

Comprehensive Water Quality Indexing Using Functional Data and Principal Component Analysis for River Narmada

Swati Mishra



Department of Civil Engineering
National Institute of Technology Rourkela

Comprehensive Water Quality Indexing using Functional Data and Principal Component Analysis for River Narmada

Dissertation submitted in partial fulfillment

of the requirements of the degree of

Master of Technology

in

Civil Engineering

by

Swati Mishra

(Roll Number: 214CE4467)

based on research carried out

under the supervision of

Prof. Kanhu Charan Patra



May, 2016

Department of Civil Engineering
National Institute of Technology Rourkela

This page is left blank intentionally



Department of Civil Engineering
National Institute of Technology Rourkela

April 20, 2016

Certificate of Examination

Roll Number: 214CE4467

Name: *Swati Mishra*

Title of Dissertation: *Comprehensive Water Quality Indexing using Functional Data and Principal Component Analysis for River Narmada.*

We the below signed, after checking the dissertation mentioned above and the official record book (s) of the student, hereby state our approval of the dissertation submitted in partial fulfillment of the requirements of the degree of *Master of Technology in Civil Engineering* at *National Institute of Technology Rourkela*. We are satisfied with the volume, quality, correctness, and originality of the work.

Prof. Kanhu Charan Patra
Supervisor



Department of Civil Engineering
National Institute of Technology Rourkela

Prof. Kanhu Charan Patra
Professor

May 28, 2016

Supervisors' Certificate

This is to certify that the work presented in the dissertation entitled *Comprehensive Water Quality Indexing using Functional Data and Principal Component Analysis for River Narmada* submitted by *Swati Mishra*, Roll Number 214CE4467, is a record of original research carried out by him under our supervision and guidance in partial fulfillment of the requirements of the degree of *Master of Technology in Civil Engineering*. Neither this dissertation nor any part of it has been submitted earlier for any degree or diploma to any institute or university in India or abroad.

Kanhu Charan Patra
Professor

Dedication

This dissertation is dedicated to the researchers who work day and night for the well-being of the society...

Signature

Declaration of Originality

I, *Swati Mishra*, Roll Number *214CE4467* hereby declare that this dissertation entitled *Comprehensive Water Quality Indexing using Functional Data and Principal Component Analysis for River Narmada* presents my original work carried out as a post-graduate student of NIT Rourkela and, to the best of my knowledge, contains no material previously published or written by another person, nor any material presented by me for the award of any degree or diploma of NIT Rourkela or any other institution. Any contribution made to this research by others, with whom I have worked at NIT Rourkela or elsewhere, is explicitly acknowledged in the dissertation. Works of other authors cited in this dissertation have been duly acknowledged under the sections “Reference”. I have also submitted my original research records to the scrutiny committee for evaluation of my dissertation.

I am fully aware that in case of any non-compliance detected in future, the Senate of NIT Rourkela may withdraw the degree awarded to me on the basis of the present dissertation.

May 27, 2016
NIT Rourkela

Swati Mishra

Acknowledgment

“Continuous effort- not strength nor intelligence- is the key to unlocking our potential”, says Winston Churchill.

Getting a seat at National Institute of Rourkela was not easy and nor was sustaining it.

The journey of my Masters at the institute has been smooth only because it was geared by my supervisor Prof. K.C. Patra. Like a father, he constantly guided, scolded and corrected me. I would like to thank our External Examiner Prof. R.K. Panda, who, in spite of his busy schedule, stayed at Rourkela and authentically commented and corrected my mistakes. I express my heartfelt gratitude to the other faculty members of Civil Engineering Department who invested their precious time in our coursework and made me aware about the upcoming technologies in Water Resources Engineering.

This is perhaps the first time that I have stayed so far away and for such a long time away from home. My parents have continuously supported me, surprised me and reprimanded me for my wrong doings. I would like to thank my brother who has been a constant light of support at the institute.

My project would have been a dismantled piece of writing had Mrs. Mrunmayee Manjari Sahoo and Mr. Janaki Ballav Swain, both PhD Scholars at the institute, not directed me towards the research and constantly scrutinized my work. It is because of them that the work has been completed in time and has been worthy of the institute.

I thank all my friends who have been with me in this journey for making it an enjoyable one.

May 27, 2016
NIT Rourkela

Swati Mishra
Roll Number: 214CE4467

Abstract

The degrading water quality status of River Narmada can be deciphered by a tool called Water Quality Indexing. From the entire river basin seven important sites are chosen. Water Quality pertaining to twenty-five parameters are analyzed from 1991-2013. The final results are averaged to get the indices for a typical year. The current study commences with some preliminary methods. A significance test is conducted to determine three relevant parameters from each site. Based on these relevant parameters, Water Quality Indexing is done using Functional Data Analysis and Principal Component Analysis and finally error analysis of the methods is done. The preliminary analysis show that Biochemical Oxygen Demand, Dissolved Oxygen, Chemical Oxygen Demand and Total and Faecal Coliform are the most crucial water quality variables along the mainstream of Narmada River Basin.

Both Principal Component Analysis and Factor Analysis yield seven components from which it is concluded that agricultural runoff and industrial wastes account for an increase in alkalinity in the river water. It is also concluded that the deterioration in Dissolved Oxygen is being rejuvenated by variables causing alkalinity in water.

Conventional Water Quality Indexing done along the seven sites shows that for Station 1,2,5,6 and 7, water quality remains excellent almost throughout the year except for Stations 2 and 4. Functional Data Analysis and Principal Component Analysis also shows similar results. From the error analysis Functional Data Analysis using Entropy Weights is found to have least error.

Thus for a typical year the minimum indices implying the worst water quality are shown in a map which can help in deciphering the water quality status of the river well.

Keywords: Water Quality Indexing; Functional Data Analysis; Entropy Weights; Relative Pollution Degree; Student's t-test; Principal Component Analysis

Contents

	Certificate of Examination	ii
	Supervisors' Certificate	iii
	Dedication	iv
	Declaration of Originality	v
	Acknowledgment	vi
	Abstract	vii
	List of Figures	ix
	List of Tables	x
1	Introduction	1
1.1	General	1
1.2	Initiation of Water Quality Indices (WQIs)	2
1.3	More on Benefits of Water Quality Indices	3
1.4	Conventional Water Quality Indices Used in Indian Rivers	3
1.5	Functional Data Analysis	4
1.6	Multivariate Analysis	4
1.7	Problem Statement	5
1.8	Thesis Outline	5
2	Literature Review	6
2.1	Regarding Study Area	6
2.2	Regarding Functional Data Analysis	8
2.3	Regarding Multivariate Analysis	10
2.4	Regarding Water Quality Indexing	12
3	Study Area and Data Collection	14
3.1	General	14
3.1.1	The River Course and Physiography	15
3.1.1.1	Stations selected for study	16
3.1.2	The Climate and Rainfall	17
3.1.3	Water Quality of Mainstream Narmada River Basin	17
3.2	Data Collection and Analysis	18
3.2.1	Discharge (in cubic metres per second)	18
3.2.2	Temperature (in Degrees Celsius)	19
3.2.3	pH	20
3.2.4	Electrical Conductivity (in micro mho/cm)	21
3.2.5	Total Dissolved Solids (in micro mho/cm)	22
3.2.6	Ammonia Ion Concentration (in mg N/L)	23
3.2.7	Nitrate and Nitrite Ion Concentration in mg N/L	23

3.2.8	Silicon Dioxide (in mg/L)	24
3.2.9	Dissolved Oxygen (in mg/L)	25
3.2.10	Biochemical Oxygen Demand (in mg/L)	26
3.2.11	Chemical Oxygen Demand (in mg/L)	26
3.2.12	Phenol Ions Concentration (in mg/L)	27
3.2.13	Total Alkalinity (in mg CaCO ₃ /L)	28
3.2.14	Total Hardness (in mg CaCO ₃ /L)	29
3.2.15	Calcium Hardness (in mg CaCO ₃ /L)	30
3.2.16	Fluoride Ion Concentration (in mg/L)	30
3.2.17	Calcium Major Ion Concentration (in mg/L)	31
3.2.18	Magnesium Ion Concentration (in mg/L)	31
3.2.19	Sodium Ion Concentration (in mg/L)	32
3.2.20	Potassium Ion Concentration (in mg/L)	33
3.2.21	Chloride Ion Concentration (in mg/L)	33
3.2.22	Sulphate Ion Concentration (in mg/L)	34
3.2.23	Carbonate Ion Concentration (in mg/L)	35
3.2.24	Bicarbonate Ion Concentration (in mg/L)	35
3.2.25	Total Coliform Count (in MPN/100mL)	36
3.2.26	Faecal Coliform Count (in MPN/100mL)	37
4	Methodology	38
4.1	Water Quality Index	38
4.1.1	Parameter Selection	38
4.1.1.1	Student's t-test	39
4.1.2	Conversion of parameters of different units into a common scale	40
4.1.2.1	Development of Sub-Indices	41
4.1.2.2	Various Types of Sub-Indices	42
4.1.3	Assigning Weights to the Sub-Indices	42
4.1.3.1	Information Entropy Method	43
4.1.3.2	Relative Pollution Degree	44
4.1.4	Sub-Indices Aggregated to Form the Final Index	45
4.1.4.1	Linear Sum Index	45
4.1.4.2	Weighted Sum Index	45
4.2	Functional Data Analysis	46
4.3	Multivariate Statistics	47
4.3.1	Principal Component Analysis	47
4.3.2	Principal Factor Analysis	48
4.3.4	Cluster Analysis	49
4.4	Miscellaneous Methodologies	50
4.5	Permissible Water Quality Criteria	51
5	Results and Discussions	54
5.1	Preliminary Analysis	54
5.1.1	Descriptive Statistics	54

