

The above table is a demonstration of the concentration of the 13 different water parameters for the 9 different sampling locations. The values listed are based on laboratory analysis of the water samples for the next period of our study i.e., from Jan’23- Mar’23.

Table 6.3: Concentrations of Water Quality Parameters of the sampling sites for Spring Season

Water quality parameters	***SITE 1***	***SITE 2***	***SITE 3***	***SITE 4***	***SITE 5***	***SITE 6***	***SITE 7***	***SITE 8***	***SITE 9***
Temperature (˚C)	29	28.4	29	29.5	28.5	29.4	28	29.2	28.6
pH	7.04	7.2	7.3	6.12	6.5	7.12	7.09	6.5	6.45

Dissolved oxygen (mg/l)	5.9	5.4	3.9	4.95	5.85	4.06	3.87	5	4.16
Total Dissolved Solids (mg/l)	240	213	220.8	198	159.3	170	143	115	184
Salinity (ppt)	0.304	0.275	0.32	0.386	0.182	0.204	0.372	0.164	0.130
Electrical Conductivity (ms/cm)	0.397	0.284	0.350	0.216	0.198	0.297	0.365	0.208	0.257
Biological Oxygen Demand (mg/l)	1.12	1.54	4.35	1.8	1.23	3.48	5	1.7	2
Total Hardness (mg/l)	123	90	127	131	136	112	105	114	86
Chloride (mg/l)	95	119	77	74	47	39	55	73	69
Iron (mg/l)	0.96	0.59	0.62	1.15	1.39	2.55	2.65	1	1.04
Nitrate (mg/l)	0.006	0.008	0.018	0.03	0.05	0.18	0.2	0.043	0.01
Lead (mg/l)	0.009	0.043	0.01	0.038	0.032	0.085	0.090	0.012	0.015
Turbidity(NTU)	9	10	10	8	10	14	13	7	7

The above table is a demonstration of the concentration of the 13 different water parameters for the 9 different sampling locations. The values listed are based on laboratory analysis of the water samples for the period from Apr’23- Jun’23.

Table 6.4: Concentrations of Water Quality Parameters of the sampling sites for Summer Season

Water quality parameters	***SITE 1***	***SITE 2***	***SITE 3***	***SITE 4***	***SITE 5***	***SITE 6***	***SITE 7***	***SITE 8***	***SITE 9***
Temperature (˚)	31	31.4	28	28.5	26	27.5	29	30	29.6
pH	7.12	7.27	7.4	6.95	7.06	8	7.89	7.10	7.3
Dissolved oxygen (mg/l)	5.45	5.23	3.5	4.12	5.3	3.1	2.93	4.67	3.85
Total Dissolved Solids (mg/l)	255	230	238	216	168.5	185	160	130	200
Salinity (ppt)	0.308	0.29	0.325	0.402	0.196	0.215	0.405	0.174	0.149
Electrical Conductivity (ms/cm)	0.4	0.283	0.350	0.218	0.199	0.292	0.368	0.200	0.260
Biological Oxygen Demand (mg/l)	1.15	1.67	4.40	2	1.2	4.89	6.7	1.8	4.2
Total Hardness (mg/l)	120	82	127	130	127	104	95	103	79
Chloride (mg/l)	98	100	84	75	48	65	90	82	70
Iron (mg/l)	0.90	0.54	0.55	1.10	1.26	1.55	1.65	0.96	0.85
Nitrate (mg/l)	0.0058	0.007	0.014	0.028	0.042	0.1	0.19	0.038	0.009

Lead (mg/l)	0.01	0.053	0.019	0.043	0.038	0.01	0.012	0.037	0.04
Turbidity(NTU)	9.5	11	10	7.5	10	15	14	6	5

The above table is a demonstration of the concentration of the 13 different water parameters for the 9 different sampling locations. The values listed are based on laboratory analysis of the water samples for the period from July'23- Sep'23.

6.1.2 Graphical output of results:

The results obtained from the testing of water quality parameters are presented in graphical form below:

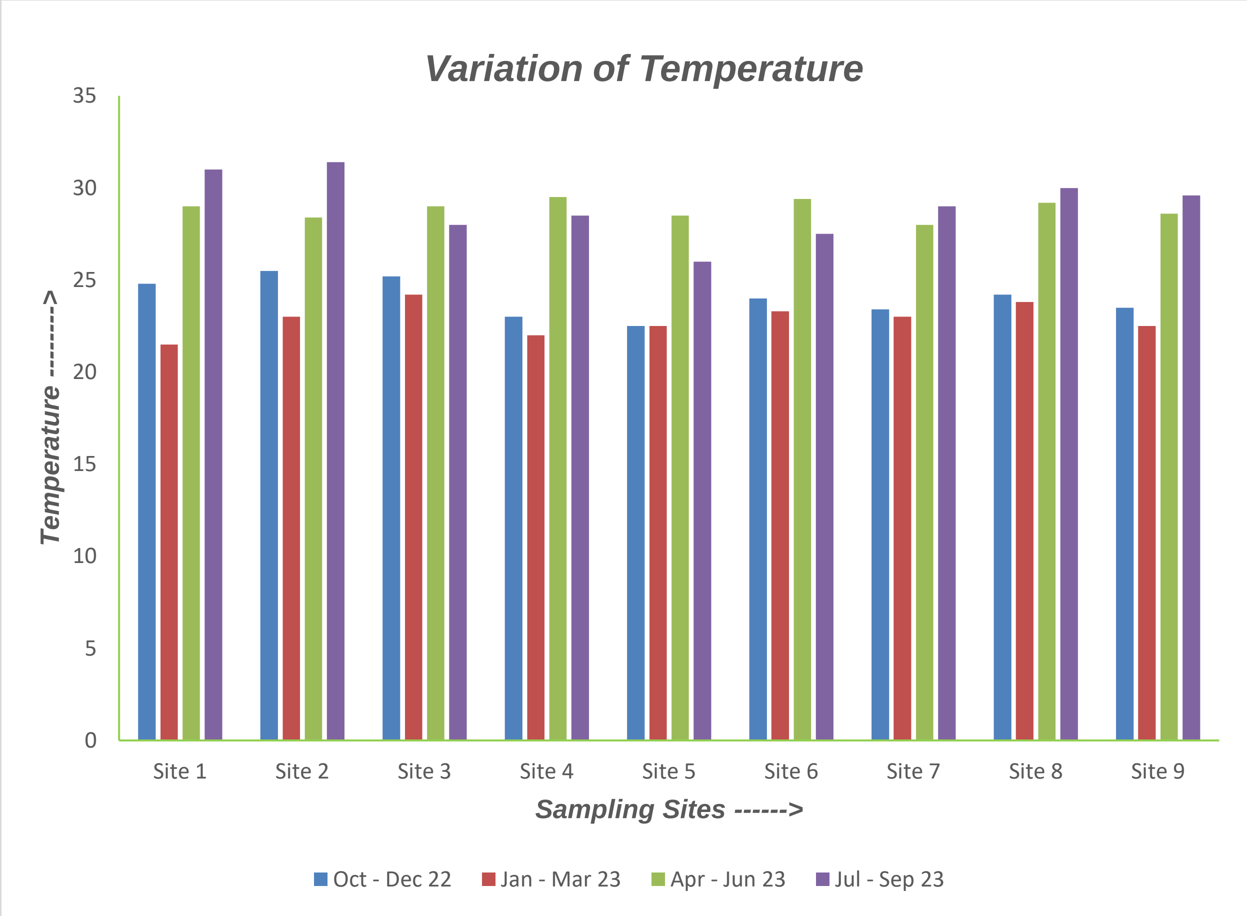


Fig 6.1: Frequency distribution of temperature variation in the sampling sites

The fig 6.1 shows the highest value of Temperature at site 2 in the period from Jul to Sep'23 and lowest value of Temperature at site 1 in the period from Jan to Mar'23.

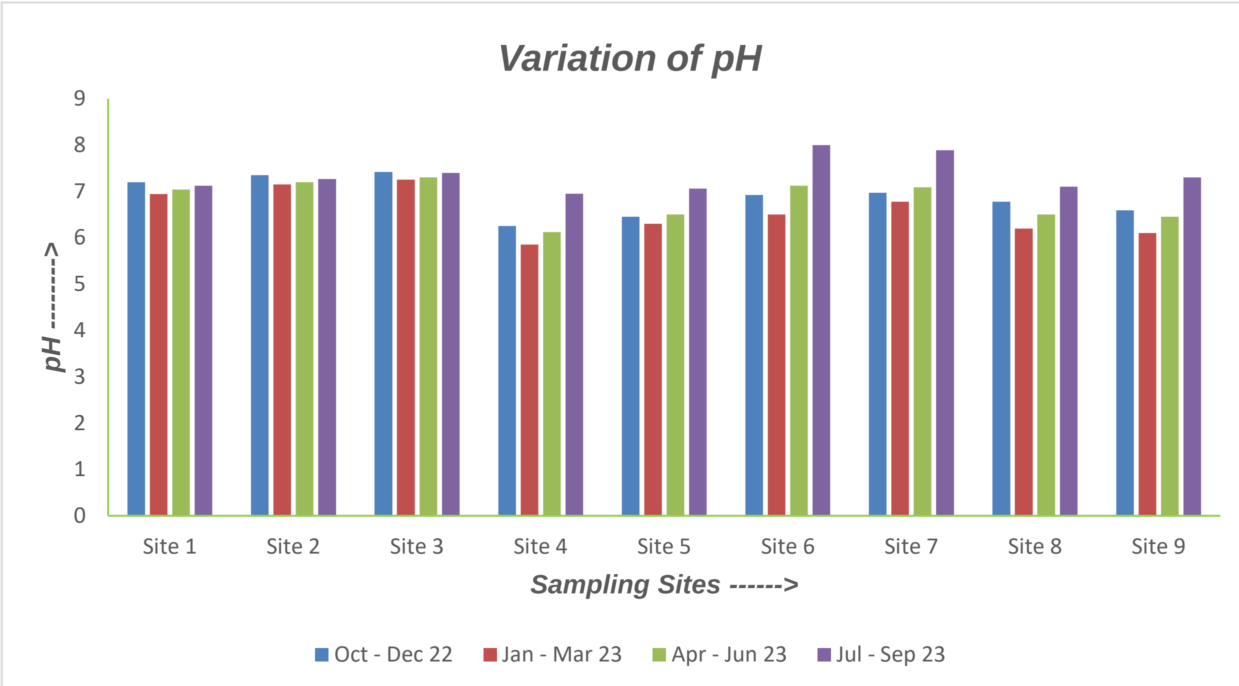


Fig 6.2: Frequency distribution of the pH concentration in the sampling sites

The fig. 6.2 shows the highest value of pH at site 6 in the period from Jul to Sep'23 and lowest value of pH at site 4 in the period from Jan to Mar'23.

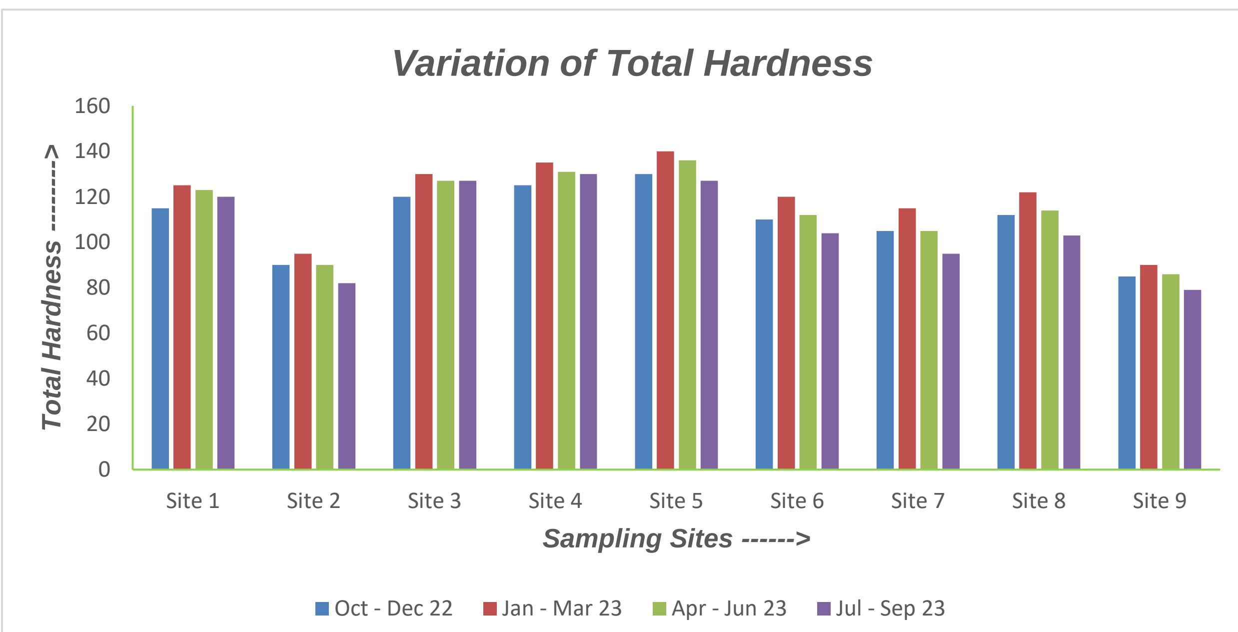


Fig 6.3: Frequency distribution of the Total Hardness concentration in the sampling sites

The fig. 6.3 shows the highest value of Total Hardness at site 5 in the period from Jan to Mar'23 and lowest value of Total Hardness at site 9 in the period from Jul to Sep'23.

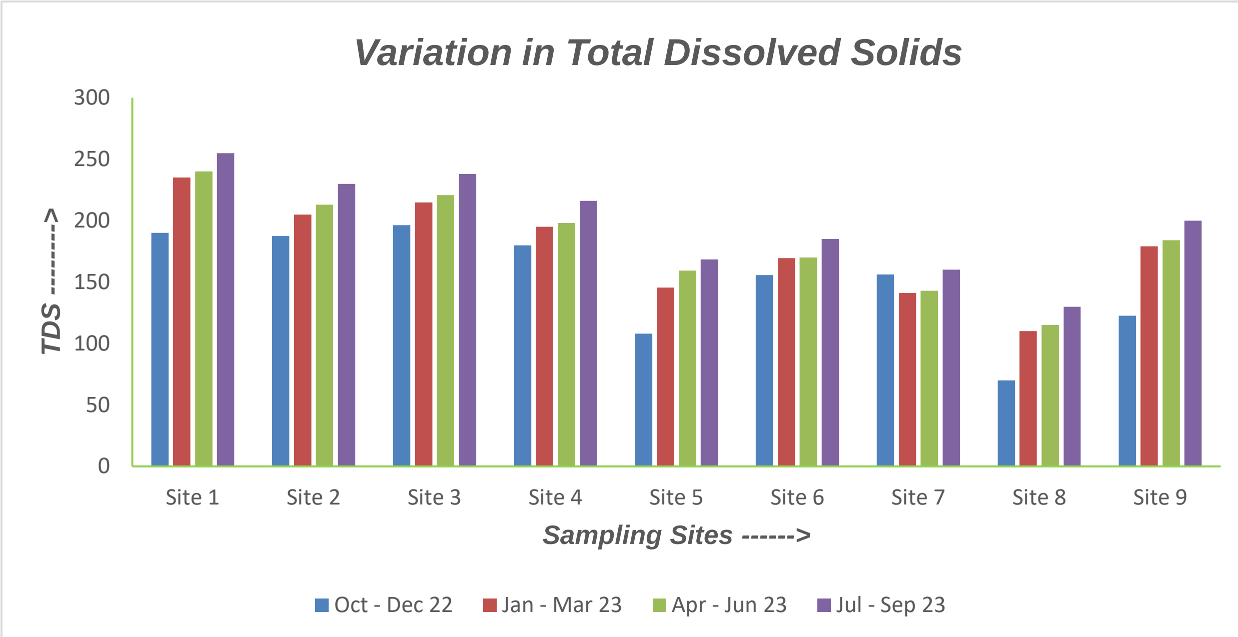


Fig 6.4: Frequency distribution of the Total dissolved solids concentration in the sampling sites

The fig. 6.4 shows the highest value of TDS at site 1 in the period from Jul to Sep'23 and lowest value of TDS at site 8 in the period from Oct to Dec'22.

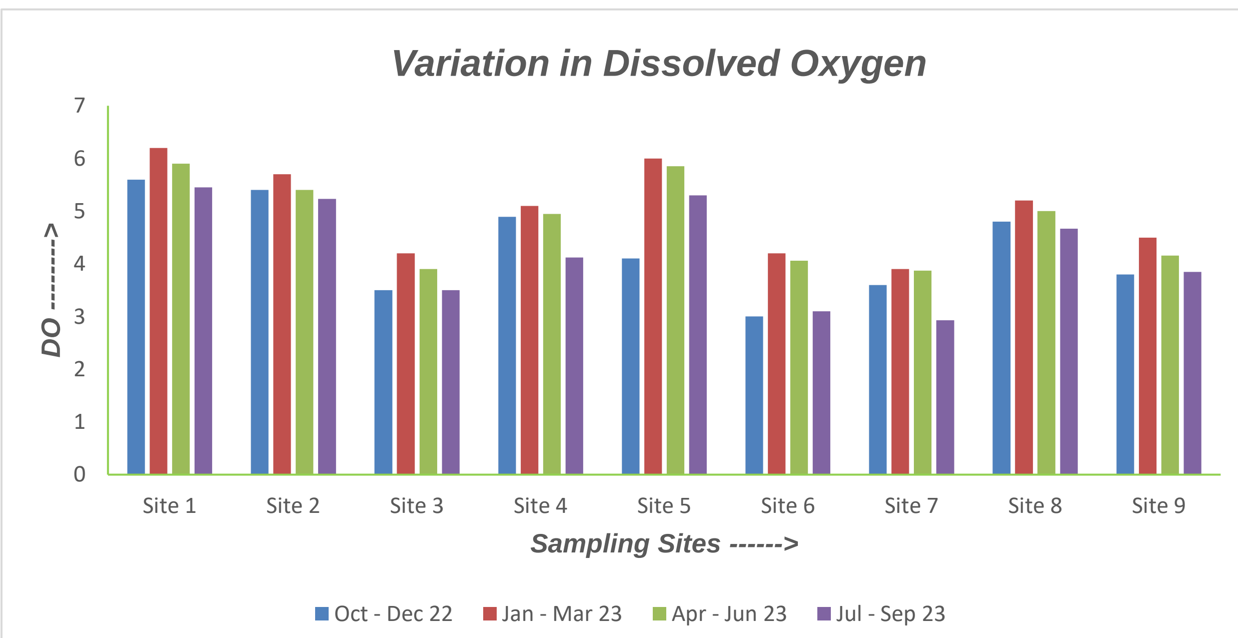


Fig 6.5: Frequency distribution of Dissolved Oxygen concentration in the sampling sites

The fig. 6.5 shows the highest value of DO at site 1 in the period from Jan to Mar'23 and lowest value of DO at site 7 in the period from Jul to Sep'23.