5.1.2	Principal Component and Principal Factor Analysis	55
5.1.2.1	Correlation Matrix	55
5.1.2.2	KMO-Sampling Adequacy	58
5.1.2.3	Total Variance Explained	58
5.1.2.4	Principal Component Matrix	59
5.1.2.5	Scree Plot	61
5.1.2.6	Rotated Factor Matrix	61
5.1.2.7	Cluster Analysis	63
5.2	Student's t-test	64
5.3	Conventional Water Quality Indexing	64
5.4	Water Quality Indexing Using Functional Data Analysis	65
5.4.1	Development of Sub-Index Curves	66
5.4.2	Assignment of Weights	70
5.4.2.1	Conventional Weights	70
5.4.2.2	Assignment of Weights by Information Entropy Method	70
5.4.2.3	Assignment of Weights by Relative Pollution Degree	74
5.4.2.4	Development of Water Quality Index Curves	78
5.5	Water Quality Index Based on Principal Component Analysis	83
5.6	Error Analysis	88
6	Conclusions	90
7	References	
	List of Figures	
	Figure 3.1 Location of Study Area	
	Figure 3.2 Variation of discharge for a typical year	19
	Figure 3.3 Variation of temperature for a typical year	20
	Figure 3.4 Variation of pH for a typical year	21
	Figure 3.5 Variation of Electrical Conductivity for a typical year	22
	Figure 3.6 Variation of Total Dissolved Solids for a typical year	22
	Figure 3.7 Variation of Ammonia for a typical year	23
	Figure 3.8 Variation of Nitrate and Nitrite for a typical year	24
	Figure 3.9 Variation of Silica for a typical year	25
	Figure 3.10 Variation of Dissolved Oxygen for a typical year	25
	Figure 3.11 Variation of BOD for a typical year	26
	Figure 3.12 Variation of COD for a typical year	27
	Figure 3.13 Variation of Phenol Ions for a typical year	28
	Figure 3.14 Variation of Total Alkalinity for a typical year	28
	Figure 3.15 Variation of Total Hardness for a typical year	29
	Figure 3.16 Variation of Calcium Hardness for a typical year	30
	Figure 3.17 Variation of Fluoride Ions for a typical year	30
	Figure 3.18 Variation of Calcium Major Ions for a typical year	31
	Figure 3.19 Variation of Magnesium Ions for a typical year	32
	Figure 3.20 Variation of Sodium Ions for a typical year	32

Figure 3.21 Variation of Potassium Ions for a typical year	33
Figure 3.22 Variation of Chloride Ions for a typical year	34
Figure 3.23 Variation of Sulphate Ions for a typical year	34
Figure 3.24 Variation of Carbonate Ions for a typical year	35
Figure 3.25 Variation of Bicarbonate Ions for a typical year	36
Figure 3.26 Variation of Total Coliform for a typical year	36
Figure 3.27 Variation of Faecal Coliform for a typical year	37
Figure 4.1 Typical t-distribution curve	40
Figure 4.2 Mechanism of Water Quality Indexing	41
Figure 5.1 A Scree Plot of Eigen Values versus Components	61
Figure 5.2 Dendrogram obtained using Hierarchical Clustering	63
Figure 5.3 Average Sub-Index Curves for Station 1	66
Figure 5.4 Average Sub-Index Curves for Station 2	67
Figure 5.5 Average Sub-Index Curves for Station 3	67
Figure 5.6 Average Sub-Index Curves for Station 4	68
Figure 5.7 Average Sub-Index Curves for Station 5	68
Figure 5.8 Average Sub-Index Curves for Station 6	69
Figure 5.9 Average Sub-Index Curves for Station 7	69
Figure 5.10 Entropy Weight Curves for Station 1	71
Figure 5.11 Entropy Weight Curves for Station 2	71
Figure 5.12 Entropy Weight Curves for Station 3	72
Figure 5.13 Entropy Weight Curves for Station 4	72
Figure 5.14 Entropy Weight Curves for Station 5	73
Figure 5.15 Entropy Weight Curves for Station 6	73
Figure 5.16 Entropy Weight Curves for Station 7	74
Figure 5.17 Relative Pollution Weight Curves for Station 1	75
Figure 5.18 Relative Pollution Weight Curves for Station 2	75
Figure 5.19 Relative Pollution Weight Curves for Station 3	76
Figure 5.20 Relative Pollution Weight Curves for Station 4	76
Figure 5.21 Relative Pollution Weight Curves for Station 5	77
Figure 5.22 Relative Pollution Weight Curves for Station 6	77
Figure 5.23 Relative Pollution Weight Curves for Station 7	78
Figure 5.24 Water Quality Index Curves for Station 1	79
Figure 5.25 Water Quality Index Curves for Station 2	79
Figure 5.26 Water Quality Index Curves for Station 3	80
Figure 5.27 Water Quality Index Curves for Station 4	81
Figure 5.28 Water Quality Index Curves for Station 5	81
Figure 5.29 Water Quality Index Curves for Station 6	82
Figure 5.30 Water Quality Index Curves for Station 7	83
Figure 5.31 Water Quality Index Curves for Station 1 using PCA	84
Figure 5.32 Water Quality Index Curves for Station 2 using PCA	84
Figure 5.33 Water Quality Index Curves for Station 3 using PCA	85
Figure 5.34 Water Quality Index Curves for Station 4 using	86

PCA	
Figure 5.35 Water Quality Index Curves for Station 5 using PCA	86
Figure 5.36 Water Quality Index Curves for Station 6 using PCA	87
Figure 5.37 Water Quality Index Curves for Station 7 using PCA	88
Figure 5.38 Minimum Water Quality Indices for River Narmada	89
List of Tables	
Table 3.1 Description of the stations studied	
Table 4.1 Permissible and Acceptable Limits of Water Quality Parameters	51
Table 5.1 Descriptive Statistics for the span 2009-2013	54
Table 5.2 Correlation Matrix	56
Table 5.3 KMO and Bartlett's Test	58
Table 5.4 Total Variance Explained	59
Table 5.5 Principal Component Matrix	60
Table 5.6 Rotated Factor Matrix	62
Table 5.7 p-values for Student's t-test	64
Table 5.8 Conventional Indices for a typical year	65
Table 5.9 Different Classes of water with their notations	65
Table 5.10 Stationary Weights of Significant Variables	70
Table 5.11 Error Analysis of the Water Quality Index Methods	88

CHAPTER I

INTRODUCTION

1.1 General

Water is an indispensable natural resource. It has played a crucial role in the triggering and sustenance of civilizations. However, till 1960s, extraction of water was more important than its quality assessment. The last three decades of the twentieth century have witnessed the importance of water quality in an alarming rate (Abbasi et.al. 2012).

The water quantity of a resource is defined by the mass of water contained in it. On the other hand, water quality has different constraints according to its use. The water required for household purpose varies from that invested in industries. Different animals require different levels of water quality; same goes with different crops. Thus, water quality depends on anything and everything that it might have gathered during its travel from clouds to the surface or water body; in any form, be it dissolved, colloidal or suspended. The tag of being a “Universal Solvent” expands the canopy of its miscibility.

The quality of water can be described by listing out the concentration of all the chemicals present in a sample. But, such an assessment would be too lengthy depending on the constituents present and would be correctly deciphered by technical experts only.

The question also arises of how to compare two samples having different chemical constituents. The danger of consuming a sample with five components having 5% more concentration than the permissible limit (and hence objectionable): pH, Electrical Conductivity, Hardness, Sodium and chloride may not be equivalent to a sample containing at least 1 MPN of Coliform Count.

Thus all the constituents present in a sample should be addressed differently but depicted under a single value. Water Quality Indices solve the vexing problem of quality assessment by following the same formula. They translate the constituents and their corresponding concentrations into a single value on the basis of a system.

The concept of using indices to represent a system is not a novel idea. Be it Dow Jones Index of New York or Sensex of stock exchange of Mumbai, indices have already reserved their place in the field of economics, trade and commerce. Even in ecology, indices have been thoroughly implemented through Shannon Index, Simpson Index and so on.

Thus, indices compositely represent a condition deduced by combining, in several ways, a number of relevant yet incommensurate measurements. This aggregation results in an ordinal number which assists in understanding and inferring the overall contribution of the measurements to the number.

Water Quality Index is an important component of environmental indices. It is used as communication tool by regulatory boards to describe the 'health' of any environment system; water, air, sediments or soil. It also helps in the evaluation of various regulatory policies on the different environmental practices.

1.2 Initiation of Water Quality Indices (WQIs)

The concept of water quality assessment had its first introduction in Germany in 1848. There, the water was considered fit or otherwise based on the presence or absence of some organisms. Since then European countries have used different classification systems for assessing water quality in their regions. There are two types of classification systems: one concerned with the

amount of pollutants present and the other concerned with the number of living communities (macroscopic or microscopic) of organisms present.

These classification systems grouped water bodies into different pollution classes. Unlike this, indices assign a numerical value to represent the condition of water quality of a water body.

The use of such a numerical scale begins with Horton's Index in 1965.

1.3 More on the benefits of Water Quality Indices

The use of indices are well-entrenched by the regulatory agencies which are responsible for checking the status of water pollution. After the development of Water Quality Indices, they are further implemented to examine trends, showcase the environmental conditions and evaluate the potential of the regulatory measures of the government. The other uses of water quality indices are resource allocation, R&D, ranking of allocations and environmental planning.

The water quality indices can also be classified into two categories; indices based on physico-chemical parameters and indices based on bioassessment. Usually water quality indices belong to the category based on physico-chemical parameters, including Horton's Index which includes coliforms as a variable. The present study is also classified under physico-chemical parameters.

1.4 Conventional Water Quality Indices Used in Indian Rivers

The first water quality index developed in India and perhaps in Asia too, is developed by Bhargava in 1983 and 1985. This method sub-divides the water quality variables into four groups namely, bacterial group, toxicants, physical and non-toxic parameters. A multiplicative

aggregation is used to combine the sub-indices obtained from different functions. This method has been applied to River Yamuna at New Delhi.

An index system, based on National Sanitary Foundation-Water Quality Index, is developed exclusively for River Ganga by Ved Prakash et.al from 2000 to 2010.

In 1995, Dhamija and Jain have used nine water quality parameters to summarize the water quality of Hanuman Lake at Jabalpur.

Sargaonkar and Deshpande in 2003, have developed “Overall Index of Pollution” to evaluate the surface water quality of Indian Rivers.

Swamee and Tyagi in 2000 and 2007 have improved the expressions of sub-indices and aggregation methods for conveying the water quality.

1.5 Functional Data Analysis

A branch of statistics and developed by Ramsay (1997), Functional Data Analysis, is a method which works on information represented by functions varying over a continuum, usually time. For this, the sampling data need not be identical and may have missing values. This highly flexible dynamic method has been thoroughly used in the fields of medicine, economy and ecology. A detailed account of this method is given in the further chapters.

1.6 Multivariate Analysis

This statistical method involves the observation and analysis where more than one variables occur as outcomes. The techniques used in the present study are Principal Component Analysis, Factor Analysis, Clustering, Discriminant Analysis and MANOVA. These methods have been explained later.

1.7 Problem Statement

The lifeline of Madhya Pradesh and one of the major rivers of India, Narmada, is being readily polluted by municipal wastes, anthropogenic sewage and industrial wastes. As a result, the water quality has deteriorated to Category 'B' i.e. 'Fit for Outdoor Bathing' along the river. Hence to create awareness among the people using its water is necessary. Water Quality Indexing is such a technique with which common people can understand the water quality status of a water body. Conventional and Modern Methods of Water Quality Indexing are used in the current research. The objectives of the research are as follows:

- i) Applying a Conventional WQI method to the river.
- ii) Apply WQI method based on Functional Data Analysis and Modern Statistical Techniques.
- iii) Compute the Final Indices along the Mainstream of the River after comparison.
- iv) Predict the Water Quality Using Clustering and MANOVA.

1.8 Thesis Outline

Chapter I acquaints one about the origin of Water Quality Indices and highlights about the methods to be used. The problem statement has been elaborated in this chapter

Chapter II reflects the work done by researchers in the study area and the methods used. It also brings out a critical review of the past works.

Chapter III narrates in detail about the study area, the water quality variables used and their variations in a typical year.

Chapter IV describes about the methodologies incorporated in the research.

Chapter V comprises of the results obtained from the research.

CHAPTER II

LITERATURE REVIEW

2.1 Regarding Study Area

Ahmad et.al (2001) state that the water quality of many rivers like the Ganges, the Sutlej, the Cauvery, Sabarmati, Krishna, Khan-Kshipra, Betwa and River Narmada are comparable. The pH in these rivers varies from 6 to 9. DO and BOD are also very poor.

Sharma et.al (2007) have studied the effect of domestic sewage and effluents from the paper mills on the water quality and ecology of River Narmada at Hoshangabad. Experimental studies are done for four stations namely, Post Office Ghat (SS1), Korighat sewage mixing point (SS2), Landianalla sewage mixing point (SS3), and Dongarwara (SS4). The water quality parameters studied are temperature, pH, nitrate, chloride, phosphate, Total Suspended Solids, DO, BOD, Coliform Count and phyto-plankton. The results are explained through the minimum, maximum and average values along with statistical evaluations like standard deviation, standard variance, standard error and 95% confidence limit. The study gives us informative data regarding the wastewater contamination and ecology of the river. In some water samples, the hydro-biological characteristics exceed the permissible limits. The results can be employed for future planning of water quality.

Telang et.al (2009) have analyzed the aftermath of mass bathing at Hoshangabad. They say that the water takes more time to rejuvenate through self-purification process. Experimental analysis is done according to APHA standards. Three ghats namely, a place near Bandhra ban, SethaniGhat and MangalwaraGhat are chosen. Water samples are

collected before, during and after sometime of the mass bathing ritual during January 2008. It is seen that the parameters considered such as, pH, DO, BOD, COD and TC are deteriorating on account of the organic pollution load. The authors have recommended awareness programmes for all the citizens who are associated with the river.

Malviya et.al (2010) have identified and characterized six sites, four at Hoshangabad, one at Handia and one at Nemawar. This is carried throughout the year. DongarwaraGhat (Site IV), at Hoshangabad is found to be more polluted than the rest. The various physico-chemical and biological parameters were determined as per methods suggested by APHA (1976). Temperature, pH and Dissolved oxygen were recorded immediately after collection of sample at the sites, while other parameters were analysed in the laboratory within 24 hours. The various sites are characterized by low COD values (at DongarwaraGhat), high BOD values (at Vivekananda Ghat (I), SethaniGhat (II), Handia Ghat (V) and NemawarGhat (VI). High values of alkalinity and chloride are found at Dongarwara and Handia.

Sharma et.al (2011) have carried out statistical analysis on parameters like pH, EC, Turbidity, Calcium Hardness, Nitrite, Sulphate, Chloride, phosphate and Dissolve Oxygen. Local anthropogenic factors, industrial effluents and agricultural are held responsible for the decreasing water quality.

Soni et.al (2013) have conducted pre-impoundmental studies on water quality of the Western part of River Narmada. The study is carried out at Omkareshwar (U/S), Omkareshwar (D/S), Maheshwar Dam, Mandleshwar and Koteshwar. The physico-chemical parameters are analyzed using APHA standards and water quality indexing is

done using Brown's Method. The WQI ranges from 69.10(very good) to 124.03 (unfit). Organic pollution and anthropogenic activities are held responsible for the deterioration. Deshkar et.al (2014) have studied the temporal changes in the fluxes of the macronutrients in the Narmada estuary from 2005-10. Temperature, salinity, nitrates and phosphates are analyzed. One-way ANOVA is done and it is found that these nutrients change over a period of time. Effluent flow, Sewage drainage, catchment runoff and tides have greater influence on the nutrients.

2.2 Regarding Functional Data Analysis

Champely and Doledec (1997) have used FDA to separate long term trends from periodic variation of pollutants in the Seine River. They have sampled 17 parameters irregularly from 1983-1991 on 212 occasions and have incorporated 9% missing values. To overcome certain difficulties in modelling and multivariate analysis, they propose a statistical approach based on data modelling by non-parametric smoother Loess and application of Functional Principal Component Analysis (FPCA). They smooth 17 curves using Periodical Trend Loess (PTL) and later use FPCA to determine the structure of long term variation in data. Results show that FPCA processed on non-parametric models helps to solve some of the problems associated with missing data and irregular sampling.

Bjornstad (1998) maps about the regional transition to cyclicity in a vole scientifically named as *Clethrionomys rufocanus* in Hokkaido, Japan. Data is collected from 1962 to 1992 in terms of time series of different lengths. Using FDA, it is found that the population of the vole is relatively stable in the west and south-west regions and is cyclic in north-east and east.

Ramsay et.al (2002) have analyzed the dynamics of a monthly production index adjusted non-seasonally. They say that a third-order Ordinary Differential Equation describes the data and the evolutionary dynamics of the 70 year span series is accordingly determined. This study thus illustrates the flexibility of using functional data analysis in the dynamics of production index.

Fabrice et.al (2005) have extended the study of classical Multilayer Perceptron to Functional Data Analysis. MLP is a feedforward ANN model and maps input data sets into approximate output data. It consists of multilayer nodes fully connected with the adjacent layers in a directed graph. Consistent results are obtained using Functional Data Analysis and hence prove the approach to be statistically well defined.

Henderson (2006) has explored the differences in nutrient and sediment trends in three dams in South East Queensland using FDA. Functional data analysis is briefly reviewed, concentrating in particular on the techniques of functional principal component analysis, functional linear modelling and functional cluster analysis.

Muniz et.al (2012) have expressed water quality monitoring data of San Esteban estuary in Spain, as curves or time-dependent functions instead of discrete values. The outliers in water quality monitoring samples are detected. Results are defined in terms of the origin of the outliers and their causes. A comparison is done between the functional approach and the vector approach.

Iglesias et.al (2015) have studied the variability of the Ebro River (Spain) water quality through a global quality index (GQI) using two methods: functional data analysis (FDA) and Stewhart-type control charts for statistical process control (SPC). The data was collected in 2008 at the El Bocal station, which is a strategic location. Temperature,

ammonium content, nitrate content, conductivity, dissolved oxygen, pH, and turbidity were measured every 15 min.

Yan et.al (2015) have developed a dynamic water quality index based on Functional Data Analysis for River Changjiang at Sanjianying for 2012 and find that the water quality index curves derived from functional data analysis are more generalized and flexible as compared to those obtained using conventional indexing. The weights for the indices are determined using importance and relative pollution degree of each of the three water quality variables. The current research work is principally based on this paper.

2.3 Regarding Multivariate Analysis

Noori et.al (2010) have determined non-principal and principal monitoring stations in River Karoon using Principal Component Analysis. Samples from 17 stations are collected from 1999-2002. Four insignificant stations are removed and all the water quality variables namely, Biochemical Oxygen Demand, Chemical Oxygen Demand, Electrical Conductivity, Nitrate Ions concentration, Sulphate Ions concentration, temperature, Chloride Ions concentration, Dissolved Oxygen, hardness, Total Dissolved Solids, pH, and turbidity, are deemed important.

Najar et.al (2011) have conducted cluster analysis, principal component analysis and factor analysis on Anchar Lake, Khushalsar Lake and Dal Lake to evaluate the water quality of thirteen parameters. The physicochemical parameters vary considerably within the sites. The hierarchical cluster analysis divides the variables into heavily, moderately and less polluted groups. Domestic and agricultural runoff and catchment geology are held responsible for the water quality on the basis of factor analysis.

Wang et.al (2012) have used cluster analysis and principal component and principal factor analysis to evaluate the 15 water quality variables for six sites at Songhua River Harbin region from 2005-2009. Cluster analysis groups the variables into high, moderate and low pollution levels. Principal Component Analysis is conducted on the three groups and five latent factors are deduced. The causes for water quality degradation are found to be organic sources, agricultural runoff, industrial effluents and natural sources.

Selle et.al (2013) have used principal components analysis to understand the processes governing the water quality of surface, spring and groundwater from Ammer catchment at South-western Germany. Solutes from land use and geology affect the water quality in surface water and ground water. Principal Component Analysis Scores are found to interpret the results more accurately.

Gomes et.al (2014) have optimized the water quality monitoring of Leca River, Portugal, with principal component analysis at seven sites for five campaigns in 2006. These methods help in determining minimum number of stations and thus support management decisions.

Bonansera et.al (2015) have employed principal component analysis and cluster analysis to evaluate the water quality data sets in Rio Tercero Reservoir at Argentina. Six stations are chosen for assessing 21 water quality parameters from 2003-2010. Four principal components and three clusters are extracted accounting for mineralization and other pollutants.

Kim et.al (2016) have applied multivariate analysis to Chilika Lagoon to determine the changes in water quality with respect to space and time post construction of the new channel. Thirty sites are analyzed on monthly basis from 1999-2009 and it is found that

salinity is an important factor for water quality. The nitrate to phosphate ratio is found to be increasing steadily. Thus multivariate analysis provides a clear understanding of the changing patterns with respect to space and time.