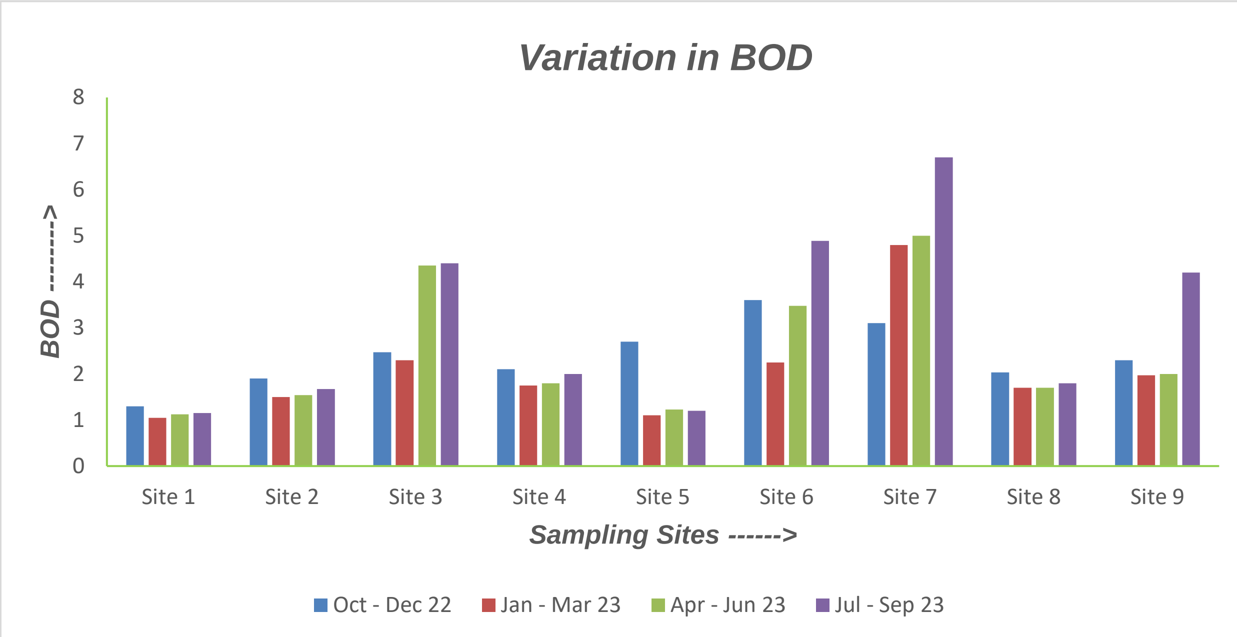


Fig 6.6: Frequency distribution of B.O.D concentration in the sampling sites

The fig. 6.6 shows the highest value of BOD at site 7 in the period from Jul to Sep’23 and lowest value of BOD at site 5 in the period from Jan to Mar’23.

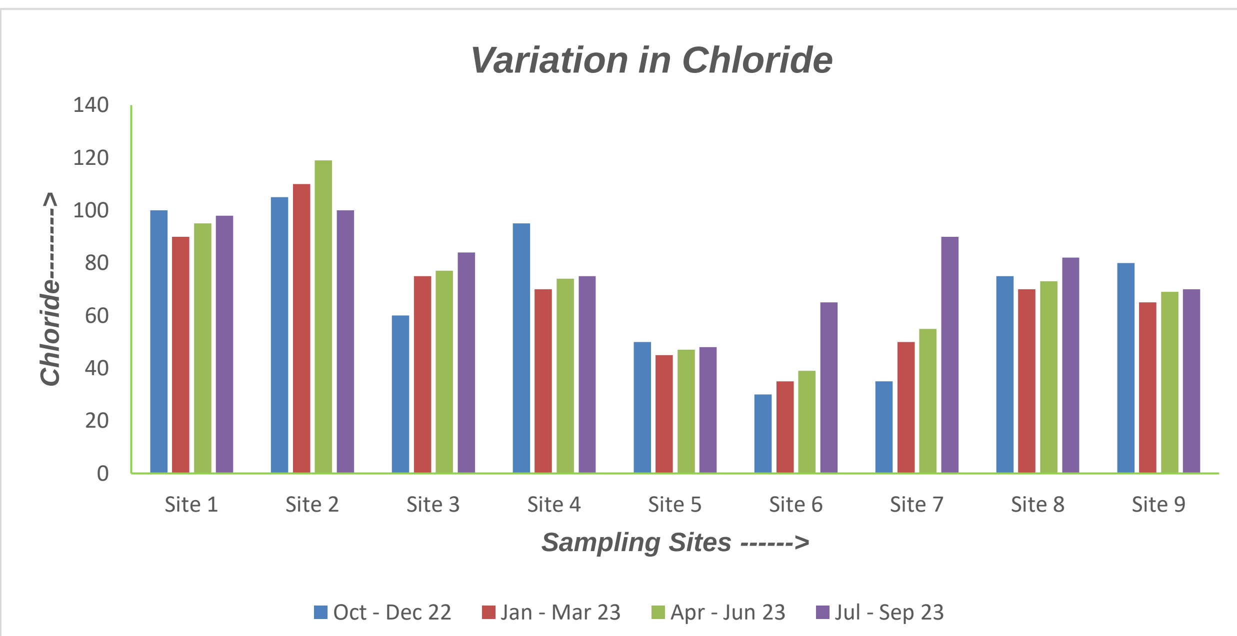


Fig 6.7: Frequency distribution of Chloride content in the sampling sites

The fig. 6.7 shows the highest value of chloride at site 2 in the period from Apr to Jun’23 and lowest value of chloride at site 6 in the period from Oct to Dec’23.

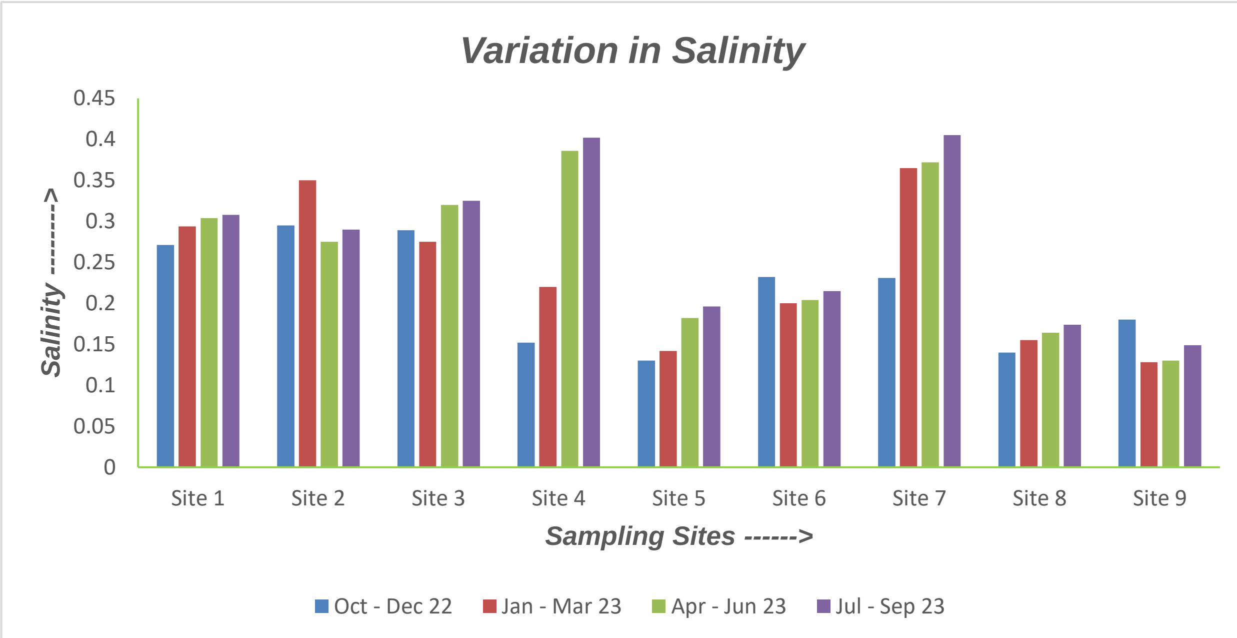


Fig 6.8: Frequency distribution of Salinity concentration in the sampling sites

The fig. 6.8 shows the highest value of salinity at site 7 in the period from Jul to Sep'23 and lowest value of salinity at site 9 in the period from Jan to Mar'23.

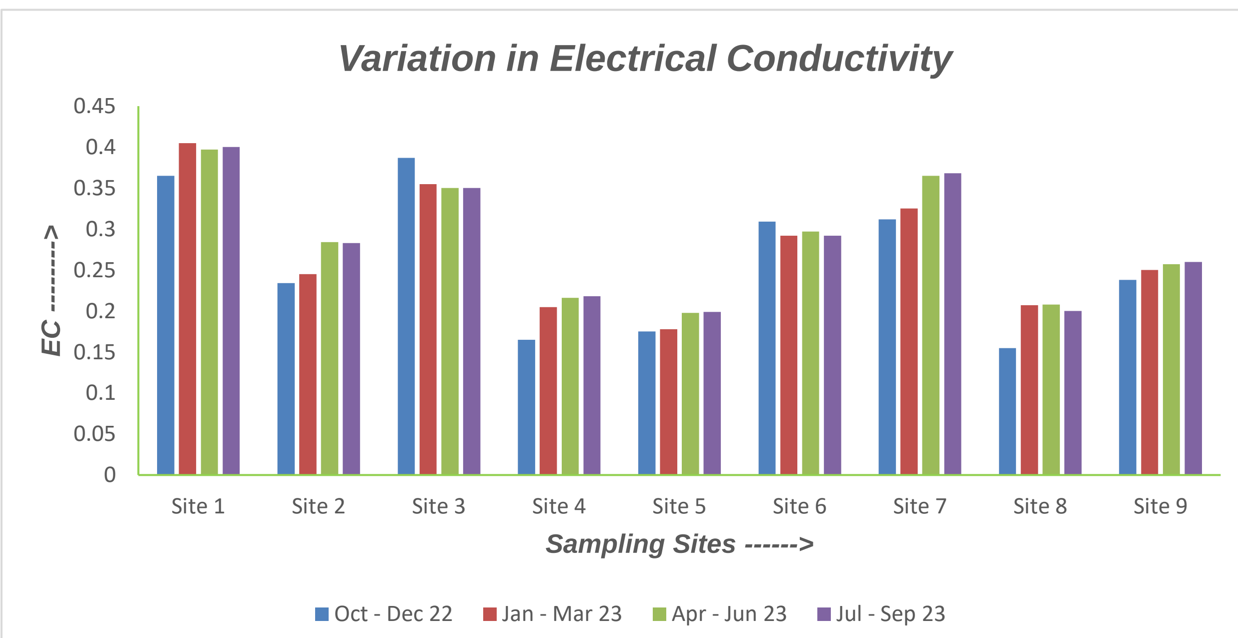


Fig 6.9: Frequency distribution of Electric Conductivity in the sampling sites

The fig. 6.9 shows the highest value of EC at site 1 in the period from Jan to Mar'23 and lowest value of EC at site 8 in the period from Oct to Dec'23.

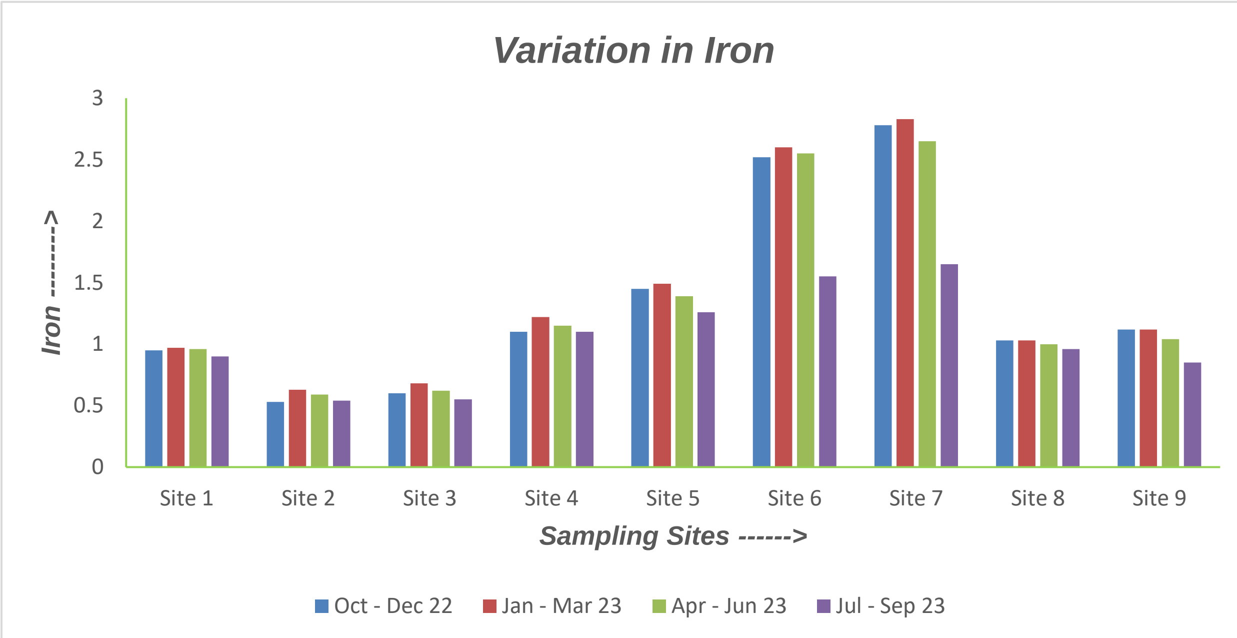


Fig 6.10: Frequency distribution of Iron concentration in the sampling sites

The fig. 6.10 shows the highest value of Iron at site 7 in the period from Jan to Mar'23 and lowest value of Iron at site 2 in the period from Oct to Dec'23.

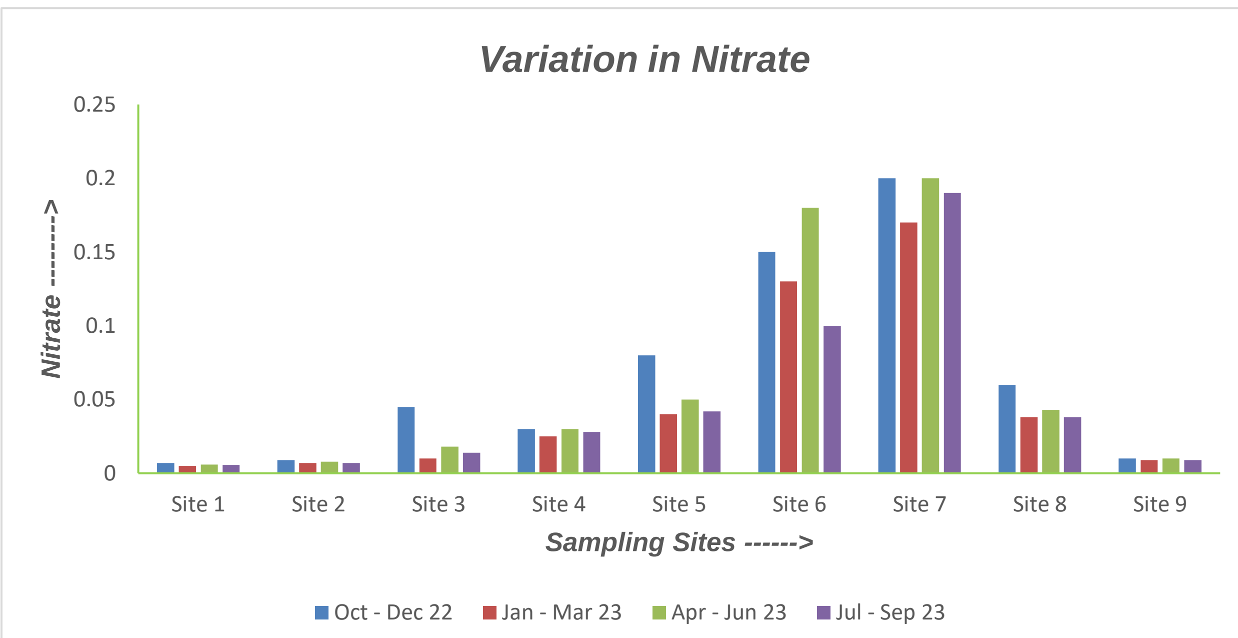


Fig 6.11: Frequency distribution of Nitrate concentration in the sampling sites

The fig. 6.11 shows the highest value of Nitrate at site 7 in the period from Oct to Dec'22 & Apr to Jun'23 and lowest value of Nitrate at site 1 in the period from Jan to Mar'23.