2.4 Regarding Water Quality Indexing

Brown et.al (1970) have finalized 11 out of 35 water quality variables through a panel of 142 persons. The variables are ranked and rating curves are developed for each variable. Weights are calculated using comparative analysis and by multiplicative aggregation formula, final indices are developed. This method is also called National Sanitary Foundation Index.

Akkaraboyina et.al (2012) have assessed the water quality indices of Godavari River Basin at Rajahmundry using eight water quality variables from 2009-2012. They have also predicted the water quality for 2012-2015 and have found the results to be adequate according to NSF indexing.

In the present study, this method has been used as conventional indexing because of its robustness in Indian Rivers.

Canadian Council of Ministers of the Environment Water Quality Index (CCME) in 2001, proposed a model to summarize the complexities of water quality data. Three elements are incorporated here: Scope (F1) which is the number of water quality variables not meeting the criteria; Frequency (F2) which is how many times the criteria have not been fulfilled; and Amplitude (F3) which is the extent to which the criteria have not been fulfilled. Terrado et.al (2010) say that CCME-WQI is most suitable for continuous monitoring.

Qian et.al (2007) have combined different multivariate methods to compute the water quality index of Florida River Lagoon in south Indiana. They have used thirteen water

quality variables and have log transformed the data set. Hierarchical Clustering, Principal Component Analysis, Exploratory Factor Analysis and Trend Analysis have been comprehensively used. The indices developed are found satisfactory.

The present study incorporates the indices developed using Principal Component Analysis used by Qian et.al (2007).

Jha et.al (2008) used entropy to evaluate the water quality of seven river systems in India namely Baitarni, Malaprabha, Brahmani, Pachin, Yamuna and Gomti using Dissolved Oxygen and Biochemical Oxygen Demand. They say that entropy can be satisfactorily used to evaluate the water quality.

Liu et.al (2009) have used fuzzy theory and information entropy method for assessment of water quality of three gorges of China. Here, exponential membership function has been used and information entropy method has been employed to modify the weights to obtain necessary information. They say that this comprehensive method gives better results than the traditional method of indexing.

Sahoo (2014) studied about the water quality of five gauging stations at Brahmani River for the span 2003-2012. Various multivariate techniques along with fuzzy methods for water quality prediction are done and Principal Component Analysis is found to give satisfactory results.

CHAPTER III

STUDY AREA AND DATA COLLECTION

3.1 General

The Narmada River Basin stretches out over the states of Madhya Pradesh, Chhattisgarh, Maharashtra and Gujarat. This major river system heralds a territory of 98,796 square kilometres which is equivalent to 3 percent of the aggregate geographical expanse of the country. The maximum length of the river is 923 kilometres and its maximum width is 161 kilometers. It expands over 72°32' E to 81°43'E longitudes and 21°27'N to 23°37'N latitudes. The Vindhya Ranges confine the river from the north, in the south it is limited by Satpura Ranges and the Maikala Ranges bound it from the north. The location of the river demarcates between the North and South India. This peninsular river flows towards the west, covering a distance of 1,312 kilometres and finally drains itself into the Arabian Sea via Gulf of Khambat, near the city of Bharuch at Gujarat. Figures 3.1 and 3.2 illustrate the location of River Narmada Basin and the various hydro-meteorological stations along River Narmada and its course with some of its tributaries.

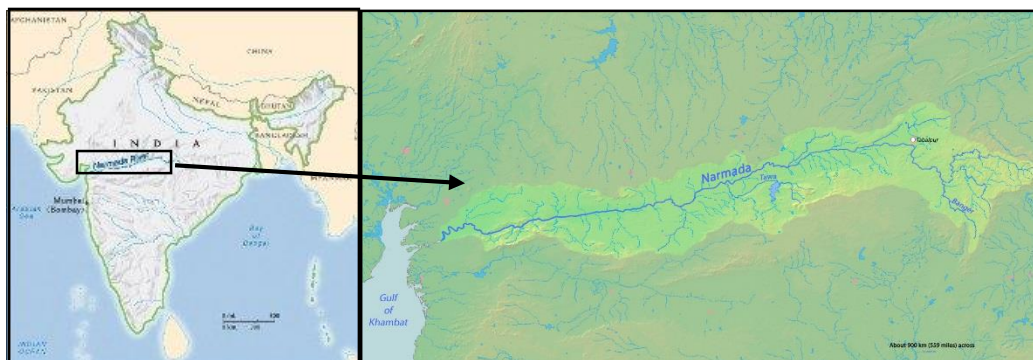


Fig 3.1: Map showing the location of the study area along with its general course and some tributaries

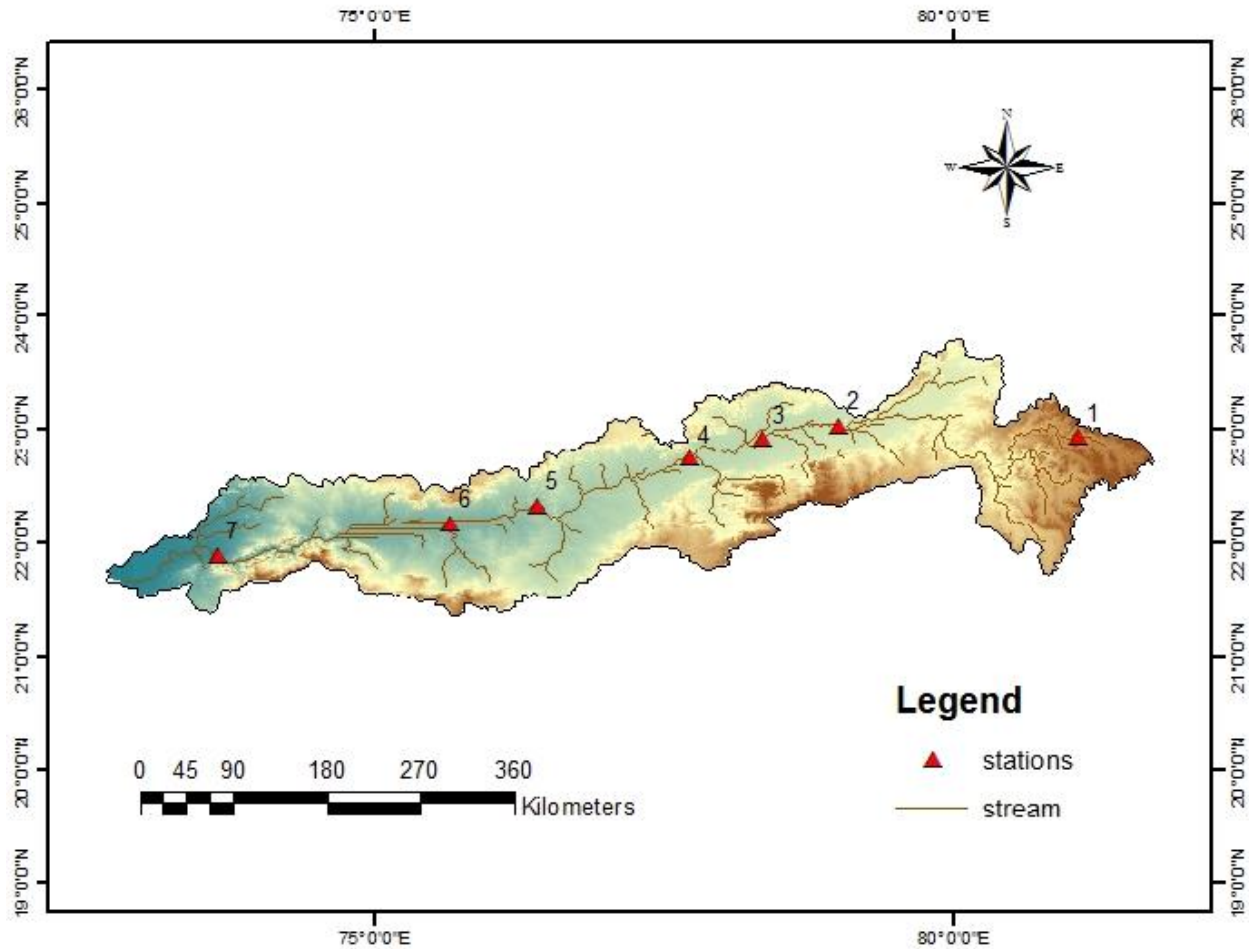


Fig 3.2: Map showing the Stations selected for the analysis

3.1.1 The River Course and Physiography

The river originates from Maikala Range near the pilgrim town of Amarkantak in Anupur district, Madhya Pradesh with an elevation of approximately, 1057 metres. Along the mainstream, for the initial 1079 kilometres, it covers Madhya Pradesh. Later on, it constitutes a boundary between Maharashtra and Madhya Pradesh for 39 kilometres and Gujarat and Maharashtra for 35 kilometres. The state of Gujarat withholds 39 kilometres of the river. The different tributaries of the river system are the Banjar, the Burhner, the Shakkar, the Sher, the Dudhi, the Ganjal, the Tawa, the Goi, the Kundi and the Karjan. These tributaries connect with the river from the left.

The tributaries of Tendoni, Hiran, Kolar, Barna, Uri, Man, Orsang and Hatni connect with the river from the right.

The Narmada River Basin can be split up into plain and hilly areas. The hilly areas are covered with forests and cover the upper and lower middle expanses. The central area between the upper and lower reaches constitute of pain regions which are suitable for cultivation.

3.1.1.1 Stations selected for the study

In the present study, seven stations have been chosen along the mainstream of the river. The sites are chosen according to the availability of data in a government portal.

The different stations chosen are illustrated in the Table 3.1.

Table 3.1: Description of the Stations Studied

Sl. No.	Name	Location	Resources/ Importance	Population	Distance
Station 1	Dindori	22°50'44"N latitude 81°4'31.8"E longitude	A hub of mineral resources like coal, bauxite, white ash, okars, high iron laterite and limestone	704,218 (according to 2011 census)	89.4 km from Amarkantak
Station 2	Barmanghat	23°1'58.8"N latitude, 79°1'14.52"E longitude	It is a holy place where River Warahi and Narmada unite. A fair is held each year during January.	7000 (according to 2008 census)	Located 211.4 km from Station 1.
Station 3	Sandhia	22°1'0.012"N latitude and 73°46' 0.012" E longitude	It is a medium sized village	767 (2011 census)	Located 551.3 km away from Station 2.
Station 4	Hoshangabad	22°44'38.76"N latitude, 77°44'13.2"E longitude	A very significant station. Of religious importance.	117956 (2011)	Located 416.1 km away from Station 3.

			Ritual of mass bathing during mid-January.		
Station 5	Handia	25°21'41.76"N latitude, 82°11'21.48"E longitude	This is a large village with many temples in the name of River Narmada	4221 (2011)	Located 537.6 km away from Station 4.
Station 6	Mandleshwar	22°10'36.84"N latitude, 75°39'36.36"E longitude	It is an important town with a hydro-electric project near Maheshwar Dam. Agriculture is an important occupation here.	11345 (2001)	Located 752.7 km away from Station 5.
Station 7	Garudeshwar	21°53'32.64"N latitude and 73°39'7.56"E longitude	An important holy place and a small village.	2452 (2011)	Located 209.4 km away from Station 6.

3.1.2 The Climate and Rainfall

The overall climate of the river basin remains humid tropical. In the east it is sub-humid and in the west it is semi-arid. In the hilly areas, humid climate is prevalent.

The Tropic of Cancer divides the basin into a small northern zone and a large southern zone. The seasons of cold-weather (temperature 17.5 to 20°C), hot weather (temperature 30 to 32.5°C), south-west monsoon (temperature 27.5 to 30°C), and post-monsoon (temperature 25 to 27.5°C) are distinct in the basin.

The normal annual rainfall in the basin is 1178 centimetres with the predominant rainy season being the south-west monsoon from June to October. The upper hilly as well as plain expanses of the basin receive maximum rainfall.

3.1.3 Water Quality of mainstream Narmada River Basin

The percentage of utilization of surface water from River Narmada is 23 which is not very high as the river is inaccessible at most of the places. The river has attained global priority for conservation of aquatic biodiversity since 1988 by Jetkins and Groombridge. The water level has also degraded as per a report in 2014. More than fifty species of fishes have been made extinct because of the degrading water quality and anthropogenic activities.

According to the Water Quality Status Report of 2012, by Central Pollution Control Board, the desirable limits are being exceeded along the main stream of the river at Mandla, Hoshangabad, and Saraswati Ghat in Madhya Pradesh and at Bharuch and Panetha in the state of Gujarat. The lowest DO level measured at Narmada is 2.2 mg/L while the Coliform Count and Faecal Coliform count reach to 9000 and 7000 MPN/100mL respectively.

3.2 Data Collection and Analysis

Monthly water quality data are collected from relevant government sites from the year 1991 to 2013 for the seven chosen stations. Following are the twenty six parameters chosen for the analysis. The variation of each estimate for a typical year has been illustrated with the help of graphs.

3.2.1 Discharge (in cubic metres per second):

The water quality estimates measured in a river system account more for the river chemistry than the temporal dynamics. Distinct seasonal changes in such a system result in prominent variations in the river discharge. The concept of yearly loading of nutrients

into the river system depends on how the river chemistry transforms according to the temporal changes in discharge. Though River Narmada is rain fed, it gets a moderately heavy discharge due to the fairly heavy yearly average rainfall, especially in the upper catchment expanse. Figure 3.3 shows the variation of discharge across all the stations for a typical year.

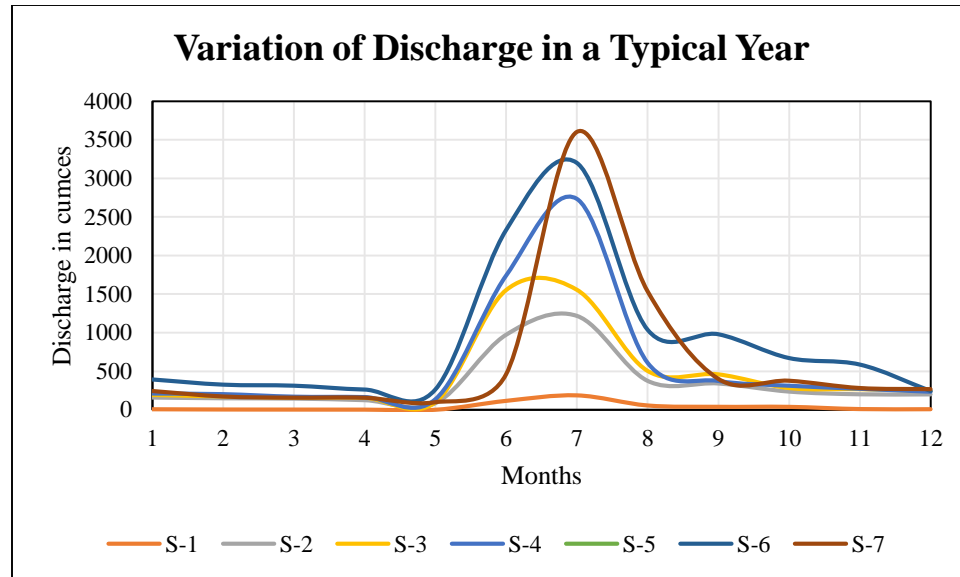


Fig 3.3: Variation of Discharge for all the sites in a typical year

3.2.2 Temperature (in Degrees Celsius):

This is the most common physical water quality variable which affects the bio-chemical characteristics and the life system in water. The concentration of Dissolved Oxygen and various biological activities depend on what the temperature is. Figure 3.4 shows the variation of temperature across all the stations for a typical year.

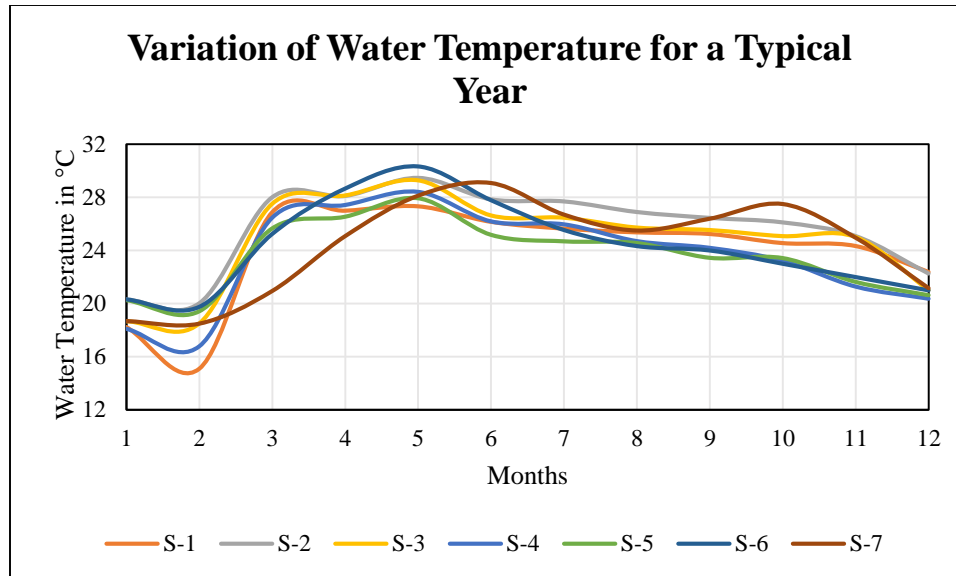


Fig 3.4: Variation of Temperature for all the sites in a typical year

3.2.3 pH

pH is another water quality estimate influenced by chemicals which in turn influences the acidity or the basicity of water. Bicarbonates of different metallic ions are responsible for alkalinity while mineral acids account for the acidity of water. Figure 3.5 shows the variation of pH across all the stations for a typical year.

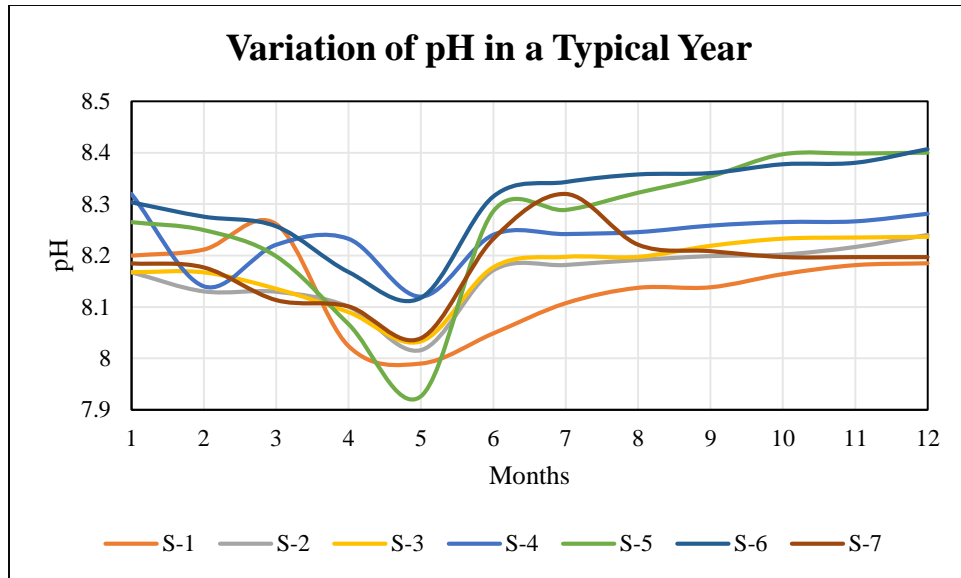


Fig 3.5: Variation of pH for all the sites in a typical year

3.2.4 Electrical Conductivity in micro mho/cm

It is the measure of the capacity of water to allow electrical flow through it. Thus it accounts for the number of ions present in the water body. It is also an approximate estimate of Total Dissolved Solids. Metallic ions are responsible for an increase in Electrical Conductivity. Figure 3.6 shows the variation of discharge across all the stations for a typical year.