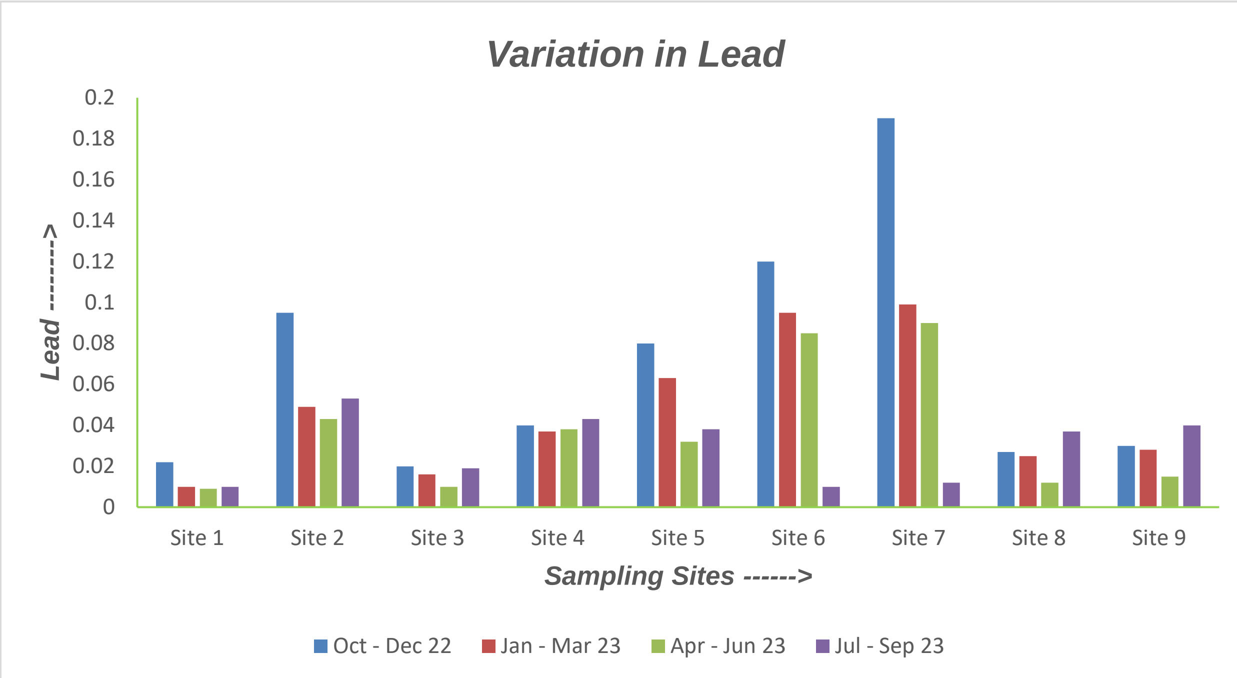


Fig 6.12: Frequency distribution of Lead concentration in the sampling sites

The fig. 6.12 shows the highest value of Lead at site 7 in the period from Oct to Dec'23 and lowest value of Lead at site 1 in the period from Apr to Jun'23.

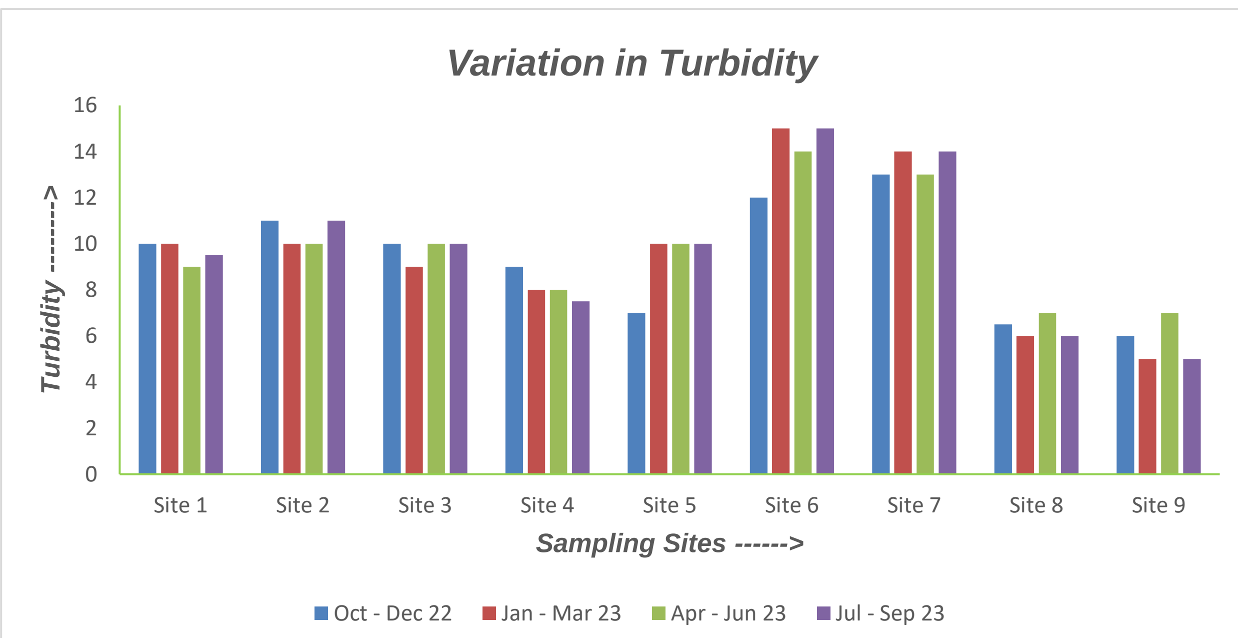


Fig 6.13: Frequency distribution of Turbidity concentration in the sampling sites

The fig. 6.13 shows the highest value of Turbidity at site 6 in the period from Jan to Mar'23 & Jul to Sep'23 and lowest value of Turbidity at site 9 in the period from Jan to Mar'23 & Jul to Sep'23.

CHAPTER 7

7.1 MULTIVARIATE DATA ANALYSIS

Multivariate data analysis (**MVDA**) is a Statistical procedure for analysis of data involving more than one type of measurement or observation. It may also mean solving problems where more than one dependent variable is analysed simultaneously with other variables.

7.1.1 Advantages and Disadvantages of Multivariate Data Analysis

Advantages

- The main advantage of multivariate analysis is that since it considers more than one factor of independent variables that influence the variability of dependent variables, the conclusion drawn is more accurate.
- The conclusions are more realistic and nearer to the real-life situation.

Disadvantages

- The main disadvantage of MVDA includes that it requires rather complex computations to arrive at a satisfactory conclusion.
- Many observations for a large number of variables need to be collected and tabulated; it is a rather time-consuming process.

There are many different techniques for multivariate analysis, and they can be divided into two categories:

- Dependence techniques
- Interdependence techniques

7.1.2 Dependence methods

Dependence methods are used when one or some of the variables are dependent on others. Dependence looks at cause and effect; in other words, can the values of two or more independent variables be used to explain, describe, or predict the value of another, dependent variable. To give a simple example, the dependent variable of "weight" might be predicted by independent variables such as "height" and "age."

In machine learning, dependence techniques are used to build predictive models. The analyst enters input data into the model, specifying which variables are independent and which ones are dependent—in other words, which variables they want the model to predict, and which variables they want the model to use to make those predictions.

7.1.3 Interdependence methods

Interdependence methods are used to understand the structural makeup and underlying patterns within a dataset. In this case, no variables are dependent on others, so you're not looking for causal relationships. Rather, interdependence methods seek to give meaning to a set of variables or to group them together in meaningful ways.

So, one is about the effect of certain variables on others, while the other is all about the structure of the dataset.

The various type of multivariate analysis techniques available in IBM SPSS Statistics are:

- Multiple linear regression
- Multiple logistic regression
- Multivariate analysis of variance (MANOVA)
- Factor analysis
- Cluster analysis
- K-Means cluster analysis
- Cluster silhouettes
- Discriminant analysis
- Optimal scaling

In this report, the multivariate data analysis techniques used are One-Way Analysis of Variance (ANOVA), Hierarchical Cluster Analysis (HCA), Factor analysis or Principal Component Analysis (PCA) and K-Means Cluster Analysis and Pearson Correlation Matrix Method.

It was carried out in the free trial version of IBM SPSS Statistics downloaded from its official website.

7.2 ANALYSIS OF VARIANCE AND CORRELATION

7.2.1 One-way ANOVA

The results obtained from the test were subjected to basic statistical analysis (One-Way ANOVA) and formation a correlation matrix. One-Way Analysis of Variance (ANOVA) of the various the parameters of the water samples are done to determine whether there are any statistically significant differences between the means of two or more independent groups. In this case study, the One-Way ANOVA is performed using SPSS Software. This analysis tests the validity of null hypothesis that the water variables concentrations did not differ with different seasons. This significance is based on the F-ratio and the p-value, obtained from the software during the analysis. If the F-ratio > 1.0 and the p-value<.05, then it is concluded as the null hypothesis is not valid and the differences are quite significant. For the null hypothesis to hold good, the value of F must be approximately equal to 1. The One-Way ANOVA values obtained for each parameter against each season are shown below.

Table 7.1: One-Way ANOVA datasheet for Autumn Season using SPSS

ANOVA						
	Cluster		**Error**		**F**	**P**
	Mean Square	**df**	**Mean Square**	**df**		
pH	5.224	1	0.397	7	13.176	0.008
DO	1.556	1	0.921	7	1.690	0.235
TEMPERATURE(0C)	5.842	1	0.308	7	18.947	0.003
TDS	3.734	1	0.609	7	6.128	0.042
SALINTY	5.507	1	0.356	7	15.462	0.006
EC	2.797	1	0.743	7	3.762	0.034
BOD	2.408	1	0.799	7	3.015	0.026
TOTAL HARDNESS	0.071	1	1.133	7	0.063	0.809
CHLORIDE	1.952	1	0.864	7	2.259	0.177
IRON	3.010	1	0.713	7	4.222	0.079

NITRATE	2.008	1	0.856	7	2.346	0.169
LEAD	0.753	1	1.035	7	0.727	0.422
TURBIDITY	0.657	1	1.049	7	0.626	0.455

From the Table 7.1 we can conclude that the parameters like pH, TEMPERATURE, TDS, SALINTY, EC, IRON and BOD have significant variation among all the nine sites in autumn season.

Table 7.2: One-Way ANOVA datasheet for Winter Season using SPSS

ANOVA						
	Cluster		**Error**		**F**	**P**
	Mean Square	**df**	**Mean Square**	**df**		
TEMPERATURE(0C)	0.287	1	1.102	7	0.260	0.626
pH	0.063	1	1.134	7	0.055	0.821
DO	3.246	1	0.679	7	4.779	0.065
TDS	0.781	1	1.031	7	0.757	0.413
SALINTY	0.688	1	1.045	7	0.658	0.444
EC	0.540	1	1.066	7	0.507	0.500
BOD	4.456	1	0.506	7	8.802	0.021
TOTAL HARDNESS	0.023	1	1.140	7	0.020	0.890
CHLORIDE	3.089	1	0.702	7	4.402	0.074
IRON	7.098	1	0.129	7	55.112	0.000
NITRATE	7.400	1	0.086	7	86.368	0.000
LEAD	6.047	1	0.279	7	21.671	0.002
TURBIDITY	5.588	1	0.345	7	16.218	0.005

From the Table 7.2 we can conclude that the parameters like DO, NITRATE, IRON, LEAD, TURBIDITY and BOD have significant variation among all the nine sites in winter season.

Table 7.3: One-Way ANOVA datasheet for Spring Season using SPSS

ANOVA						
	Cluster		**Error**		**F**	**p.**
	Mean Square	**df**	**Mean Square**	**df**		
TEMPERATURE(0C)	0.214	1	1.112	7	0.193	0.674
pH	1.235	1	0.966	7	1.277	0.296

DO	2.595	1	0.772	7	3.361	0.009
TDS	1.096	1	0.986	7	1.111	0.327
SALINTY	0.238	1	1.109	7	0.214	0.658
EC	0.993	1	1.001	7	0.992	0.352
BOD	3.924	1	0.582	7	6.740	0.036
TOTAL HARDNESS	0.232	1	1.110	7	0.209	0.661
CHLORIDE	2.690	1	0.759	7	3.546	0.102
IRON	7.163	1	0.120	7	59.900	0.000
NITRATE	7.631	1	0.053	7	144.744	0.000
LEAD	6.699	1	0.186	7	36.045	0.001
TURBIDITY	5.993	1	0.287	7	20.907	0.003

From the Table 7.3 we can conclude that the parameters like DO, NITRATE, IRON, LEAD, TURBIDITY and BOD are showing significant variation among all the nine sites in spring season.

Table 7.4: One-Way ANOVA datasheet for Summer Season using SPSS

ANOVA						
	Cluster		**Error**		**F**	**p**
	Mean Square	**df**	**Mean Square**	**df**		
TEMPERATURE(0C)	0.490	1	1.073	7	0.457	0.521
pH	6.866	1	0.162	7	42.385	0.000
DO	4.118	1	0.555	7	7.426	0.030
TDS	1.010	1	0.999	7	1.011	0.348
SALINTY	0.371	1	1.090	7	0.340	0.578
EC	0.921	1	1.011	7	0.911	0.372
BOD	4.712	1	0.470	7	10.033	0.016
TOTAL HARDNESS	0.422	1	1.083	7	0.390	0.552
CHLORIDE	0.024	1	1.139	7	0.021	0.888
IRON	5.218	1	0.397	7	13.127	0.008
NITRATE	6.525	1	0.211	7	30.961	0.001
LEAD	3.140	1	0.694	7	4.523	0.041
TURBIDITY	5.151	1	0.407	7	12.656	0.009

From the Table 7.4 we can conclude that the parameters like DO, pH, NITRATE, IRON, LEAD, TURBIDITY and BOD are showing significant variation among all the nine sites in summer season.

7.3 Hierarchical Cluster Analysis (HCA)

HEIRARCHICAL CLUSTER ANALYSIS is a procedure that attempts to identify relatively homogeneous groups of cases (or variables) based on selected characteristics, using an algorithm that starts with each case (or variable) in a separate cluster and combines clusters until only one is left. It is an unsupervised pattern recognition technique, and its algorithms produce a sequence of nested partitions including similar groups. Clusters in HCA are formed sequentially, starting with the most similar pair of variables and forming higher clusters step by step. Cluster process formation is repeated until a single cluster containing all the variables are obtained. The result of the clustering can be displayed in a **tree like structure, called a dendrogram**. The dendrogram can be broken at different levels to yield different clusters of the data set. However, it should be noted that the decision of the final cluster is rather arbitrary.

The hierarchical agglomerative clustering methods differ in the way they calculate the similarity between two clusters i.e., single link, complete link, group average and WARDS METHOD. The former methods depend on calculating the similarity between two patterns using a distance measure. The most popular distance method is the EUCLIDEAN distance (Abu-Khalaf et.al, 2013).

7.3.1 Considerations for HCA

Statistics. Agglomeration schedule, distance (or similarity) matrix, and cluster membership for a single solution or a range of solutions.

Plots: Dendrograms and icicle plots.

Data: The variables can be quantitative, binary, or count data. Scaling of variables is an important issue--differences in scaling may affect our cluster solution(s). If the variables have large differences in scaling (for example, one variable is measured in dollars and the other is measured in years), one should consider standardizing them (this can be done automatically by the Hierarchical Cluster Analysis procedure).

Case order: If tied distances or similarities exist in the input data or occur among updated clusters during joining, the resulting cluster solution may depend on the order of cases in the file. You may want to obtain several different solutions with cases sorted in different random orders to verify the stability of a given solution.

Assumptions: The distance or similarity measures used should be appropriate for the data analysed (see the Proximities procedure for more information on choices of distance and similarity measures). Also, you should include all relevant variables in your analysis. Omission of influential variables can result in a misleading solution. Because hierarchical cluster analysis is an exploratory method, results should be treated as tentative until they are confirmed with an independent sample.

Hierarchical Cluster Analysis was used to identify groups of similarity between sampling areas, which was applied to the data obtained from the four seasons from nine sites. The best results for the variable relationship between the methods used for correlation and distance combination were obtained using Ward's method and Euclidean distance. In this study a Q mode hierarchical cluster analysis (HCA) was applied to classify samples according to their parameters. The similarity measurements together with Ward's method for linkage produces the most distinctive groups where each member within the group is more similar to its fellow members than to any member outside the group. The Ward's method is distinct from the other methods, because it uses an analysis of variance approach to evaluate the distance between the clusters. Cluster membership in this method is assessed by calculating the total sum of squared deviations from the mean of a cluster. The criteria for fusion are that it should produce the smallest possible increase in the error sum of squares. The Wards method with squared Euclidean distance is used as a dissimilarity measure has been found to provide meaningful dendrogram of clusters with the proximity or similarity of clusters measured with a rescaled distance.

7.3.2 Output for Hierarchical Cluster Analysis for Autumn Season

Case Processing Summary[a,b]

Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
9	100.0	0	.0	9	100.0

a. Squared Euclidean Distance used

b. Ward Linkage

Proximity Matrix									
	Squared Euclidean Distance								
Case	site 1: near ASTU	site 2: Near MCA building	site 3: Tetelia	site 4: Chakardeo village	Site 5: Mikirpara Rani	site 6: Boragaon dumping site A	site 7: Boragaon dumping site B	site 8: GIMT Azara	site 9: Dharapur
site 1	.000	658.362	1669.552	231.364	9468.377	6127.628	5489.112	15048.030	5851.450
site 2	658.362	.000	3005.755	1393.36	10974.272	7058.685	6123.043	15213.179	4881.450
site 3	1669.552	3005.755	.000	1522.04	7997.681	2659.294	2467.664	16231.244	7046.974
site 4	231.364	1393.364	1522.037	.000	7239.401	5063.695	4588.684	12676.990	5119.056
site 5	9468.377	10974.272	7997.681	7239.40	.000	3096.431	3212.555	2397.368	3143.572
site 6	6127.628	7058.685	2659.294	5063.70	3096.431	.000	52.408	9394.644	4248.005
site 7	5489.112	6123.043	2467.664	4588.68	3212.555	52.408	.000	9128.093	3599.801
site 8	15048.030	15213.179	16231.244	12677.0	2397.368	9394.644	9128.093	.000	3533.122
site 9	5851.450	4881.450	7046.974	5119.06	3143.572	4248.005	3599.801	3533.122	.000

Fig 7.1: The figure shows the similarity between the cases (here sites are taken as case number) based on the squared Euclidean Distance. It was done with the WARD'S method.