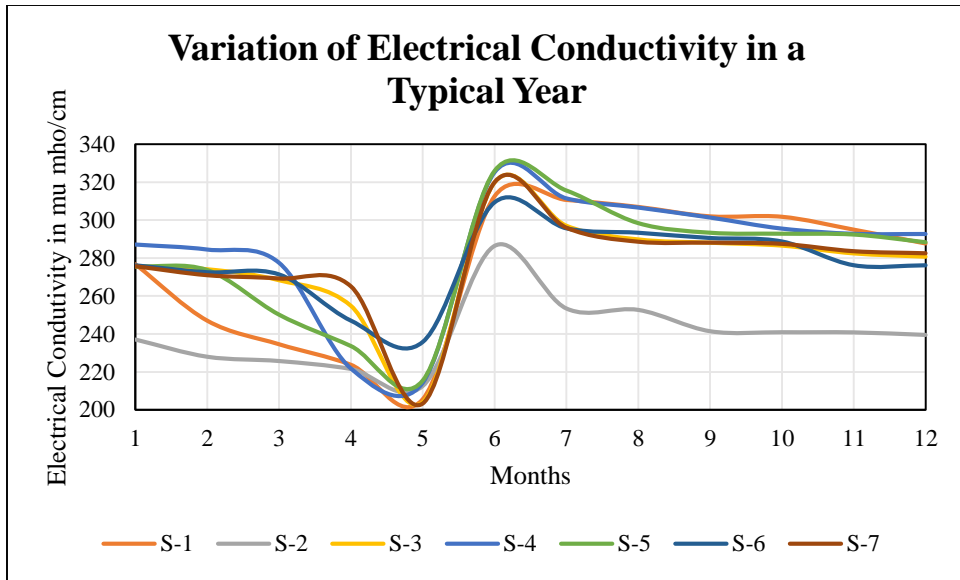


Fig 3.6: Variation of Electrical Conductivity for all the sites in a typical year

3.2.5 Total Dissolved Solids in micro mho/cm

This parameter constitutes of all the ions, both organic and inorganic, which are less than 2 microns. Figure 3.7 shows the variation of Total Dissolved Solids across all the stations for a typical year.

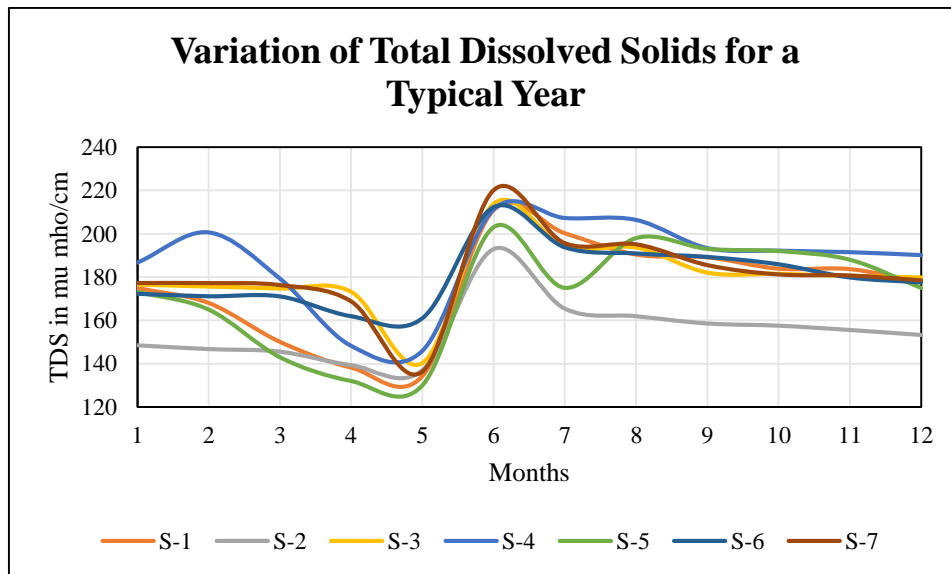


Fig 3.7: Variation of Total Dissolved Solids for all the sites in a typical year

3.2.6 Ammonia Ion Concentration in mg N /L:

This is a chemical water quality variable that reduces dissolved oxygen and harmful for aquatic life even when its concentration reaches 0.06 mg/L. The inflow of sewage results in accumulation of the parameter. Figure 3.8 shows the variation of Ammonia Ion Concentration across all the stations for a typical year.

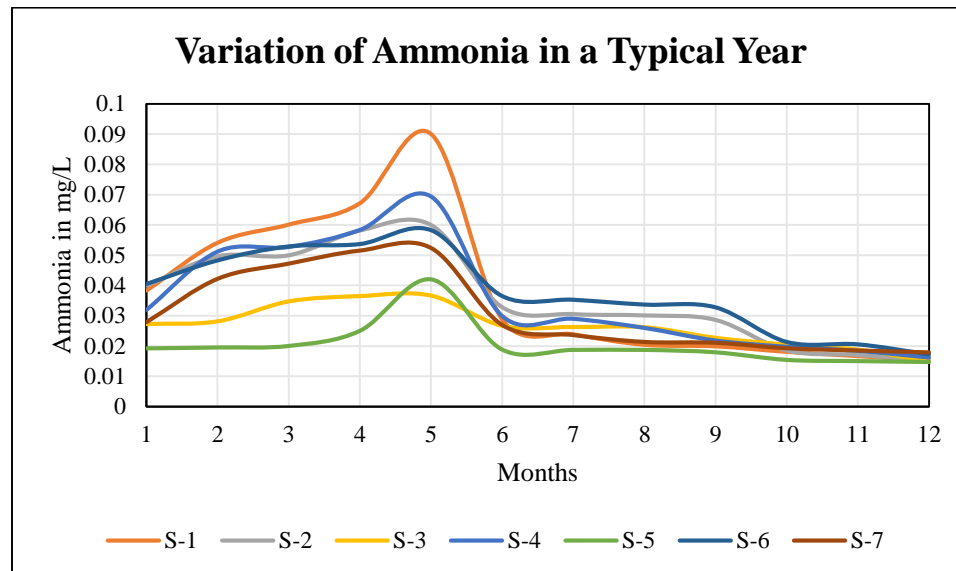


Fig 3.8: Variation of Ammonia Ion concentration for all the sites in a typical year

3.2.7 Nitrate and Nitrite Ions Concentration in mg N/L:

Nitrate ion concentration accounts for agricultural waste, waste water, human and animal wastes. Nitrite can be produced by Nitrosomonas Bacteria in stagnated water conditions with limited oxygen supply. Figure 3.9 shows the variation of Nitrate and Nitrite Ions Concentration across all the stations for a typical year.

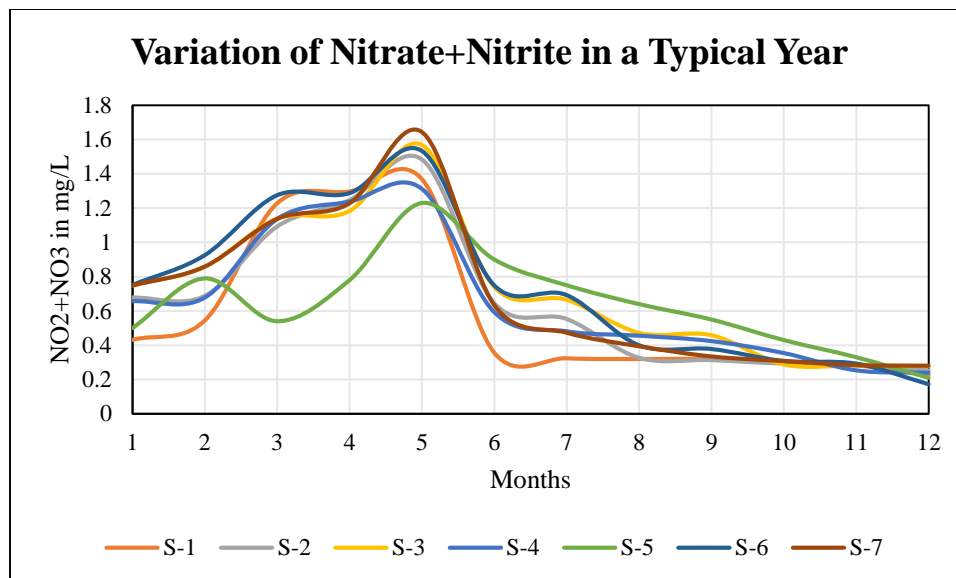


Fig 3.9: Variation of Nitrate and Nitrite Ion concentration for all the sites in a typical year

3.2.8 Silicon Dioxide in mg/L:

Disintegration of minerals and rocks account for the presence of silica in water. This estimate is an important growth component for microscopic plants like diatoms and also serve as a source of food for fishes. Though termed as corrosion inhibitors, if the concentration of silica exceeds certain limit, it makes the water unhealthy for consumption. Figure 3.10 shows the variation of Silicon Dioxide across all the stations for a typical year.

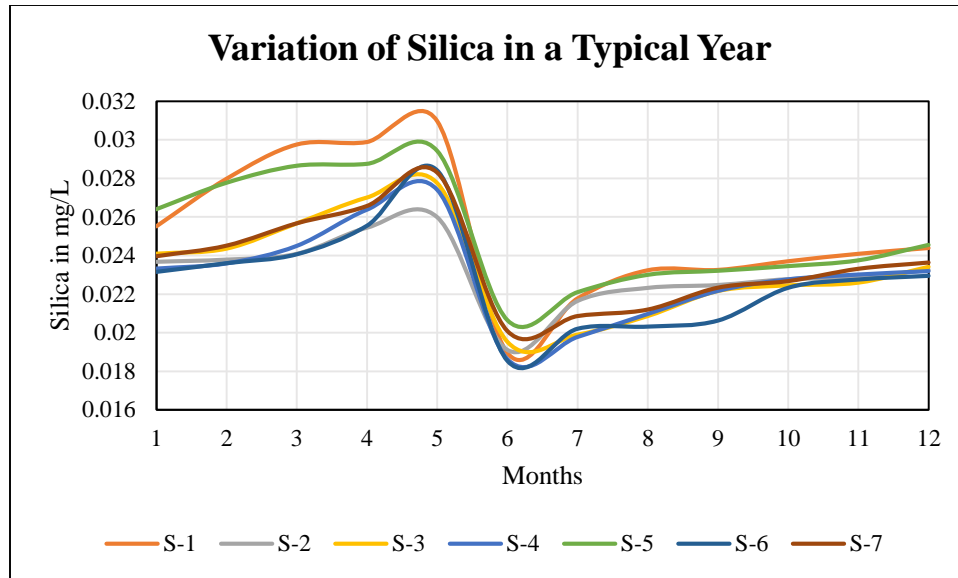


Fig 3.10: Variation of Silica concentration for all the sites in a typical year

3.2.9 Dissolved Oxygen in mg/L:

This is an important chemical parameter in water quality analysis. A certain amount of oxygen gets dissolved in a water body, is acclaimed as Dissolved Oxygen and accounts for the life within water. This parameter is sensitive to an increase in organic or inorganic pollution. Figure 3.11 shows the variation of Dissolved Oxygen across all the stations for a typical year.

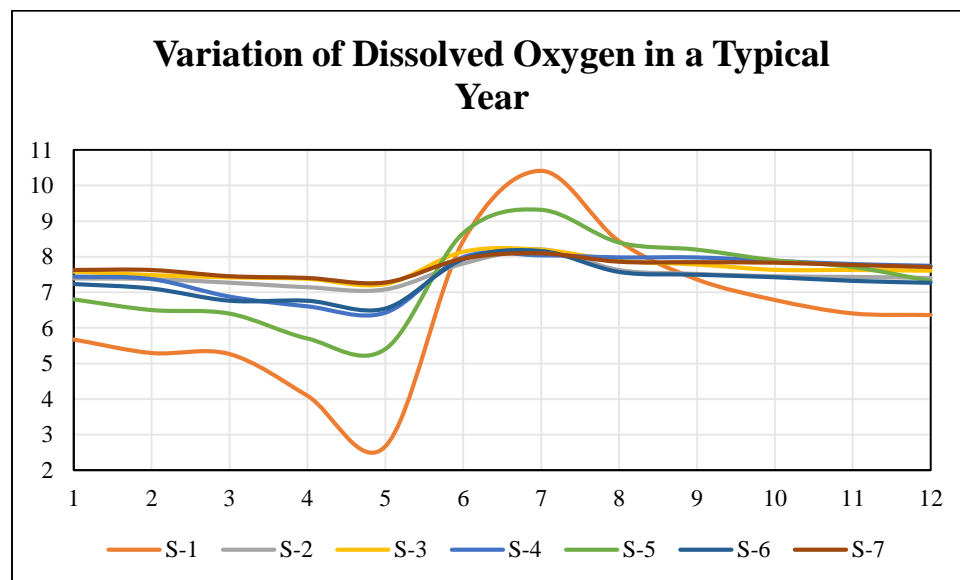


Fig 3.11: Variation of Dissolved Oxygen concentration for all the sites in a typical year

3.2.10 Biochemical Oxygen Demand in mg/L:

This parameter accounts for the amount of pollutants and organisms present in water. It is the quantity of DO consumed by microbes to catabolize organic matter in a water body. Hence an increase in BOD is harmful for the aquatic ecosystem. Figure 3.12 shows the variation of Biochemical Oxygen Demand across all the stations for a typical year.

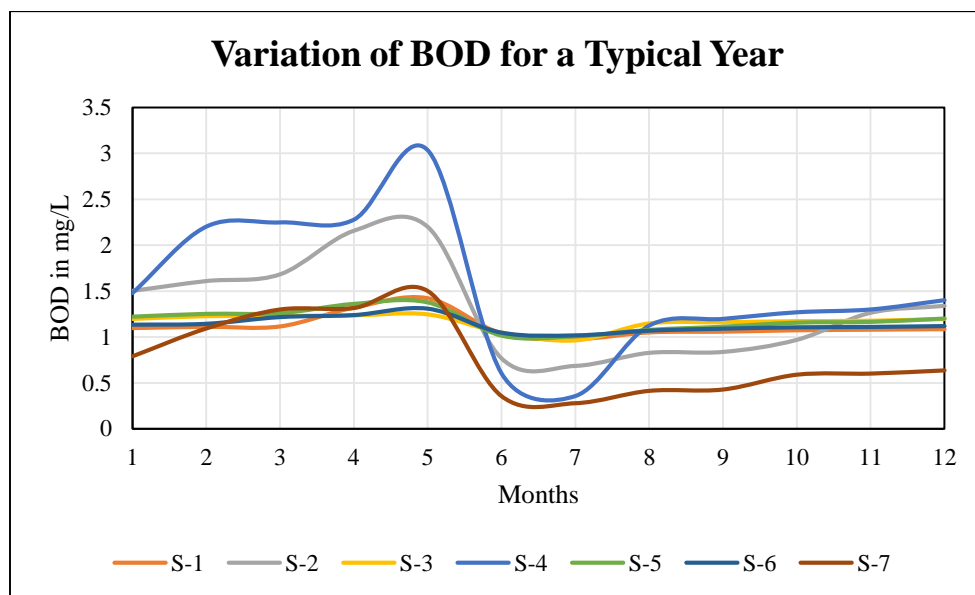


Fig 3.12: Variation of Biochemical Oxygen Demand for all the sites in a typical year

3.2.11 Chemical Oxygen Demand in mg/L:

This chemical estimate accounts for the amount of oxidizable organics present in a water body due to discharge of wastewaters. An increase in COD decreases the amount of DO present and hence affects the aquatic life. Figure 3.13 shows the variation of Chemical Oxygen Demand across all the stations for a typical year.

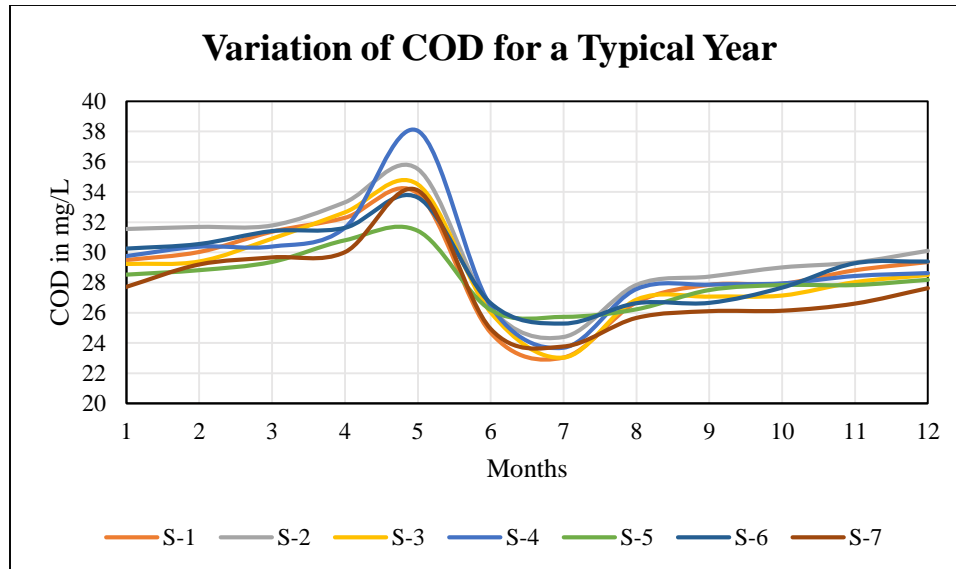


Fig 3.13: Variation of Chemical Oxygen Demand for all the sites in a typical year

3.2.12 Phenol Ions Concentration in mg/L:

Initially employed as a disinfectant, this organic pollutant accounts for toxicity when available above its permissible limits. This variable enters into a water body because of incessant disposal of effluents from various industries. Figure 3.14 shows the variation of Phenol Ions Concentration across all the stations for a typical year.

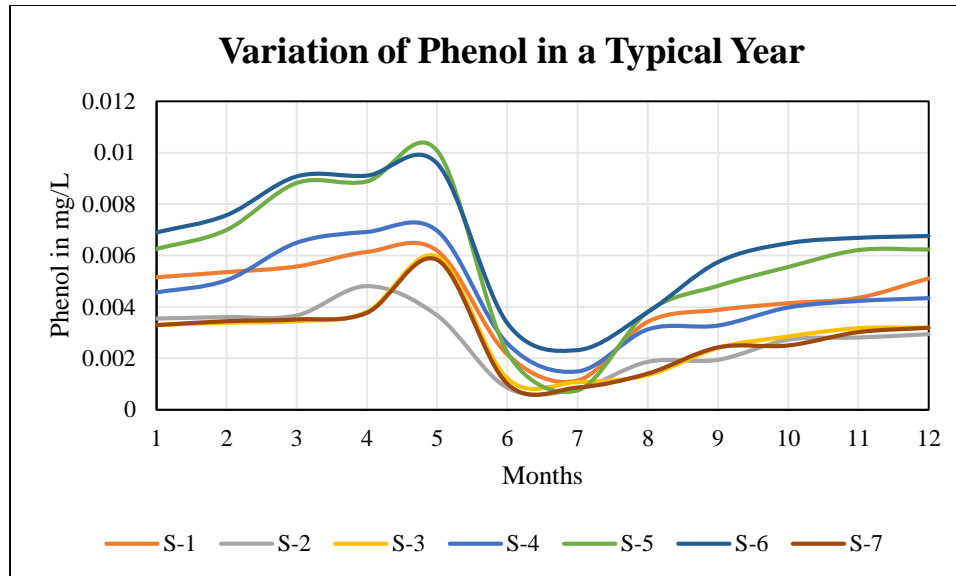


Fig 3.14: Variation of Phenol Ion Concentration for all the sites in a typical year

3.2.13 Total Alkalinity in mg CaCO₃ /L:

This parameter accounts for the resistance of water body to change its pH and is identified by the concentration of carbonates, hydroxides and bicarbonates. Figure 3.15 shows the variation of Total Alkalinity across all the stations for a typical year.