7.3.3 Ward Linkage for Autumn Season

Agglomeration Schedule						
	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
1	6	7	26.204	0	0	7
2	1	4	141.886	0	0	3
3	1	2	787.234	2	0	5
4	5	8	1985.918	0	0	6
5	1	3	3344.996	3	0	8
6	5	9	5170.999	4	0	7
7	5	6	10481.308	6	1	8
8	1	5	23068.038	5	7	0

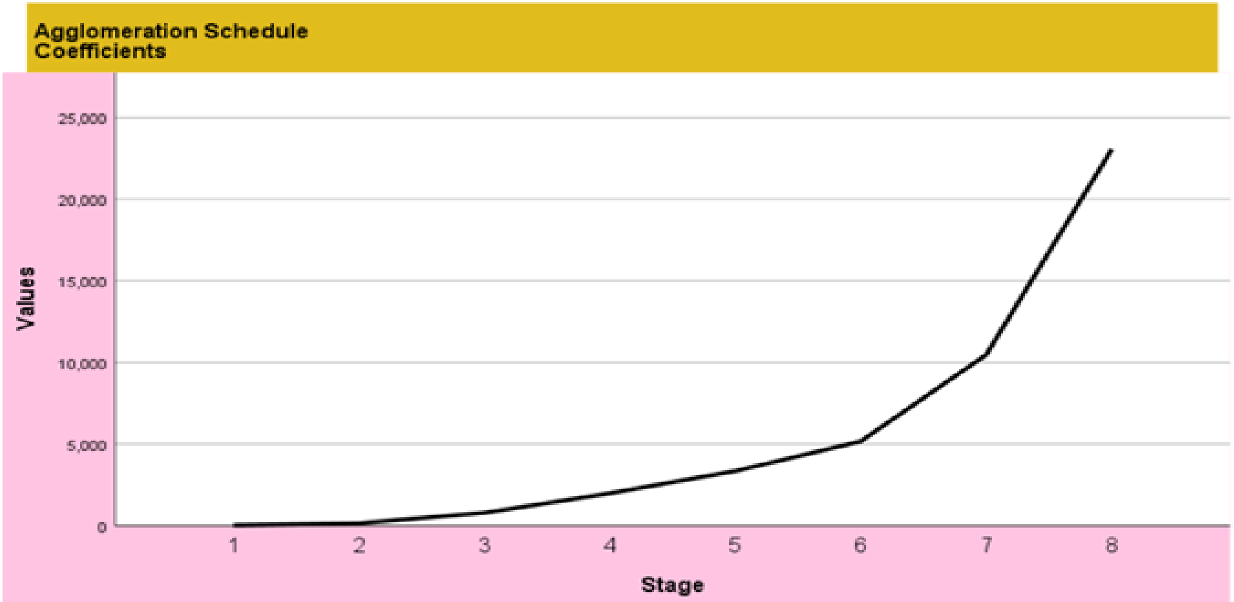


Fig 7.2: The figure shows the plot between the agglomeration coefficients and the different stages.

7.3.4 Dendrogram Plot for Autumn Season for Nine Sites

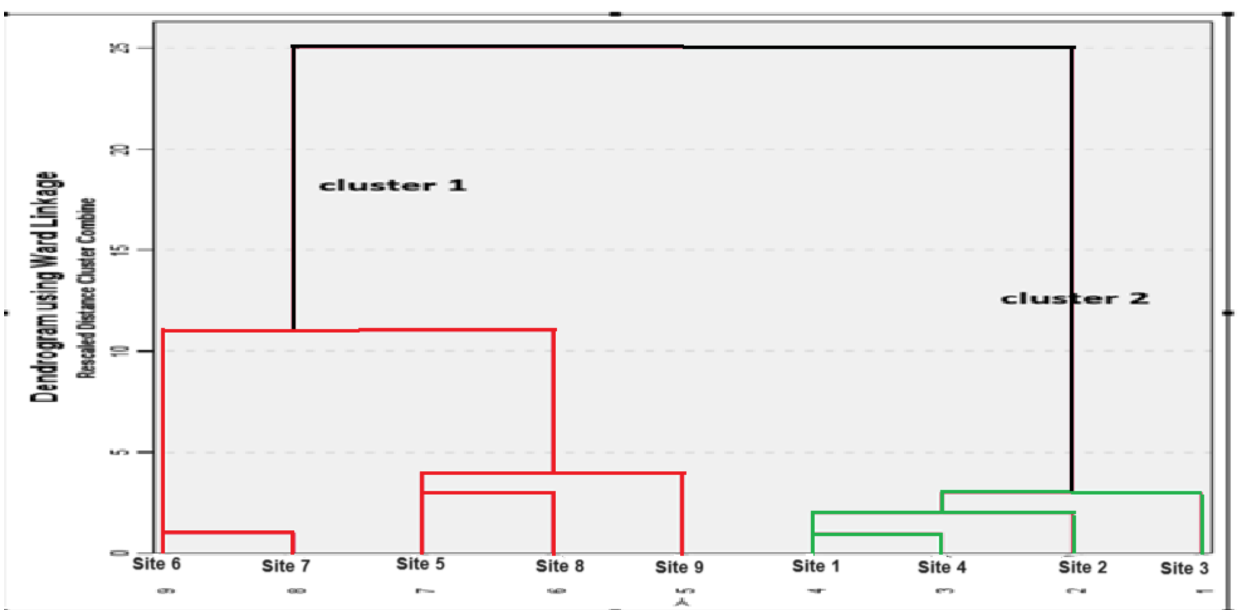


Fig 7.3: The figure shows a tree-like structure which depicts clusters between the sites and the rescaled Euclidean distance. It demonstrates two clusters. The first cluster consists of site 6, site 7, site 5, site 8 and site 9 and site 1, site 4, site 2 and site 3 belong to cluster 2.

7.3.5 Output for Hierarchical Cluster Analysis for Winter Season

Case Processing Summary[a,b]

Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
9	100.0	0	.0	9	100.0

a. Squared Euclidean Distance used

b. Ward Linkage

Proximity Matrix

	Squared Euclidean Distance								
Case	site 1:Near ASTU	site 2:Near MCA building	site 3 : Tetelia	site 4 : Chakardeo village	site 5 : mikirpara rani	site 6 : Boragaon dumping site A	Site 7 : Boragaon damping site B	Site 8: GIMT Azara	site 9: Dharapur
site 1	.000	658.362	1669.552	231.364	9468.377	6127.628	5489.112	15048.030	5851.450
site 2	658.362	.000	3005.755	1393.364	10974.272	7058.685	6123.043	15213.179	4881.450
site 3	1669.552	3005.755	.000	1522.037	7997.681	2659.294	2467.664	16231.244	7046.974
site 4	231.364	1393.364	1522.037	.000	7239.401	5063.695	4588.684	12676.990	5119.056
site 5	9468.377	10974.272	7997.681	7239.401	.000	3096.431	3212.555	2397.368	3143.572
site 6	6127.628	7058.685	2659.294	5063.695	3096.431	.000	52.408	9394.644	4248.005
site 7	5489.112	6123.043	2467.664	4588.684	3212.555	52.408	.000	9128.093	3599.801
site 8	15048.030	15213.179	16231.244	12676.990	2397.368	9394.644	9128.093	.000	3533.122
site 9	5851.450	4881.450	7046.974	5119.056	3143.572	4248.005	3599.801	3533.122	.000

This is a dissimilarity matrix

Fig 7.4: The figure shows the similarity between the cases (here sites are taken as case number) based on the squared Euclidean Distance. It was done with the WARD'S method.

7.3.6 Ward Linkage for Winter Season

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	4	29.626	0	0	2
2	3	8	112.158	1	0	4
3	6	7	241.080	0	0	5
4	1	3	513.963	0	2	7
5	5	6	876.988	0	3	8
6	2	9	1916.140	0	0	7
7	1	2	3766.028	4	6	8
8	1	5	6667.177	7	5	0

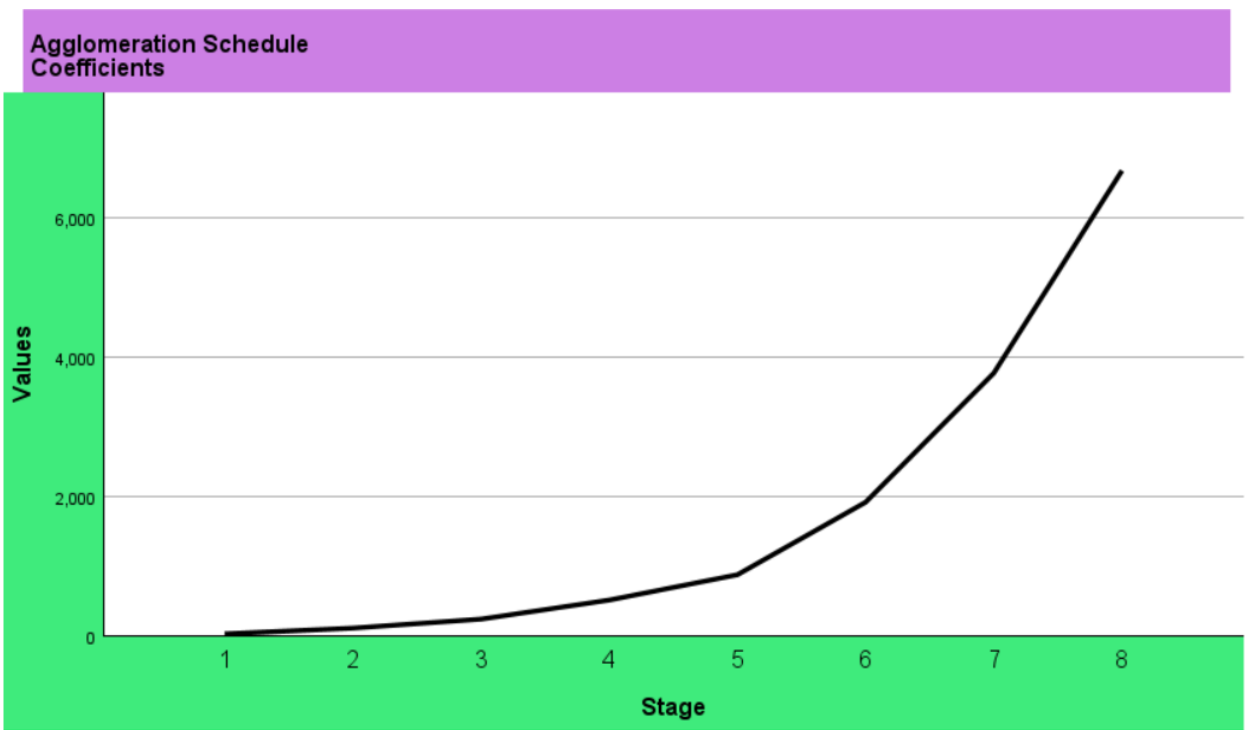


Fig 7.5: The figure shows the plot between the agglomeration coefficients and the different stages.

7.3.7 Dendrogram Plot for Winter Season for Nine Sites

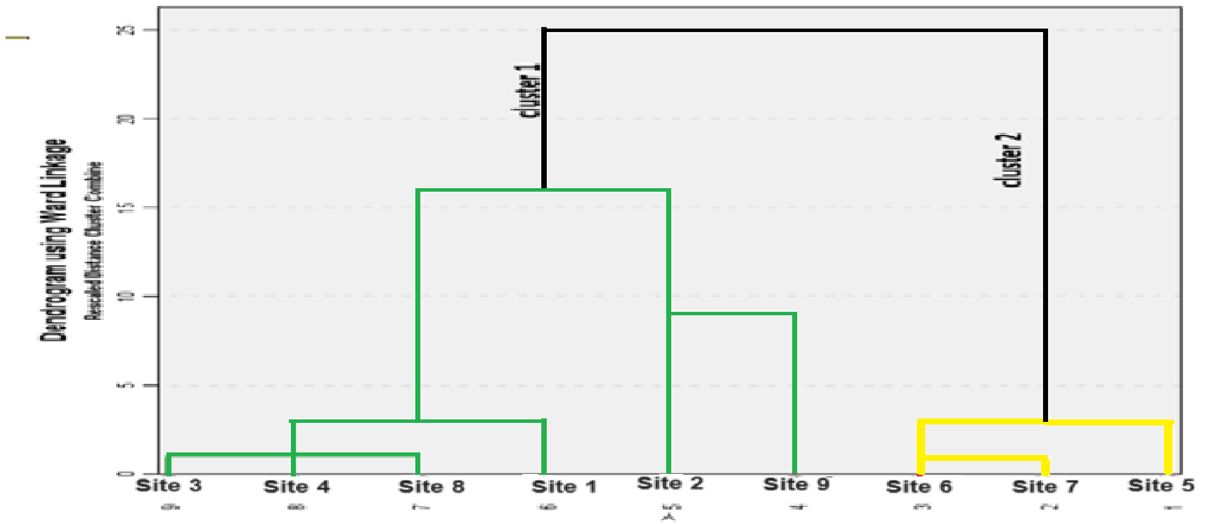


Fig 7.6: The figure shows a tree-like structure which depicts clusters between the sites and the rescaled Euclidean distance. It demonstrates two clusters. The first cluster consists of site 3, site 4, site 1, site 8 and site 9 and site 2 and site 5, site 6 and site 7 belong to cluster 2.

7.3.8 Output for Hierarchical Cluster Analysis for Spring Season

Case Processing Summary[a]

Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
9	100.0	0	.0	9	100.0

a. Ward Linkage

Proximity Matrix

Case	Squared Euclidean Distance								
	site 1 :Near ASTU	site 2: Near MCA Building	site 3: Tetelia	site 4: Chakardeo village	site 5: Mikirpara Rani	site 6 : Boragaon dumping site A	site 7 : Boragaon dumping site B	site 8 : GIMT Azara	site 9 : Dharapur
site 1	.000	2395.964	724.259	2272.538	8987.288	8193.706	11372.084	16195.536	5189.366
site 2	2395.964	.000	3204.365	3937.978	10185.147	8759.444	9248.783	12306.504	3368.577
site 3	724.259	3204.365	.000	558.392	4778.313	4270.388	7035.471	11396.742	3114.934
site 4	2272.538	3937.978	558.392	.000	2258.069	2412.649	4103.901	7180.320	2248.663
site 5	8987.288	10185.147	4778.313	2258.069	.000	781.327	1320.099	3133.076	3606.682
site 6	8193.706	8759.444	4270.388	2412.649	781.327	.000	1039.351	4240.913	1826.610
site 7	11372.084	9248.783	7035.471	4103.901	1320.099	1039.351	.000	1241.776	2286.558
site 8	16195.536	12306.504	11396.742	7180.320	3133.076	4240.913	1241.776	.000	5562.164
site 9	5189.366	3368.577	3114.934	2248.663	3606.682	1826.610	2286.558	5562.164	.000

This is a dissimilarity matrix

Fig 7.7: The figure shows the similarity between the cases (here sites are taken as case number) based on the squared Euclidean Distance. It was done with the WARD'S method.

7.3.9 Ward Linkage for Spring Season

Agglomeration Schedule					
	Cluster Combined			Stage Cluster First Appear	
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2
1	3	4	279.196	0	
2	5	6	669.859	0	
3	7	8	1290.747	0	
4	1	3	2196.614	0	
5	5	9	3877.491	2	
6	1	2	5965.801	4	
7	5	7	8281.085	5	
8	1	5	20081.993	6	

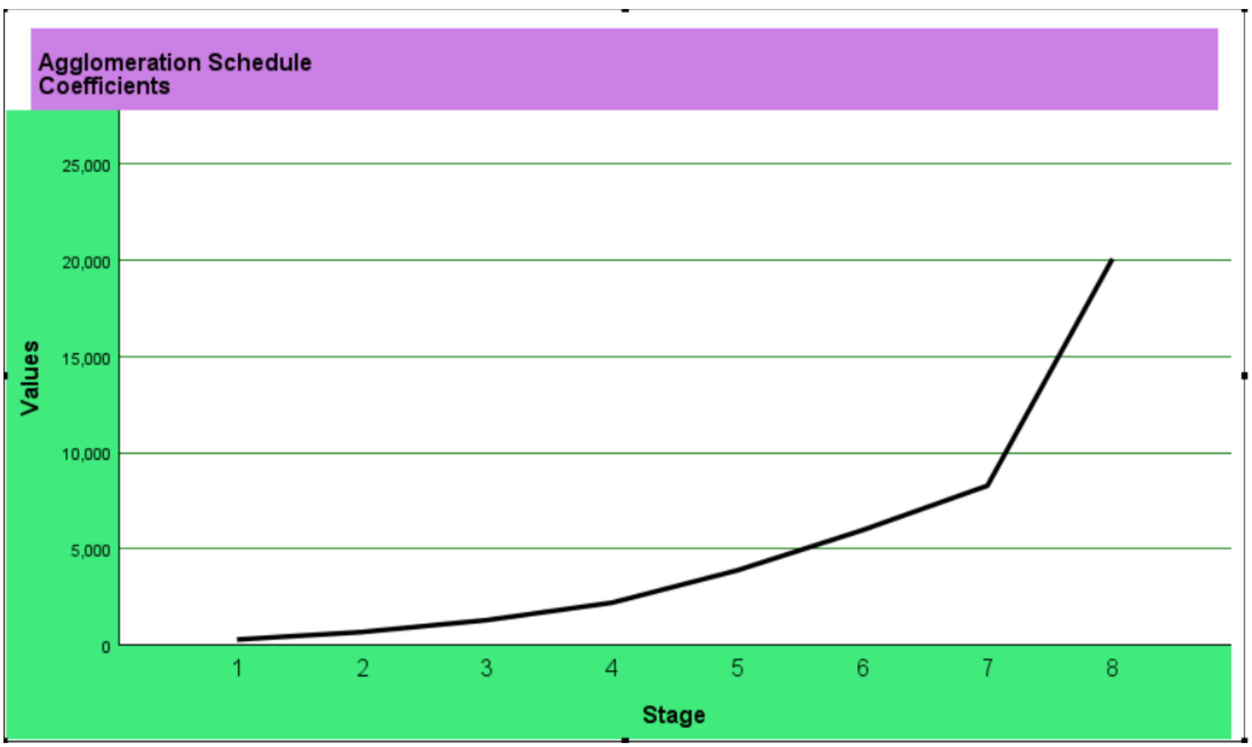


Fig 7.8: The figure shows the plot between the agglomeration coefficients and the different stages.

7.3.10 Dendrogram Plot for Spring Season for Nine Sites

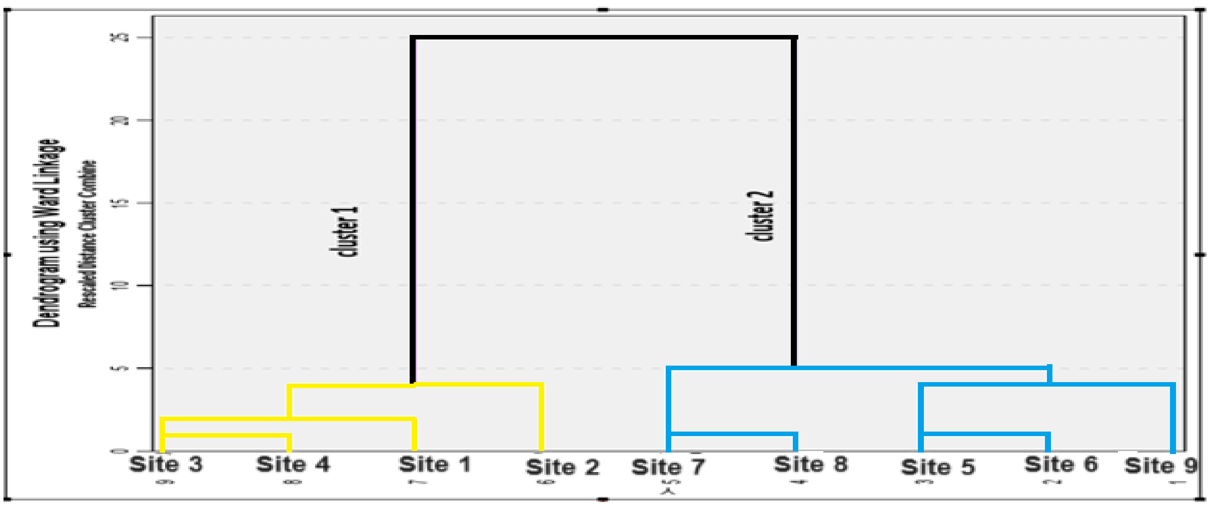


Fig 7.9: The figure shows a tree-like structure which depicts clusters between the sites and the rescaled Euclidean distance. It demonstrates two clusters. The first cluster consists of site 3, site 4, site 1 & site 2 and site 9, site 8, site 5, site 6 & site 7 belong to cluster 2.

7.3.11 Output for Hierarchical Cluster Analysis for Summer Season

Case Processing Summary[a]

Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
9	100.0	0	.0	9	100.0

a. Ward Linkage

Proximity Matrix

Case	Squared Euclidean Distance: site 1: Near ASTU	site 2: Near MCA building	site 3 : Tetelia	site 4: Chakardeo village	site 5: Mikirpara Rani	site 6: Boragaon dumping site A	Site 7: Boragaon dumping site B	Site 8: GIMT Azara	site 9: Dharapur
site 1	.000	2075.897	557.819	2162.854	10056.713	6308.236	9776.603	16184.345	5524.153
site 2	2075.897	.000	2368.030	3147.434	8542.216	3781.685	5216.023	10792.517	1856.663
site 3	557.819	2368.030	.000	587.174	6144.390	3726.033	7168.103	12272.433	3971.862
site 4	2162.854	3147.434	587.174	.000	3008.863	1804.994	4653.242	8178.937	2894.624
site 5	10056.713	8542.216	6144.390	3008.863	.000	1136.937	2922.053	3247.099	3829.545
site 6	6308.236	3781.685	3726.033	1804.994	1136.937	.000	1337.627	3415.436	981.443
site 7	9776.603	5216.023	7168.103	4653.242	2922.053	1337.627	.000	1121.243	2345.555
site 8	16184.345	10792.517	12272.433	8178.937	3247.099	3415.436	1121.243	.000	5627.650
site 9	5524.153	1856.663	3971.862	2894.624	3829.545	981.443	2345.555	5627.650	.000

This is a dissimilarity matrix

Fig 7.10: The figure shows the similarity between the cases (here sites are taken as case number) based on the squared Euclidean Distance. It was done with the WARD'S method.

7.3.12 Ward Linkage for Summer Season

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	1	3	278.909	0	0	4
2	6	9	769.631	0	0	5
3	7	8	1330.253	0	0	7
4	1	4	2153.959	1	0	6
5	5	6	3645.879	0	2	7
6	1	2	5268.065	4	0	8
7	5	7	7917.719	5	3	8
8	1	5	18747.381	6	7	0

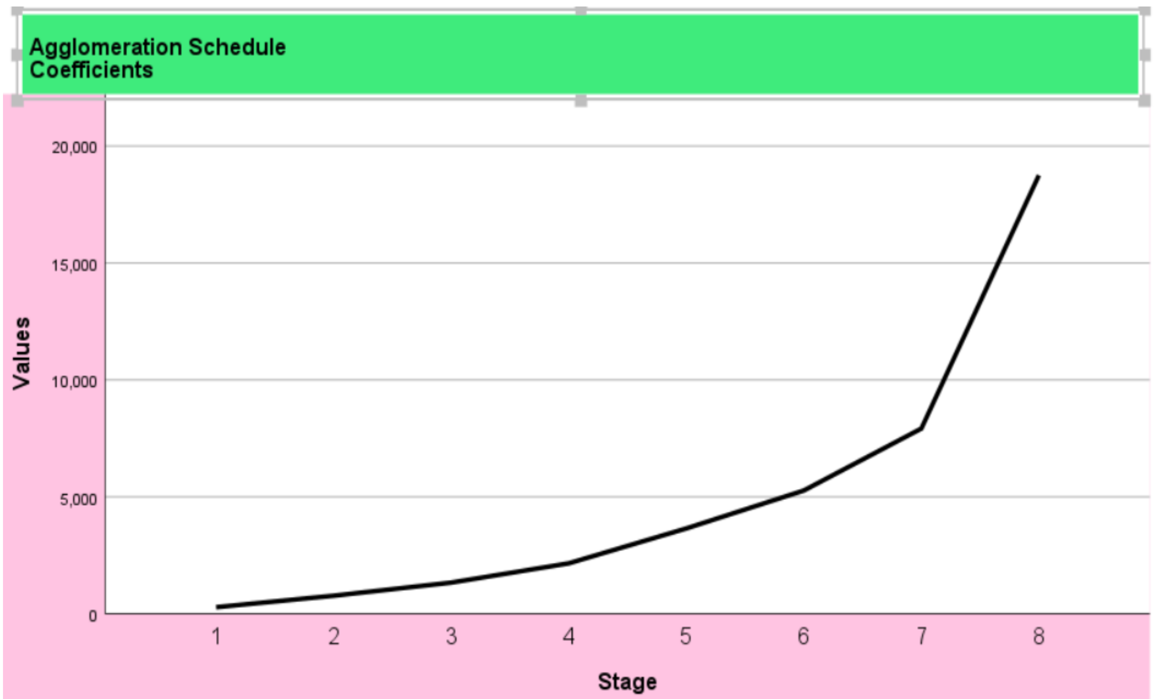


Fig 7.11: The figure shows the plot between the agglomeration coefficients and the different stages.

7.3.13 Dendrogram Plot for Summer Season for Nine Sites

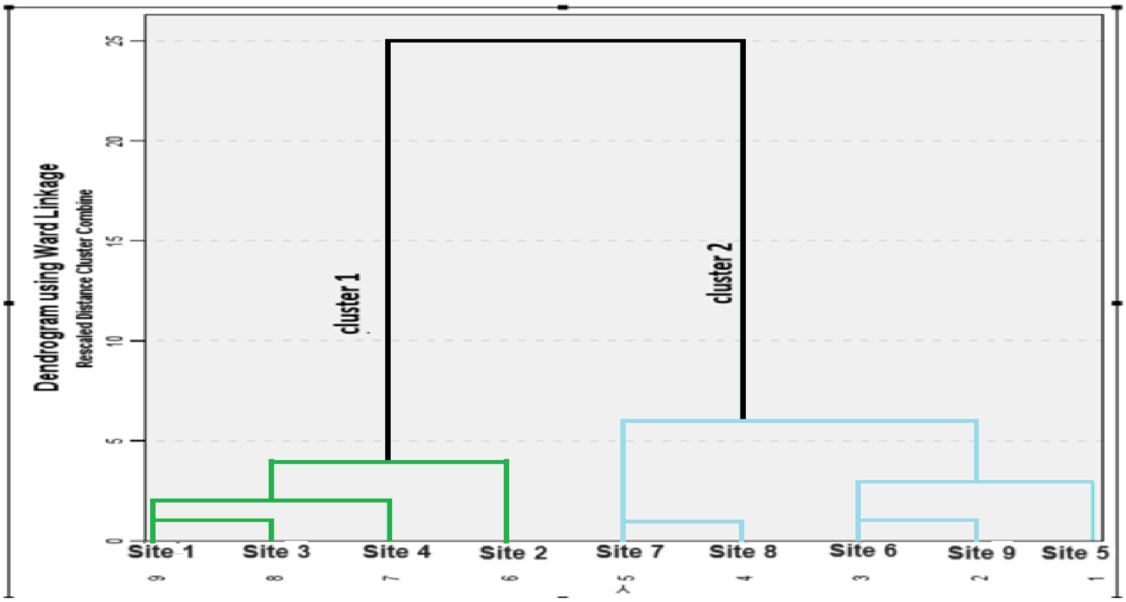


Fig 7.12: The figure shows a tree-like structure which depicts clusters between the sites and the rescaled Euclidean distance. It demonstrates two clusters. The first cluster consists of site 3, site 4, site 1 & site 2 and site 9, site 8, site 5, site 6 & site 7 belong to cluster 2.

7.4 PRINCIPAL COMPONENT ANALYSIS (PCA):

Principal Component Analysis is an unsupervised learning algorithm that is used for the dimensionality reduction in machine learning. It is a statistical process that converts the observations of correlated features into a set of linearly uncorrelated features with the help of orthogonal transformation. These new transformed features are called the **Principal Components**. It is one of the popular tools that is used for exploratory data analysis and predictive modelling. It is a technique to draw strong patterns from the given dataset by reducing the variances.

PCA generally tries to find the lower-dimensional surface to project the high-dimensional data.

PCA works by considering the variance of each attribute because the high attribute shows the good split between the classes, and hence it reduces the dimensionality. Some real-world applications of PCA are image processing, movie recommendation system, optimizing the power allocation in various communication channels**.** It is a feature extraction technique, so it contains the important variables and drops the least important variable.

The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, *the* principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables.

The PCA algorithm is based on some mathematical concepts such as:

- Variance and Covariance
- Eigenvalues and Eigen factors

Some common terms used in PCA algorithm:

- **Dimensionality:** It is the number of features or variables present in the given dataset. More easily, it is the number of columns present in the dataset.
- **Correlation:** It signifies that how strongly two variables are related to each other. Such as if one changes, the other variable also gets changed. The correlation value ranges from -1 to +1. Here, -1 occurs if variables are inversely proportional to each other, and +1 indicates that variables are directly proportional to each other.
- **Orthogonal:** It defines that variables are not correlated to each other, and hence the correlation between the pair of variables is zero.
- **Eigenvectors:** If there is a square matrix M, and a non-zero vector v is given. Then v will be eigenvector if Av is the scalar multiple of v.
- **Covariance Matrix:** A matrix containing the covariance between the pair of variables is called the Covariance Matrix.
- **Scree plot:** It is a graphical that shows the explained variance per newly defined component (principal component). The measure of the plot can be the percentage or the absolute value of the explained variance (eigenvalues). It is common in practice that the first few principal components explain the major amount of variance.

Principal component analysis is used to decrease the dimensional space of the large dataset in order to improve the clustering. In PCA analysis, there are three components, 13 physico-chemical parameters were categorized. PCA's classified the factor loadings as 'strong', 'moderate' and 'weak', matching to absolute loading values of >0.75, 0.75-0.50 and 0.50-0.30, respectively (Liu et al., 2003). PCA is done for four different seasons i.e., Autumn, Winter, Spring and Summer. The results of calculations are shown in Table 7.5,7.6,7.7 & 7.8. According to Hair et al. (2009), the choice of the number of major components to be retained in the number of major components released before a clear break between scree.

7.4.1 PCA for Autumn

PCA revealed that three components explain 84.620% of the total variance, with the salinization process and anthropogenic activities being the main factors controlling the surface water quality variability. The PCA results are shown in Table 7.5. The PCA approach identified three components that have the most critical loading (Fig. 7.14).

Table 7.5: PCA Analysis for Autumn Season

Variable	**PC 1**	**PC 2**	**PC 3**
EIGEN VALUES	5.392	4.432	1.177
% of Variance	41.477	34.091	9.052
Cumulative %	41.477	75.568	**84.620**
TEMPERATURE(0C)	**0.810**	-0.277	-0.223
pH	**0.896**	0.003	-0.099
DO	-0.030	**-0.959**	-0.048
TDS	**0.783**	-0.296	0.235
SALINTY	**0.970**	-0.104	0.125
EC	**0.893**	0.276	0.015
BOD	-0.086	**0.868**	0.395
TOTAL HARDNESS	-0.162	0.089	-0.078
CHLORIDE	0.054	**-0.938**	-0.289
IRON	-0.158	0.740	0.697
NITRATE	-0.055	**0.757**	0.658
LEAD	0.025	0.377	**0.886**
TURBIDITY	0.642	0.135	**0.737**

SCREE PLOT OBTAINED FROM PCA

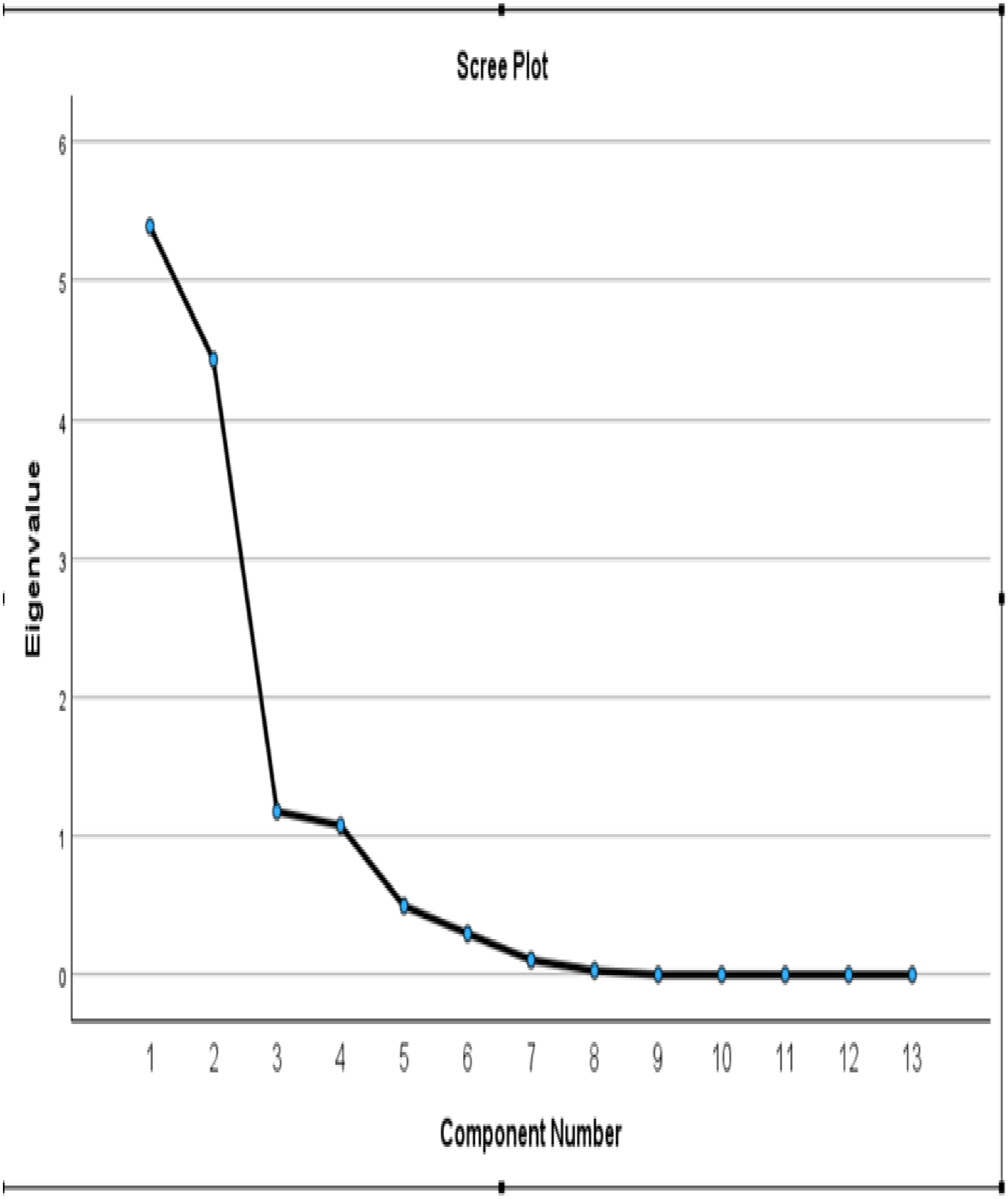


Fig 7.13: The above plot represents the component number along with the eigen values in our study area. The eigen values greater than 1 are considered as the principal components. Here, in our study area the principal number of components present is two set of components.

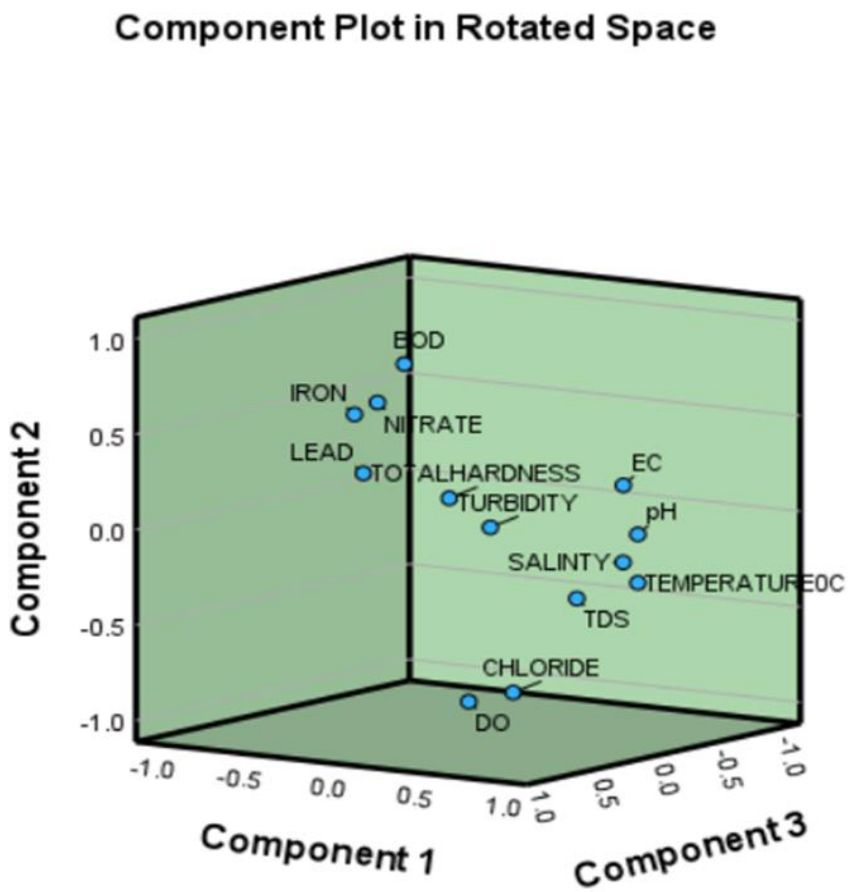


Fig 7.14: Component plot

Fig 7.14 shows the Statistical assessment of water quality parameters for pollution source identification for autumn season.

From the table 7.5, the first component (PCA1) explains 41.477% of the total variance and encompasses the following main parameters EC, TDS, salinity, and pH were strongly related. The significant variables (EC, TDS, salinity, pH) within PCA1 followed the same direction and showed a major increase related to salinity. PCA1 demonstrates that the salinization process is the main factor controlling the surface water quality variability and the importance of mineralization process.

The second component (PCA2) explains 34.091% of the total variance and was assembled by DO, BOD, Chloride, Iron and Nitrate showing high correlations among themselves towards the same direction (Table 7.5). PCA2 demonstrates the high concentration of BOD and sets an inverse correlation with DO which seems an increased cause of pollution and a higher concentration of iron and nitrate also effecting the water quality.

The third component (PCA3) accounts for 9.052% of the total variance and describes the significant contributions of Lead and Turbidity (Table 7.5), disclosing good correlations among themselves. It shows that higher lead concentration effecting turbidity of surface water.

7.4.2 PCA for Winter Season

The winter PCA reveals the three components explaining **78.536**% of the total variance. The PCA results are shown in Table 7.6. The PCA approach identified three components that have the most critical loading (Fig. 7.16).

Table 7.6: PCA Analysis for Winter Season

Variable	PC 1	PC 2	PC 3
EIGEN VALUES	5.314	3.286	1.609
% of Variance	40.881	25.275	12.380
Cumulative %	40.881	66.155	**78.536**
TEMPERATURE(0C)	-0.006	0.027	0.948
pH	-0.080	**0.904**	0.220
DO	-0.499	-0.030	-0.702
TDS	-0.401	0.682	-0.359
SALINTY	0.229	**0.856**	-0.035
EC	0.077	**0.845**	0.055
BOD	**0.724**	0.203	0.446
TOTAL HARDNESS	0.052	-0.059	-0.064
CHLORIDE	-0.694	0.470	-0.201
IRON	**0.984**	-0.085	0.050
NITRATE	**0.967**	0.015	0.212
LEAD	**0.935**	-0.076	0.038

TURBIDITY	**0.807**	0.454	-0.056

SCREE PLOT OBTAINED FROM PCA

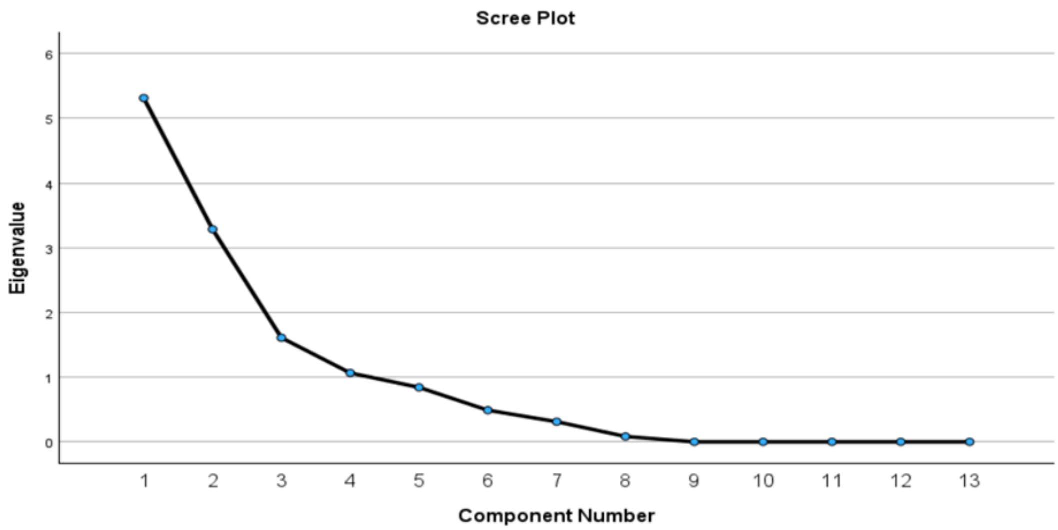


Fig 7.15: Scree-plot for the principal component model of the monitoring data

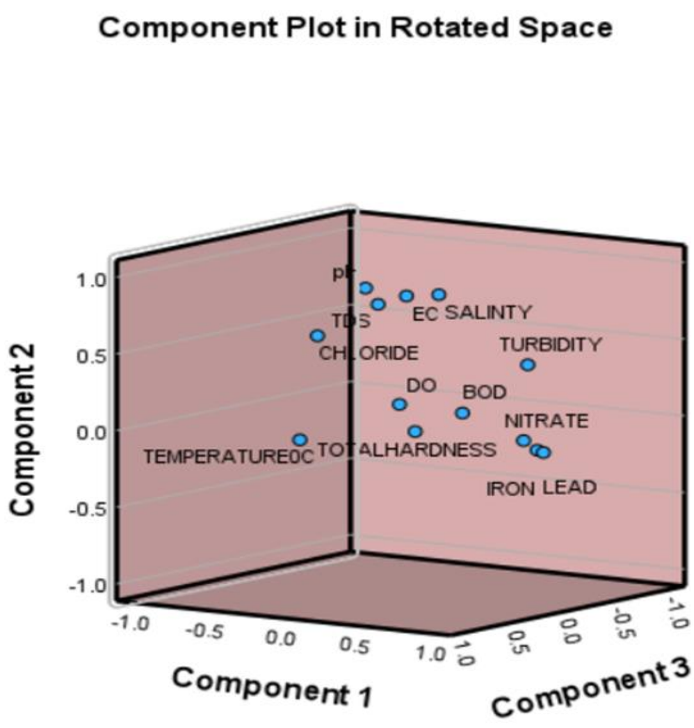


Fig 7.16: Component plot

Fig 7.16 shows the Statistical assessment of water quality parameters for pollution source identification for winter season.

From the table 7.6, the first component (PCA1) explains **40.881**% of the total variance and encompasses the following main parameters iron, nitrate, lead and turbidity were strongly related. PCA1 demonstrates metallic pollution due to which turbidity is also increased.

The second component (PCA2) explains **25.275**% of the total variance and was assembled by pH, Salinity and EC showing high correlations among themselves towards the same direction (Table 7.6).

7.4.3 PCA for Spring Season

The spring PCA reveals the three components explaining **76.485**% of the total variance. The PCA results are shown in Table 7.7.

Table 7.7: PCA Analysis for Spring Season

Variable	PC 1	PC 2	PC 3
EIGEN VALUES	**5.394**	**2.887**	**1.663**
% of Variance	41.489	22.207	12.789
Cumulative %	41.489	63.696	**76.485**
TEMPERATURE(0C)	-0.235	-0.131	0.145
pH	0.183	**0.754**	0.204
DO	-0.278	-0.047	**-0.943**
TDS	-0.502	0.717	-0.135
SALINTY	0.151	0.724	-0.017
EC	0.098	**0.857**	0.220
BOD	0.508	0.364	**0.751**
TOTAL HARDNESS	0.105	0.087	-0.283
CHLORIDE	-0.695	0.430	-0.290

IRON	**0.949**	-0.048	0.192
NITRATE	**0.946**	0.033	0.282
LEAD	**0.904**	0.128	0.111
TURBIDITY	**0.834**	0.428	0.123

SCREE PLOT OBTAINED FROM PCA

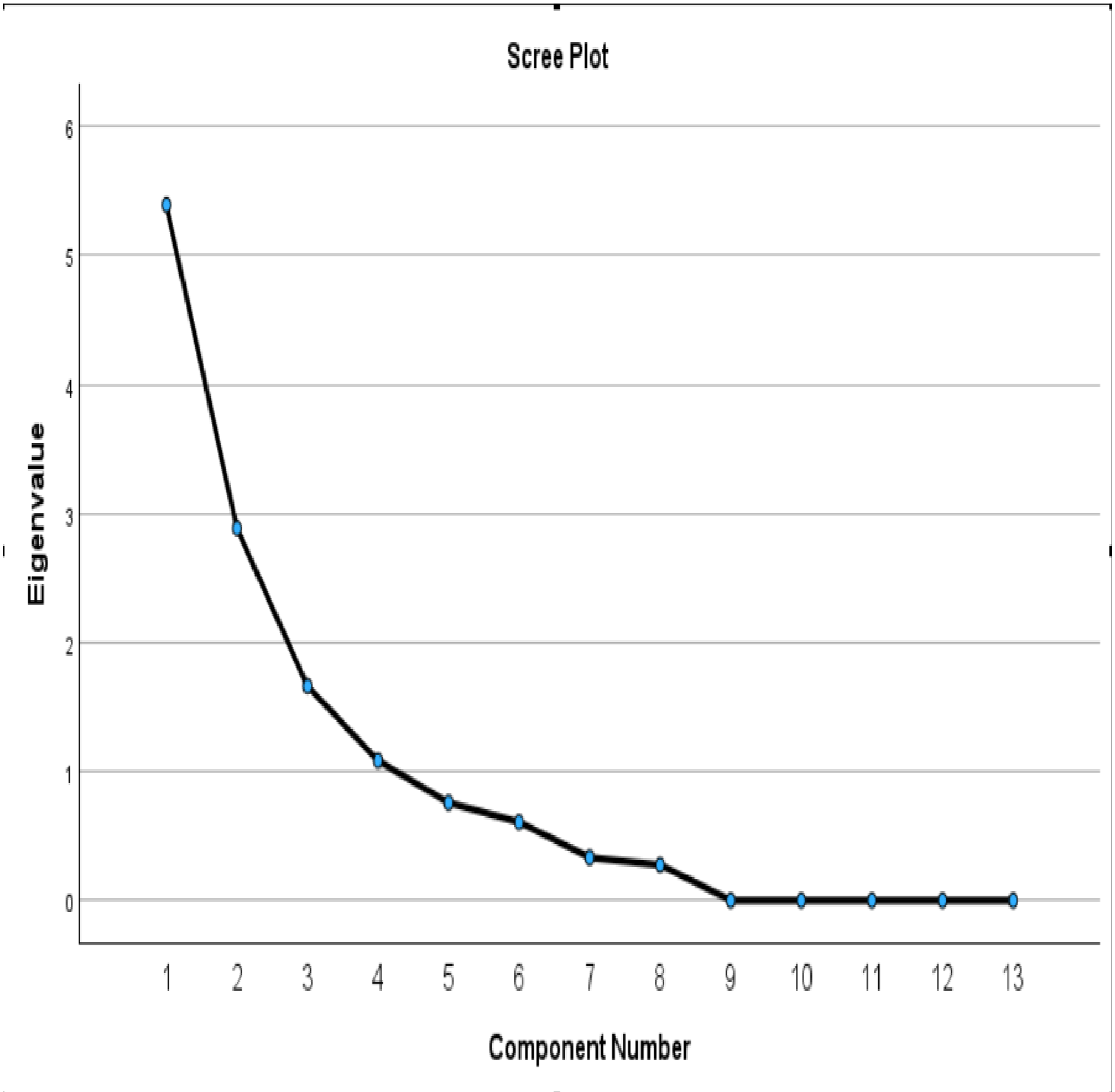


Fig 7.17: Scree-plot for the principal component model of the monitoring data

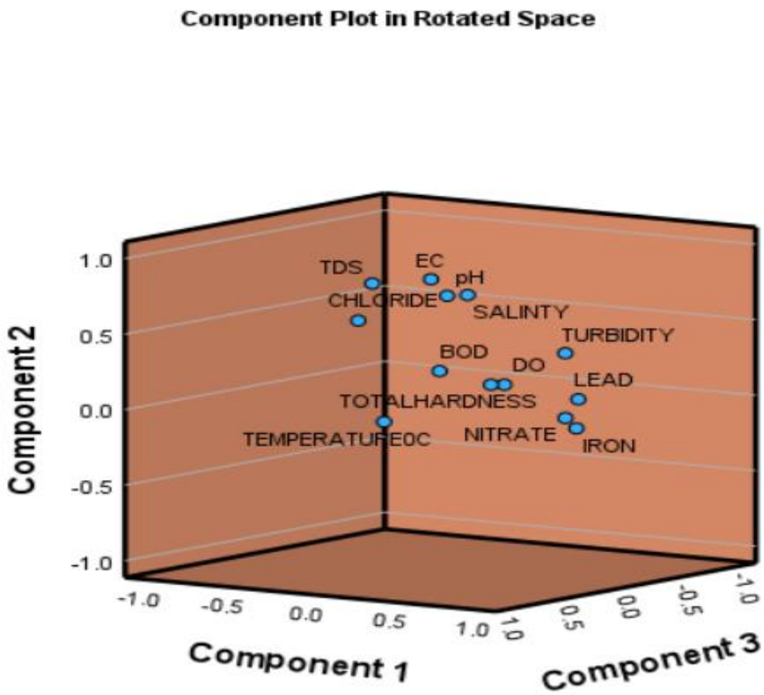


Fig 7.18: Component plot

Fig 7.18 shows the Statistical assessment of water quality parameters for pollution source identification for spring season.

From the table 7.7, the first component (PCA1) explains **41.489**% of the total variance and encompasses the following main parameters iron, nitrate, lead and turbidity were strongly related. PCA1 demonstrates metallic pollution due to which turbidity is also increased.

The second component (PCA2) explains **22.207**% of the total variance and was assembled by pH, and EC showing high correlations among themselves towards the same direction (Table 7.7).

The third component (PCA3) accounts for **12.789**% of the total variance and describes the significant contributions of DO and BOD. It sets an inverse correlation among them.

7.4.4 PCA for Summer Season

The summer PCA reveals the three components explaining **79.485**% of the total variance. The PCA results are shown in Table 7.8.

Table 7.8: PCA Analysis for Summer Season

Variable	PC 1	PC 2	PC 3
EIGEN VALUES	5.276	3.175	1.882
% of Variance	40.584	24.423	14.477
Cumulative %	40.584	65.007	**79.485**
TEMPERATURE(0C)	-0.288	0.255	0.868
pH	**0.939**	0.093	0.135
DO	**-0.826**	-0.014	0.023
TDS	-0.407	**0.750**	0.133
SALINTY	0.185	0.734	-0.080
EC	0.354	**0.826**	0.284
BOD	**0.897**	0.064	0.140
TOTAL HARDNESS	-0.271	0.388	**-0.844**
CHLORIDE	-0.076	0.612	0.713
IRON	**0.748**	-0.184	-0.411
NITRATE	**0.917**	-0.016	-0.094
LEAD	-0.623	-0.525	0.202
TURBIDITY	0.705	0.370	-0.154

SCREE PLOT OBAINED FROM PCA

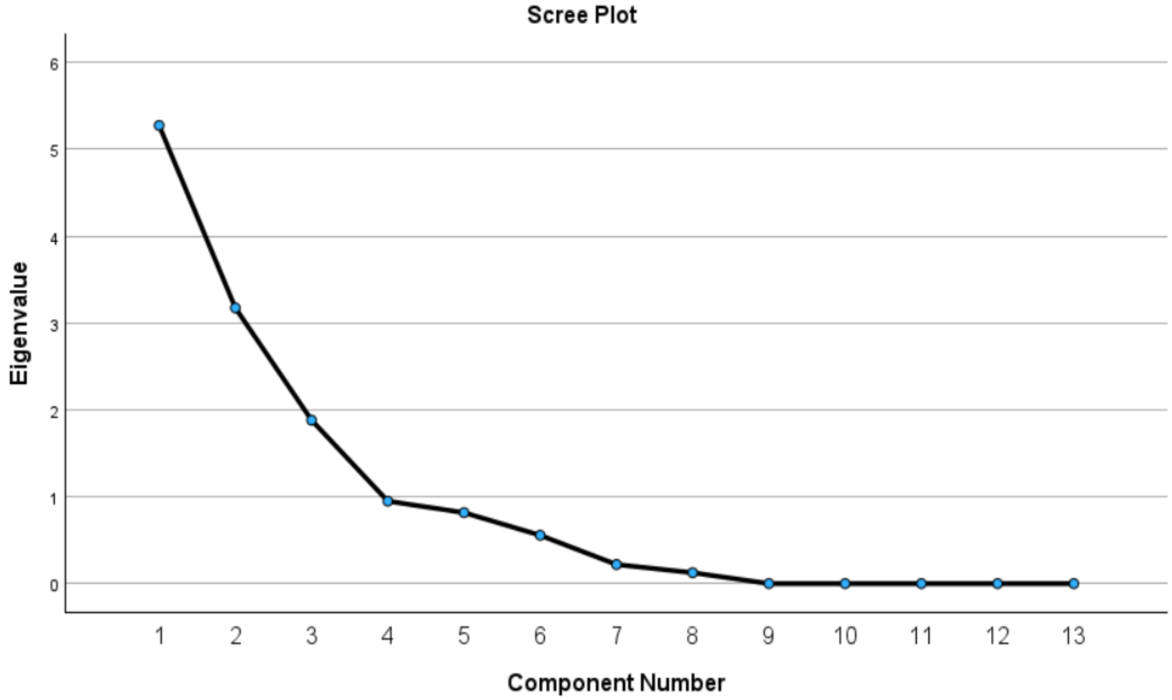


Fig 7.19: Scree-plot for the principal component model of the monitoring data

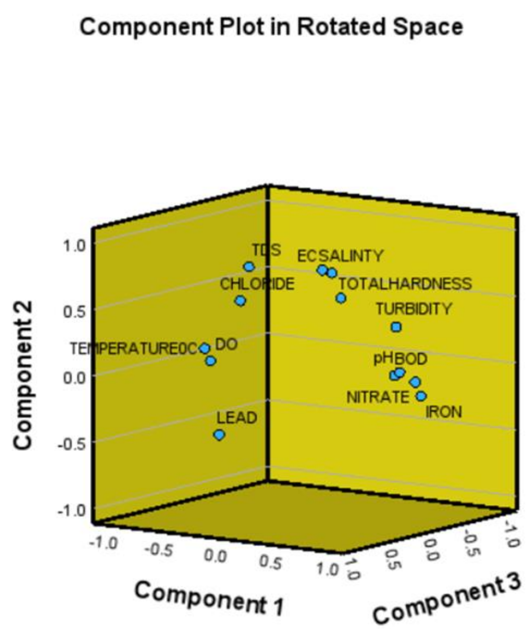


Fig 7.20: Component plot

Fig 7.20 shows the Statistical assessment of water quality parameters for pollution source identification for summer season.

From the table 7.8 the first component (PCA1) explains **40.584**% of the total variance and encompasses the following main parameters pH, BOD, iron, nitrate were strongly related and DO is negatively related with BOD.

The second component (PCA2) explains **24.423**% of the total variance and was assembled by TDS and EC showing high correlations among themselves towards the same direction (Table 7.8).

The third component (PCA 3) demonstrates a strong negative value of total hardness.

7.5 K-MEANS CLUSTER ANALYSIS:

K-Means Clustering is an unsupervised learning algorithm which groups the unlabelled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process, as if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on.

It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabelled dataset on its own without the need for any training.

It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters.

The algorithm takes the unlabelled dataset as input, divides the dataset into k-number of clusters, and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm.

The k-means clustering algorithm mainly performs two tasks:

- Determines the best value for K center points or centroids by an iterative process.
- Assigns each data point to its closest k-center. Those data points which are near to the particular k-center, create a cluster.

Hence each cluster has datapoints with some commonalities, and it is away from other clusters.

The spatial distribution, the descriptive statistics (Table 7.9), and the graphical representation of the means (Fig. 7.21) of the four detected seasons signify the spatial variability of the hydrochemistry among the seasons. The first season Autumn (9 observations) exhibits higher concentration of nitrate(0.07mg/l) and lead(0.07mg/l) relative to other seasons. The nitrate concentration is under permissible limit (<45mg/l) but lead concentration seems to be beyond permissible limit (>0.01mg/l). The second season exhibits higher concentration of DO (5mg/l) which signifies lesser pollution in winter season. Total hardness is also found in higher concentration (119.11mg/l). As per BIS total hardness ranges between 60mg/l to 120mg/l are classified as soft water. The winter season also consumes higher concentration of iron (1.5mg/l) which goes beyond the permissible limit of iron (>0.3 mg/l). The third season i.e., spring season samples are highly loaded with EC concentration (0.29 mg/l) and turbidity (9.78 NTU). Lastly the summer season's samples are bringing highest pH level (7.34), TDS (198.06 mg/l), salinity (0.27 mg/l), BOD (3.11 mg/l) and chloride content (79.11 mg/l). The highest BOD concentration notifies the highest pollution in summer season among all the four seasons.

Table 7.9: Statistics of the four seasons detected from the k-means clustering analysis

Column1	TEMPERATURE(^{0}C)	pH	DO	TDS	SALINTY	EC	BOD	TOTAL HARDNESS	CHLORIDE	IRON	NITRATE	LEAD	TURBIDITY
AUTUMN (n=9)													
MAXIMUM	25.50	7.42	5.60	196.20	0.30	0.39	3.60	130	105	2.78	0.20	0.19	13
MINIMUM	22.50	6.25	3	70	0.13	0.16	1.30	85	30	0.53	0.01	0.02	6
MEAN	24.01	6.90	4.30	151.80	0.21	0.26	2.39	110.22	70	1.34	**0.07**	**0.07**	9.39
Std. Deviation	1.01	0.39	0.91	43.29	43.29	0.09	0.68	15.01	27.84	0.79	0.07	0.06	2.47
WINTER (n=9)													
MAXIMUM	24.20	7.25	6.20	235	0.37	0.41	4.80	140	110	2.83	0.17	0.10	15
MINIMUM	21.50	5.85	3.90	110	0.13	0.18	1.05	90	35	0.63	0.01	0.01	5
MEAN	22.87	6.56	**5**	177.23	0.24	0.27	2.05	**119.11**	67.78	**1.40**	0.05	0.05	9.67
Std. Deviation	0.85	0.49	0.85	39.89	0.09	0.08	1.12	16.94	23.06	0.79	0.06	0.03	3.28
SPRING (n=9)													
MAXIMUM	29.50	7.30	5.90	240	0.39	0.40	5	136	119	2.65	0.20	0.09	14
MINIMUM	28	6.12	3.87	115	0.13	0.20	1.12	86	39	0.59	0.01	0.01	7
MEAN	28.84	6.81	4.79	182.57	0.26	**0.29**	2.47	113.78	72	1.33	0.06	0.04	**9.78**
Std. Deviation	0.50	0.42	0.82	39.93	0.09	0.07	1.43	17.56	24.44	0.76	0.08	0.03	2.44
SUMMER (n=9)													
MAXIMUM	31.40	8	5.45	255	0.41	0.40	6.70	130	100	1.65	0.19	0.05	15
MINIMUM	26	6.95	2.93	130	0.15	0.20	1.15	79	48	0.54	0.01	0.01	5
MEAN	29	**7.34**	4.24	**198.06**	**0.27**	0.28	**3.11**	107.44	**79.11**	1.04	0.05	0.03	9.18
Std. Deviation	1.72	0.37	0.97	40.79	0.10	0.07	1.98	19.60	16.62	0.39	0.07	0.02	3.34

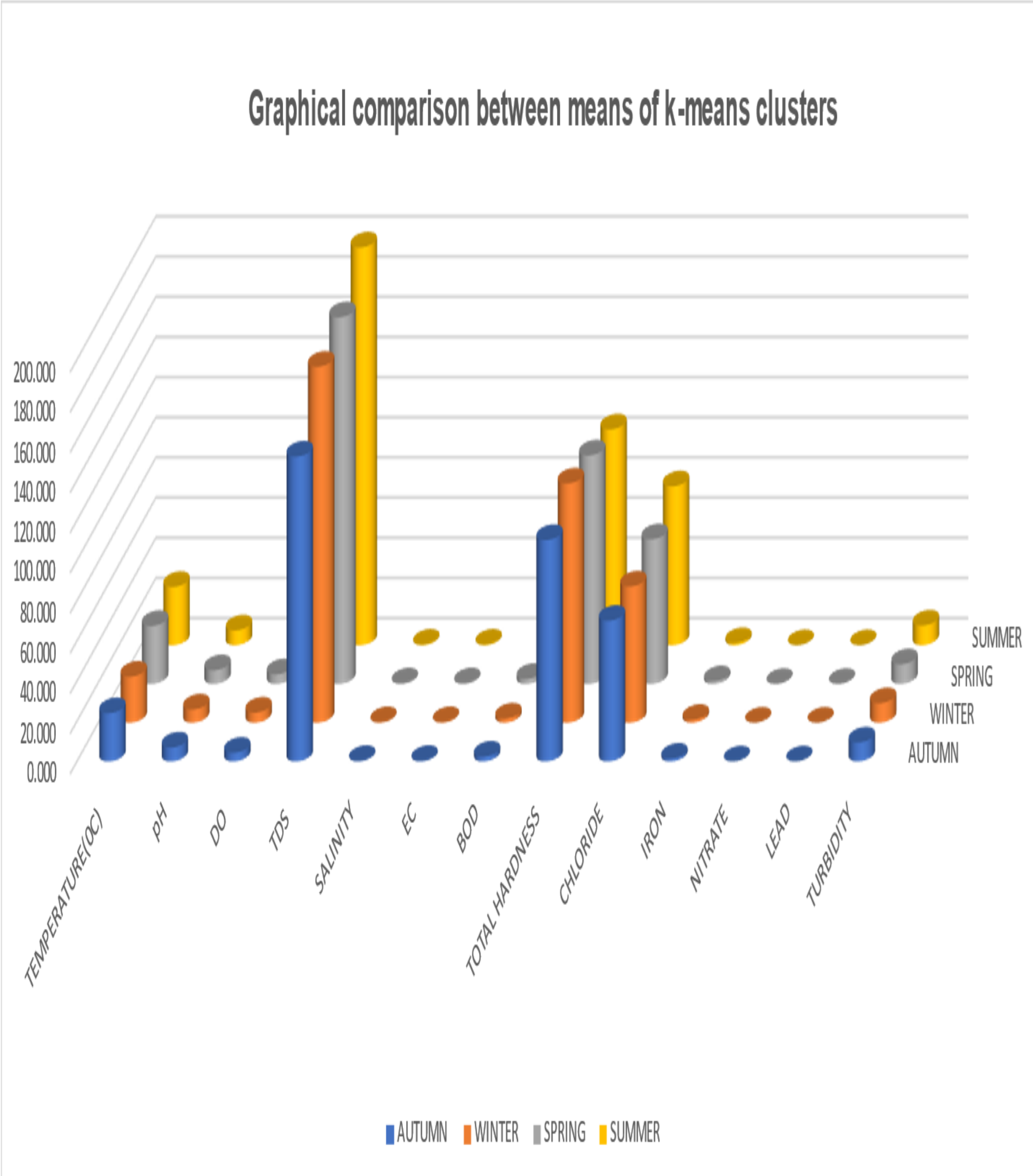


Fig 7.21: Graphical comparison between means of k-means clusters

7.6 Correlation-Matrix (Pearson Correlation Method)

A statistical correlation analysis was performed for the data obtained from the test samples, to check the possible correlation between the sample parameters of Deepor Beel. A correlation matrix is a tabular presentation of the correlation coefficients between the variables i.e., each cell in the table shows the correlation between two variables. In this study, a correlation matrix was formed with the results of all the 13 parameters which is presented in table 7.10,7.11,7.12 & 7.13. The correlation matrix was formed through the Pearson correlation analysis method and the significance test selected was two- tailed test. The correlations were presented based on .01 and .05 significances. A positive correlation indicates the occurrence of two variables in parallel; whereas a negative correlation indicates the extent to which one variable increases as the other decreases.

Table 7.10: Correlation matrix of the water quality parameters in Autumn Season

	TEMPERATURE(0C)	pH	DO	TDS	SALINTY	EC	BOD	TOTAL HARDNESS	CHLORIDE	IRON	NITRATE	LEAD	TURBIDITY
TEMPERATURE(0C)	1												
pH	0.901218609	1											
DO	0.299712993	0.035375	1										
TDS	0.485278587	0.485321	0.141191	1									
SALINTY	0.831690926	0.91419	0.059684	0.780734	1								
EC	0.5358606	0.755325	-0.2753	0.643485	0.813377	1							
BOD	-0.386550239	-0.15429	-0.89909	-0.15898	-0.11412	0.096042	1						
TOTAL HARDNESS	-0.353969443	-0.39715	-0.01509	0.001904	-0.33112	-0.06977	0.06082	1					
CHLORIDE	0.382956692	0.083294	0.880635	0.262068	0.131687	-0.21806	-0.92209	-0.221294751	1				
IRON	-0.49029338	-0.23052	-0.61653	-0.16624	-0.14967	0.126018	0.796581	0.032796048	-0.80475	1			
NITRATE	-0.356798569	-0.08796	-0.65792	-0.16255	-0.06718	0.164856	0.835404	0.134160615	-0.88492	0.931298	1		
LEAD	-0.226751639	0.028878	-0.39253	0.047634	0.134644	0.100343	0.673726	-0.169041467	-0.60773	0.810776	0.842215	1	
TURBIDITY	0.327115942	0.467349	-0.18616	0.656947	0.681648	0.599974	0.373239	-0.053134425	-0.31332	0.494379	0.55893	0.692105	1

Table 7.11: Correlation matrix of the water quality parameters in Winter Season

	TEMPERATURE(0C)	pH	DO	TDS	SALINTY	EC	BOD	TOTAL HARDNESS	CHLORIDE	IRON	NITRATE	LEAD	TURBIDITY
TEMPERATURE(0C)	1												
pH	0.307450541	1											
DO	-0.555752502	-0.01504	1										
TDS	-0.371513646	0.482207	0.224035	1									
SALINTY	0.003095494	0.748126	-0.0877	0.405901	1								
EC	-0.036765915	0.692907	-0.19907	0.578741	0.581778	1							
BOD	0.313652373	0.15485	-0.7872	-0.32572	0.46074	0.271926	1						
TOTAL HARDNESS	-0.043180113	-0.14696	0.199787	-0.07547	-0.15384	-0.04233	-0.15981	1					
CHLORIDE	-0.141582969	0.454395	0.483917	0.574678	0.457333	0.193735	-0.37966	-0.357525766	1				
IRON	0.021536216	-0.16196	-0.53544	-0.45838	0.097742	0.077436	0.696684	0.088168544	-0.78219	1			
NITRATE	0.20199933	-0.02071	-0.59752	-0.50473	0.241031	0.113047	0.796072	0.057395448	-0.6866	0.967148	1		
LEAD	0.119312412	-0.04752	-0.42388	-0.45485	0.187033	-0.14127	0.60552	-0.024807409	-0.63724	0.893522	0.894539	1	
TURBIDITY	0.053916387	0.39241	-0.27503	0.002294	0.51271	0.368768	0.480043	0.178499286	-0.39947	0.752909	0.773778	0.795281	1

Table 7.12: Correlation matrix of the water quality parameters in Spring Season

	TEMPERATURE(0C)	pH	DO	TDS	SALINTY	EC	BOD	TOTAL HARDNESS	CHLORIDE	IRON	NITRATE	LEAD	TURBIDITY
TEMPERATURE(0C)	1												
pH	-0.292845199	1											
DO	0.031390403	-0.25529	1										
TDS	0.134798546	0.34084	0.21625	1									
SALINTY	-0.024302398	0.233133	-0.09555	0.389845	1								
EC	-0.235958784	0.783021	-0.24811	0.503057	0.483367	1							
BOD	-0.217949135	0.522847	-0.86139	-0.16789	0.386822	0.494928	1						
TOTAL HARDNESS	0.440875938	-0.21176	0.323253	0.066862	0.304291	-0.1143	-0.0779	1					
CHLORIDE	-0.097107667	0.221354	0.408365	0.5853	0.244504	0.22083	-0.39736	-0.311264998	1				
IRON	-0.151147086	0.131901	-0.43792	-0.49594	0.090691	0.167087	0.556044	-0.025625438	-0.74372	1			
NITRATE	-0.177639564	0.277332	-0.52434	-0.52033	0.150218	0.218553	0.680878	-0.058052665	-0.67463	0.969838	1		
LEAD	-0.243663929	0.275069	-0.4093	-0.33777	0.268988	0.155474	0.556199	-0.181653208	-0.46095	0.881219	0.908976	1	
TURBIDITY	-0.216348717	0.639053	-0.3665	-0.07558	0.27133	0.444312	0.645046	0.016217854	-0.42787	0.76847	0.829749	0.851049	1

Table 7.13: Correlation matrix of the water quality parameters in Summer Season

	TEMPERATURE(0C)	pH	DO	TDS	SALINTY	EC	BOD	TOTAL HARDNESS	CHLORIDE	IRON	NITRATE	LEAD	TURBIDITY
TEMPERATURE(0C)	1												
pH	-0.175246306	1											
DO	0.325103598	-0.76888	1										
TDS	0.32640073	-0.20294	0.228801	1									
SALINTY	0.10596272	0.10145	-0.24045	0.36748	1								
EC	0.334218545	0.459874	-0.23939	0.53262	0.483815	1							
BOD	-0.209490224	0.856016	-0.93678	-0.21362	0.226895	0.419957	1						
TOTAL HARDNESS	-0.526221581	-0.36356	0.151414	0.199456	0.331127	-0.03393	-0.32536	1					
CHLORIDE	0.862070574	0.027296	0.109246	0.43129	0.491654	0.626353	0.013081	-0.284015075	1				
IRON	-0.488337354	0.561564	-0.46574	-0.57486	0.114502	-0.01772	0.460611	0.055959261	-0.41371	1			
NITRATE	-0.277970555	0.759589	-0.64837	-0.54117	0.334189	0.245094	0.731206	-0.158699502	-0.05098	0.847107	1		
LEAD	0.171776614	-0.62832	0.466164	-0.06686	-0.23696	-0.72858	-0.52452	-0.242951283	-0.10924	-0.47357	-0.513	1	
TURBIDITY	-0.258373456	0.78161	-0.38844	0.003317	0.366383	0.474758	0.479597	0.009343552	0.06473	0.5413	0.674163	-0.5744	1

7.7 WATER QUALITY INDEX

Water quality index indicates single number like a grade that express overall water quality index at certain area and time. It gives general idea of the possible problem with water in a particular region to public.

7.7.1 Calculation of WQI by Weighted Arithmetic Water Quality Index Method

Weighted arithmetic water quality index method classified the water quality according to the degree of purity by using the most commonly measured water quality variables. The method has been widely used by the various scientists and the calculation of WQI was made by using the following equation:

$$WQI = \sum Q_i W_i / \sum W_i$$

The quality rating scale (Qi) for each parameter is calculated by using this expression:

Qi =100[(Vi- Vo)/ (Si- Vo)]

Where,

Vi is estimated concentration of i^{th} parameter in the analysed water.

Vo is the ideal value of this parameter in pure water Vo = 0 (except pH =7.0 and DO = 14.6 mg/l)

Si is recommended standard value of i^{th} parameter

The unit weight (Wi) for each water quality parameter is calculated by using the following formula:

Wi =K /Si

Where, K = proportionality constant and can also be calculated by using the following equation:

K = 1/ ∑ (1/Si)

The rating of water quality according to this WQI is given in Table 7.14.

Table 7.14: Water Quality Rating as per Weighted Arithmetic Water Quality Index Method

WQI	***RATING OF WATER QUALITY***
0-25	Excellent
25-50	Good
51-75	Poor
76-100	Very poor
>100	Unsuitable for drinking purpose

Table 7.15: BIS Standards for Various Water Quality Parameters for Drinking Purpose

PARAMETERS	***BIS STD (Sn)***
P^H	8.5
DO (mg/l)	4
TDS (mg/l)	500
EC (ms/cm)	300

BOD (mg/l)	4
Total Hardness (mg/l)	300
Chloride (mg/l)	250
Iron (mg/l)	0.3
Nitrate (mg/l)	50
Lead (mg/l)	0.1
Turbidity (NTU)	5

Table 7.16: Weighted Arithmetic Water Quality Index Values

SEASON	*WQI (Rating)*								
	SITE 1	*SITE 2*	*SITE 3*	*SITE 4*	*SITE 5*	*SITE 6*	*SITE 7*	*SITE 8*	*SITE 9*
AUTUMN	**96.89** (Very Poor)	**115.79** (Unfit for drinking)	**67.47** (Poor)	**120.63** (Unfit for drinking)	**175.76** (Unfit for drinking)	**289.5** (Unfit for drinking)	**359.59** (Unfit for drinking)	**105.64** (Unfit for drinking)	**114.19** (Unfit for drinking)
WINTER	**90.27** (Very Poor)	**91.03** (Very Poor)	**71.01** (Poor)	**127.24** (Unfit for drinking)	**167.77** (Unfit for drinking)	**278.57** (Unfit for drinking)	**300.35** (Unfit for drinking)	**103.64** (Unfit for drinking)	**112.66** (Unfit for drinking)
SPRING	**88.41** (Very Poor)	**83.66** (Very Poor)	**63.31** (Poor)	**122.64** (Unfit for drinking)	**138.24** (Unfit for drinking)	**268.12** (Unfit for drinking)	**279.81** (Unfit for drinking)	**92.57** (Very Poor)	**97.54** (Very Poor)
SUMMER	**84.31** (Very Poor)	**86.85** (Very Poor)	**63.72** (Poor)	**122.52** (Unfit for drinking)	**132.39** (Unfit for drinking)	**136.66** (Unfit for drinking)	**146.43** (Unfit for drinking)	**107.01** (Unfit for drinking)	**100.76** (Unfit for drinking)

From the above table, it is seen that the Weighted Arithmetic WQI values of Deepor Beel ranges from 63.31 to 359.59. According to Weighted Arithmetic Water Quality Index values water sample of Site 7 i.e., Boragaon Dumping Site (ii) is most polluted among all the collected water sample, which falls under "Unfit for drinking" rating.

7.8 DISCUSSION

As per the work titled "Assessment of water quality trends in Deepor Beel, Assam, India" by the author Ritabrata Roy et al. analyzed total 12 parameters by collecting sample water from 10 different sites for a period of 3 years. From their study the PH value is observed to be higher in summer days than winter. EC was found to be higher during post monsoon than at any other time of the year probably due to input of various salts through runoff during monsoon. DO value was lower during summer and comparatively higher during winter. Total suspended solids were seen to be increased in post monsoon season. TDS increased in monsoon compared to other seasons of a year.

As per the author S. Priya et al. journal titled “Analysis of water Quality in selected sections along Kanyakumari District, Tamil Nadu, India” various water quality parameters like Electrical Conductivity, hardness, PH, Alkalinity showed some variations during the months of sample collection and among the stations. It was observed that EC value was high during July and low during the month of October. Analysis of variance was carried out between seasons. which showed significant difference. TDS values are observed higher in summer days. Statistical analysis by Two-way ANOVA on TDS of water as a function of variation between stations are statistically significant i.e., F – ratio > F- critical and $p<0.05$. Analysis of PH and alkalinity was carried out among the different seasons. Statistical analysis by ANOVA on PH of water as a function of variation between stations and seasons are statically insignificant. The analysis done on alkalinity parameter showed statistically significant difference between the different season in the year.

As per author Uncumusaoğlu et. Al journal titled “Statistical assessment of water quality parameters for pollution source identification in Bektaş Pond (Sinop, Turkey)”, has researched on 21 physico-chemical and 7 heavy metal parameters obtained from four different sampling points for one year in water of Bektaş Pond by using multivariate statistical methods such as ANOVA, Pearson correlation (PC), Hierarchical cluster analysis (HCA) and Principal Component Analysis (PCA) to determine water quality as well as suitability of water for aquatic life. From his study, it was revealed that main pollution source of the pond water is non-point pollution i.e., agricultural pollution and soil leaching for this region. Result of HCA shows no significant difference between the stations but has a significant difference between seasons. A suggested solution to the problems is “best environmental practice” principle may be applied to minimize the out-of-source pollution and to efficiently use and control stocks of freshwater resources.

As per Tiri et. al (2015), in their work titled “Assessment of the quality of water by hierarchical cluster and variance analyses of the Koudiat Medouar Watershed, East Algeria”, have studied about the spatial and temporal variation of water surface quality of the Koudiat Medouar Watershed, Eastern Algeria, to which end, multivariable method such as HCA and ANOVA was used. The overall evaluation during the study period showed alkaline nature of water in the area.

Higher EC value was observed in water collected from sampling station 2. The ANOVA results indicate that all of the water quality parameters are significant except for Na, K and HCO_3 in the station 1 and EC in the station 2 and pH and NO_3 in the last station.

As per the author N.norom et. al in their article "**Multivariate statistical approach and water quality assessment of natural springs and other drinking water sources in Southeastern Nigeria**", have investigated the physico-chemical and trace elements contents of ground and surface water sources used for domestic purposes. In this connection, multivariate statistical analysis such as ANOVA, HCA etc., were used to characterize potentially toxic elements (PTEs) in natural springs and other drinking water sources based on likely origin, inter relationships, extent of involvement in water contamination and PTE's ability to cause harm to unsuspecting consumers. The result showed Fe levels were above its permissible limit in about 92% of samples from streams. Overall, Fe, Al, Mn, Se and Zn were the dominant elements. Moreover, Water Quality Index approach indicated that all drinking water sources had either excellent or good water quality with the exception of a borehole, which had poor water quality

In my study area near Deepor Beel, nine stations were selected for the analysis to be carried out. Nine water samples were collected at a depth of 1m to 1.5m in depth were collected using standard methods of collections and procedures (Islam M et. al, 2014). These samples were tested for thirteen different physico-chemical and heavy metal water parameters. Various multivariate data analyses were carried out for these nine different groundwater samples. These analyses were being interpreted in IBM SPSS software package downloaded from the official website of IBM. Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA), K- Means Cluster analysis and One way ANOVA., the observed PH ranges from 5.85 to 8. Highest value of pH i.e., Acidic water is noted in winter season and in summer season, water is found to be alkaline. Total hardness of the sampling sites is within the permissible limits. Total Hardness of water is highest in site 5, whereas, other sampling sites contain lesser value of Total Hardness. The D.O. values range from 2.93 mg/l to 6.20 mg/l which is below the permissible limit (6.5-8 mg/l), with Site 1 having the highest value in winter season and Site 7 having the lowest value in summer season. The T.D.S. values are ranged from 70 mg/l to 255 mg/l which is within the permissible limit as stipulated by BIS Standards, with Site 1 recording the highest value in summer season and Site 8 recording the lowest value in autumn season. Salinity content in the sampling stations is in the range of 0.128 to

0.405 mg/l and it is within the permissible limit among all the stations throughout the seasons. The highest value was recorded in site 7 in summer season and lowest value was recoded in site 9 in winter season. The E.C. concentration is ranged from 0.155 mg/l to 0.405 mg/l which is within the permissible limit as stipulated by BIS Standards, with Site 1 recording the highest value in winter season and Site 8 recording the lowest value in autumn season. The B.O.D. values range from 1.05 mg/l to 6.70 mg/l, with Site 7 recording the highest value in summer season and Site 1 recording the lowest value in winter season. The Chloride Content range from 30 mg/l to 119 mg/l, which are within permissible limits, with Site 2 recording the highest value in spring season and Site 6 recording the lowest value in autumn season. The Iron Content range from 0.53 mg/l to 2.83 mg/l, which are beyond permissible limits, with Site 7 recording the highest value in winter season and Site 2 recording the lowest value in autumn season. The Nitrate Content range from 0.20 mg/l to 0.005 mg/l, which are within permissible limits, with Site 7 recording the highest value in autumn and spring season and Site 1 recording the lowest value in winter season. The Lead Content range from 0.009 mg/l to 0.19 mg/l, with Site 7 recording the highest value autumn season and Site 1 recording the lowest value in spring season. The Turbidity values range from 5 NTU to 15 NTU, which are beyond permissible limits, with Site 6 recording the highest value in winter & summer season and Site 9 recording the lowest value in winter & summer season.

One-Way ANOVA analysis in this study was carried out on basis of temporal variation. The parameters like pH, Temperature, TDS, Salinity, EC, Iron and BOD and DO have significant variation among all the nine sites in the four seasons. The F-ratio of these parameters are greater than 1.0 and $P=0$. As F-ratio >> 1, so there is significant variation in the water variables at different period of testing i.e., values of the samples vary significantly from one another at different seasons. Moreover, Nitrate, chloride and lead have F value less than 1 and p-value>.05, which indicates that it is nearer to hold the null hypothesis, and there are no significant differences in the values in different period of sampling.

Pearson correlation matrix is plotted in Table 7.10, 7.11, 7.12 and 7.13, which shows the various correlation coefficients among the parameters. A matrix is plotted to check the linear dependency among the parameters. The tables Shows the Pearson Correlation Matrix for autumn, winter, spring and summer season respectively. The Pearson Correlation Coefficient ranges from -1 to +1. When the value of coefficient is 0, we conclude that there is no correlation between two variables. When coefficient of variable is +1, there is a perfectly positive linear correlation. When the value is -1, we can conclude that there is perfectly negative linear correlation i.e., if one variable increases the other one tends to decrease.

Table 7.10 shows that there is positive linear dependency between P^H and hardness, chloride, dissolved oxygen, B.O.D as this are coloured dark green. Cells coloured dark red shows a negative dependency i.e., one variable is inversely dependent to other one. Dissolved oxygen and B.O.D show negative correlation coefficient with chloride content as the cells are red. This means these parameters shows weak relation to each other. Similarly, from the remaining tables the green-coloured cells are indicating positive correlation between the parameters and red coloured cells are showing the parameters are inversely proportional to each other.

From the dendogram of HCA analysis, the variation of two cluster depicts the pollution. The cluster showing higher variation is more polluted than that of the other one, which means the sites, which belong to higher cluster are more polluted. In fig 7.3 (Autumn Season), the cluster 1 showing higher variation than cluster 2. So, we can conclude that the sites (Site 5,6,7,8 & 9) belonging to this cluster are having similar properties in the tested parameters and these sites are more polluted than the sites belonging to cluster 2 (Site 1,2,3 & 4). In fig 7.6 (Winter Season), the cluster 1 showing higher variation than cluster 2. So, we can conclude that the sites (Site 1,2,3,4,8 & 9) belonging to this cluster are having similar properties in the tested parameters and these sites are more polluted than the sites belonging to cluster 2 (Site 5,6 & 7). In fig 7.9 (Spring Season), the cluster 2 showing higher variation than cluster 1. So, we can conclude that the sites (Site 5,6,7,8 & 9) belonging to this cluster are having similar properties in the tested parameters and these sites are more polluted than the sites belonging to cluster 1 (Site 1,2,3 & 4). In fig 7.12 (Summer Season), the cluster 2 showing higher variation than cluster 1. So, we can conclude that the sites (Site 5,6,7,8 & 9) belonging to this cluster are having similar properties in the tested parameters and these sites are more polluted than the sites belonging to cluster 1 (Site 1,2,3 & 4).

From the Principal Component Analysis, we can predict the parameters which are affecting more and the parameters which are less effective. From the table 7.5, it is seen that pH, TDS, Salinity, EC, BOD, Nitrate, Turbidity and Lead are strongly effective and DO & Chloride are less effective in autumn season. From the table 7.6, it is seen that pH, Salinity, EC, BOD, Iron, Nitrate, Turbidity and Lead are strongly effective in winter season. From the table 7.7, it is seen that pH, EC, BOD, Iron, Nitrate, Turbidity and Lead are strongly effective and DO is less effective in spring season. From the table 7.8, it is seen that pH, TDS, EC, BOD, Iron and Nitrate are strongly effective and DO & Total Hardness are less effective in summer season.

From the K-Means analysis, spatial variability of the hydrochemistry among the seasons are shown. The first season Autumn (9 observations) exhibits higher concentration of nitrate(0.07mg/l) and lead(0.07mg/l) relative to other seasons. The nitrate concentration is under permissible limit (<45mg/l) but lead concentration seems to be beyond permissible limit (>0.01mg/l). The second season exhibits higher concentration of DO (5mg/l) which signifies lesser pollution in winter season. Total hardness is also found in higher concentration (119.11mg/l). As per BIS total hardness ranges between 60mg/l to 120mg/l are classified as soft water. The winter season also consumes higher concentration of iron (1.5mg/l) which goes beyond the permissible limit of iron (>0.3 mg/l). The third season i.e., spring season samples are highly loaded with EC concentration (0.29 mg/l) and turbidity (9.78 NTU). Lastly the summer season's samples are bringing highest pH level (7.34), TDS (198.06 mg/l), salinity (0.27 mg/l), BOD (3.11 mg/l) and chloride content (79.11 mg/l). The highest BOD concentration notifies the highest pollution in summer season among all the four seasons.

From the WQI analysis, we have checked the quality of water in nine different sites for four seasons. Sites are showing different water quality index in different season. From the table 7.16, it is seen that some sites are performing very poorly in terms of quality throughout all seasons, whereas, the two sites (site 6 and site 7) from Boragaon Dumping Station show the water quality rating as "Unfit for Consumption", which means water samples collected from these sites are not usable for any purpose, not good even for aquatic life.

The analyses and statistical tests conducted have resulted in the protection of the lake water only depending on the control of the amount and content of the fertilizers used in agriculture activities and the effect of pH changes on the aquatic ecosystem due to the sudden temperature changes as a result of changing the climate. A suggested solution to the problems is "best environmental practice" principle should be applied to minimize the out-of-source pollution and to efficiently use and control stocks of freshwater resources.

CHAPTER 8

CONCLUSION

Several approaches are used to analyze the water quality for Deepor Beel, such as Pearson Correlation, Multivariate Analysis of variance (one way ANOVA) with Discriminant Analysis, Principal Component Analysis and Factor Analysis, Hierarchical Cluster Analysis (HCA), K-Means Cluster Analysis and Overall Water Quality Index (WQI) by Weighted Arithmetic Water Quality Index Method. Conclusions from the present research work can be listed as follows: -

- Multivariate statistical techniques (ANOVA) are used to examine spatial and temporal variations in water quality. Discriminant analysis on the gauging stations show that there is small difference between the sampling stations throughout four seasons investigated. This suggests that the observed variability in the water quality (especially with respect to pH, Electrical Conductivity, and Biological Oxygen Demand) is brought on by anthropogenic activities, mostly the effluents of industries, runoff from agricultural lands and waste water from residential areas into the lake.
- Hierarchical Cluster Analysis groups the sampling sites into two clusters for each season based on similarity sites. The spatial pattern shows that some of the sites had the lowest level of pollution while other sampling sites have higher level of pollution. According to the water quality parameters; Electrical Conductivity, TH, FC and BOD affect more than that of the other parameters.
- The Pearson Correlation Analysis shows that there is moderate correlation between the parameters due to changes in land use, mining and improper effluent discharge in the river. When parameters exhibit strong or moderate correlation, explicit numerical representation of the input and output parameters is almost impossible and WQI may not effectively characterize the quality of water. Therefore, it is vital to convert correlated parameters into uncorrelated parameters for efficient forecasting of water quality. PCA provides a suitable method to transform correlated parameters into uncorrelated parameters.
- It is seen that pH value in spring, summer and autumn season is within permissible limit as per BIS. But in winter, the values are below the permissible range which indicates that the water is slightly acidic, so, the water in this season is not good for aquatic life as well as drinking purpose. Hardness is found higher in winter season than in summer. The higher

the concentration of hardness, greater the risk to aquatic life and human health. But it is seen to be under permissible limits. Hence, less effective. TDS is varying lower to higher from autumn, winter, spring and summer season respectively. DO and BOD are sharing a strong negative correlation. DO is found higher in winter and spring season, whereas, BOD calculation for these two seasons seemed to be lower, which indicates water is not polluted in these two seasons. However, the other two seasons i.e., autumn and summer is recorded to be most polluted seasons. Nitrate and lead exhibits higher concentration in autumn season. Nitrate is within permissible limits, but lead concentration showing beyond permissible limits. Higher iron concentration observed in winter season, which goes beyond permissible limit.

- The Water Quality Index (WQI) values for the gauging stations vary from poor to unfit for consumption. The WQI values are found higher in autumn season and lower in spring season. The ranges of water quality parameters were varied from within the range to beyond the range as per Weighted Arithmetic Water Quality Index Method. The site near Tetelia (site 3) shows comparatively good quality than the other sites and the water can be used for domestic purposes. However, the two sites at Boragaon dumping station (site 6 and site 7) are showing water quality as unfit for consumption. This is due to hazardous waste disposal and its pollution arose from the waste disposal. The water quality of these two sites is not only harmful for domestic purpose but also for aquatic life.
- This makes it clear that Deepor Beel requires a better management framework placed into effect as soon as possible in order to enhance and survive. The results of the study demonstrated that even if anthropogenic interventions in the wetland have not been particularly significant, it will already be too late if timely and effective action is not done to restrict future contaminations.

REFERENCES

1) Ali S, Singh R. K. and Kalamdhad A. S. (2015), “Water quality analysis of Disposal site and its adjacent area of Guwahati, Assam, India”, *International Research Journal of Environment Sciences* Vol. 4(5), 12-17, May (2015).
2) Ansari, N.A. (2017) “Seasonal Variations in Physicochemical Characteristics of Water Samples of Surajpur Wetland, National Capital Region, India”, *Int.J. Curr.Microbiol. App.Sci*(2017) 6(2): 971-987.
3) Barman D., Roy B. and Roy S. (2015), “Seasonal Variation of Physico-Chemical characteristics of Wetlands in the West Garo Hill, Meghalaya,India”. *International Research Journal of Biological Sciences.* Vol 4(1), January (2015): 60-65.
4) Bhattacharyya K. G. and Kapil N. (2010), “Impact of urbanization on the quality of water in a natural reservoir: a case study with the DeeporBeel in Guwahati city, India”. *Water and Environment Journal*. ISSN 1747-6585.
5) BIS, (Bureau of Indian Standards) 1993. Specification for drinking water. Indian Standards (IS: 10500), New Delhi.
6) Bordoloi R. Abujam S.K.S. and Paswan G. (2012), “Limnological Study of a Closed Wetland – Potiasola from Jorhat district,Assam”. *J.Bio. Innov* (5), 2012: 132-141.
7) Deb S., Saikia J. and Kalamdhad A.S. (2019), “Ecology of DeeporBeel wetland, a Ramsar site of Guwahati, Assam with special reference to algal community”, *ejbps*, 2019, Volume 6, Issue 5, 232-243.
8) Islam M., Ahmed A.M., Barman B., Dakua S. and Debnath D. (2014), “Studies on physico-chemical properties of water in some selected sites of Deepor Beel (Ramsar site), Assam, India”.*The Clarion* Volume 3 Number 2 (2014) pp. 25-32.
9) Boyacioglu, H. (2006).Surface water quality assessment using factor analysis. Water SA. 32(3), 383-393.
10) Chakrabarty, S. and Sarma, H.P. (2011), “A statistical approach to multivariate analysis of drinking water quality in Kamrup district, Assam, India.” Archives of Applied Science Research. 3 (5), 258-264.
11) Aydin Uncumusaoglu A. (2018), “Statistical assessment of water quality parameters for pollution source identification in Bektaş Pond (Sinop, Turkey)” Global NEST Journal, Vol 20, No 1, pp 151-160

12) Ammar Tiri, Noureddine Lahbari and Abderrahmane Boudoukha (2015), “Assessment of the quality of water by hierarchical cluster and variance analyses of the Koudiat Medouar Watershed, East Algeria”, Appl Water Sci (2017) 7: 4197-4206

13) Innocent C. Nnorom, Ewuzie and Sunday O. Eze (2019), “Multivariate statistical approach and water quality assessment of natural springs and other drinking water sources in Southeastern Nigeria”, Heliyon