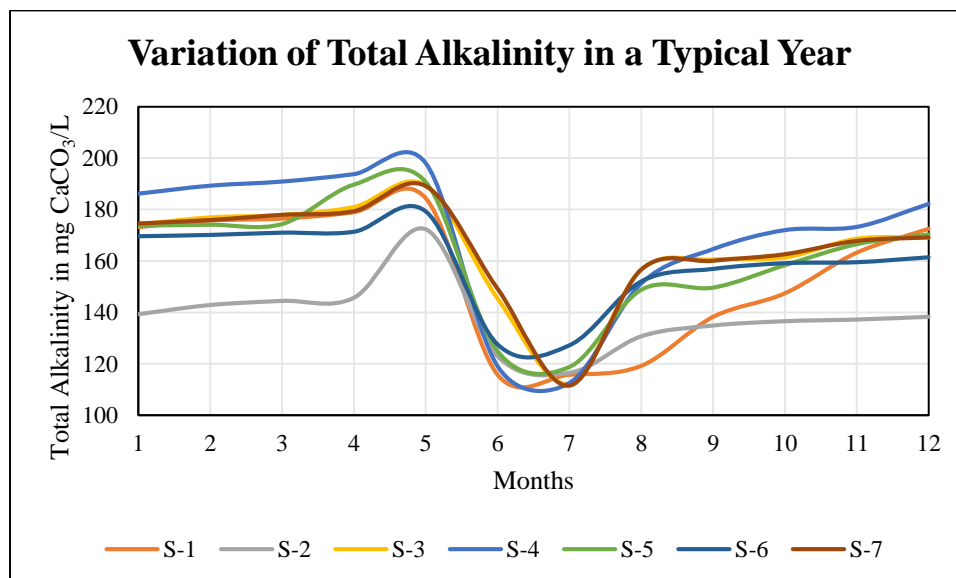


Fig 3.15: Variation of Total Alkalinity for all the sites in a typical year

3.2.14 Total Hardness in mg CaCO₃ /L:

This estimate is reflected by the concentration of Calcium and Magnesium Ions present. Hard water is difficult to use and hence should be properly treated. Figure 3.16 shows the variation of Total Hardness across all the stations for a typical year.

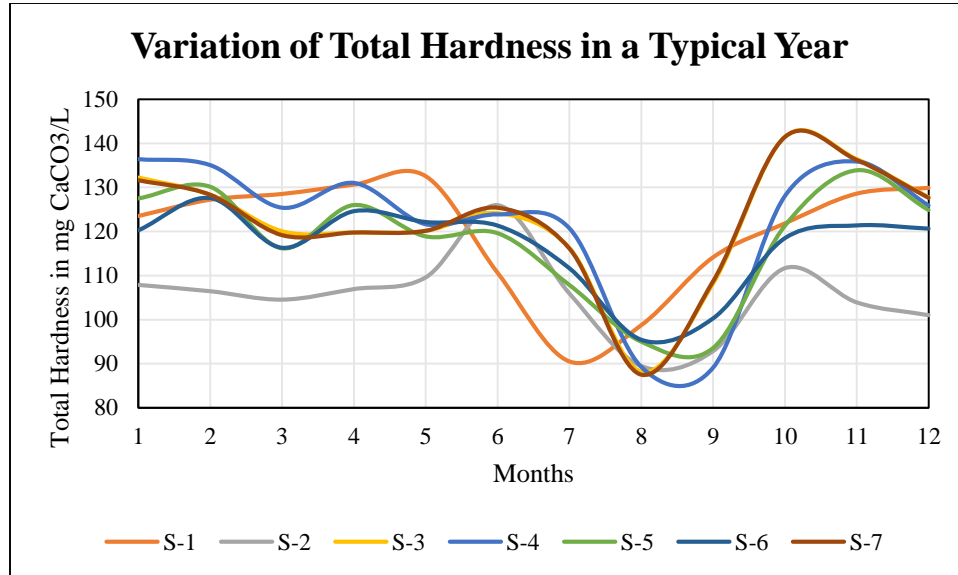


Fig 3.16: Variation of Total Hardness for all the sites in a typical year

3.2.15 Calcium Hardness in mg CaCO₃ /L:

This hardness accounts only for the calcium compounds. Figure 3.17 shows the variation of Calcium Hardness across all the stations for a typical year.

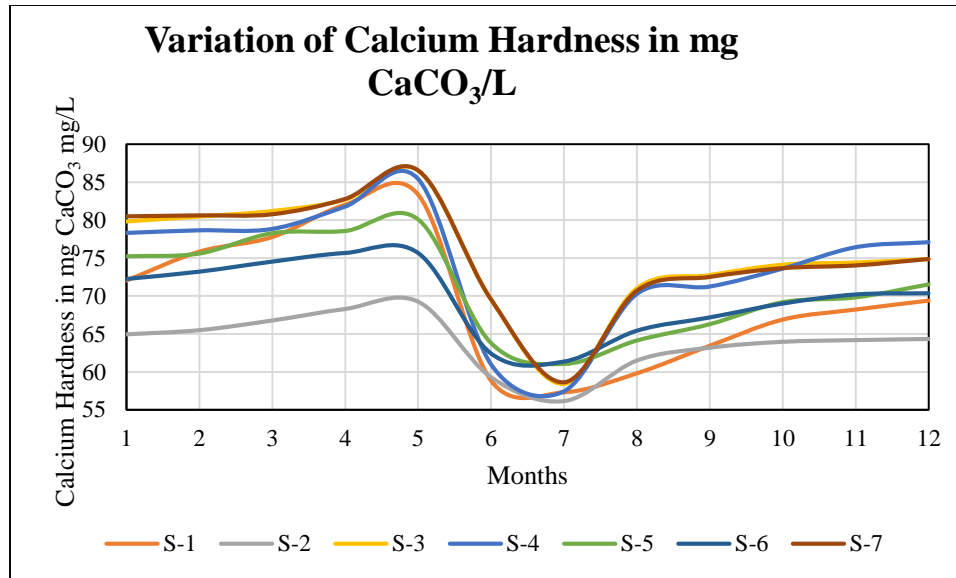


Fig 3.17: Variation of Calcium Hardness for all the sites in a typical year

3.2.16 Fluoride Ion Concentration in mg/L:

This variable when between 1.0 mg/L and 1.5 mg/L is good. If less than the permissible range, it causes tooth decay and if more causes skeletal and dental fluorosis. The main source of this pollutant are pesticides and pharmaceuticals. Figure 3.18 shows the variation of Fluoride Ions Concentration across all the stations for a typical year.

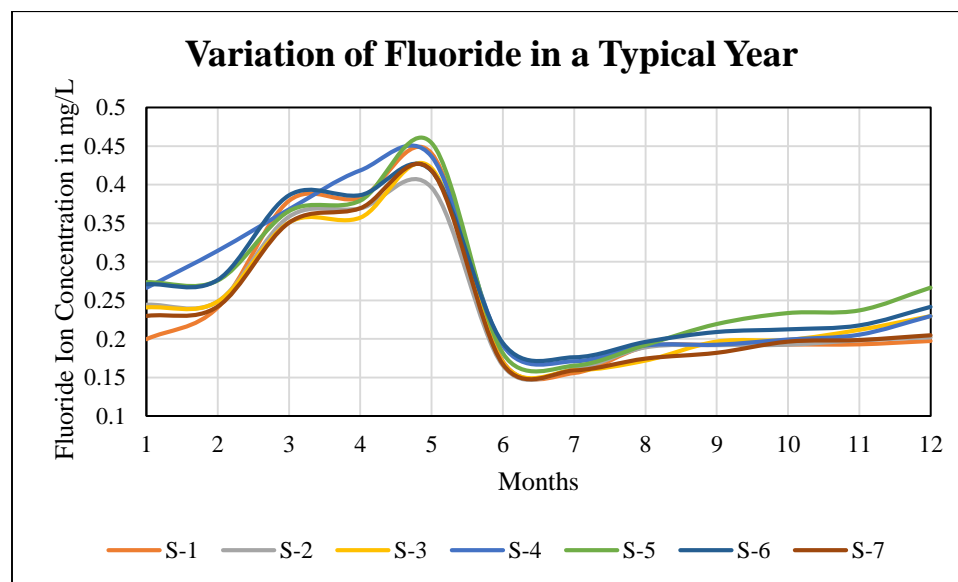


Fig 3.18: Variation of Fluoride Ion Concentration for all the sites in a typical year

3.2.17 Calcium Major Ion Concentration in mg/L:

Originating from rocks like that of limestone and marble, this variable is a measure of the hardness of water. The maximum limit of calcium ions in drinking water is 50 mg/L. Figure 3.19 shows the variation of Calcium Major Ions Concentration across all the stations for a typical year.

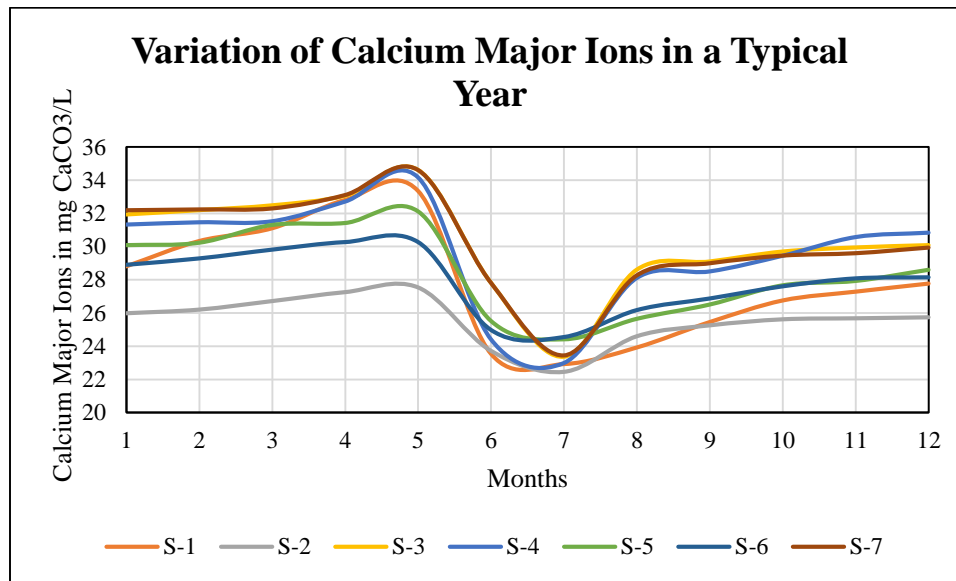


Fig 3.19: Variation of Calcium Major Ion Concentration for all the sites in a typical year

3.2.18 Magnesium Ion Concentration in mg/L:

A ubiquitous ion in sea water, this parameter accounts for hardness along with Calcium. It occurs due to mineral resources, fertilizers and industrial wastes. Figure 3.20 shows the variation of Magnesium Ion Concentrations across all the stations for a typical year.

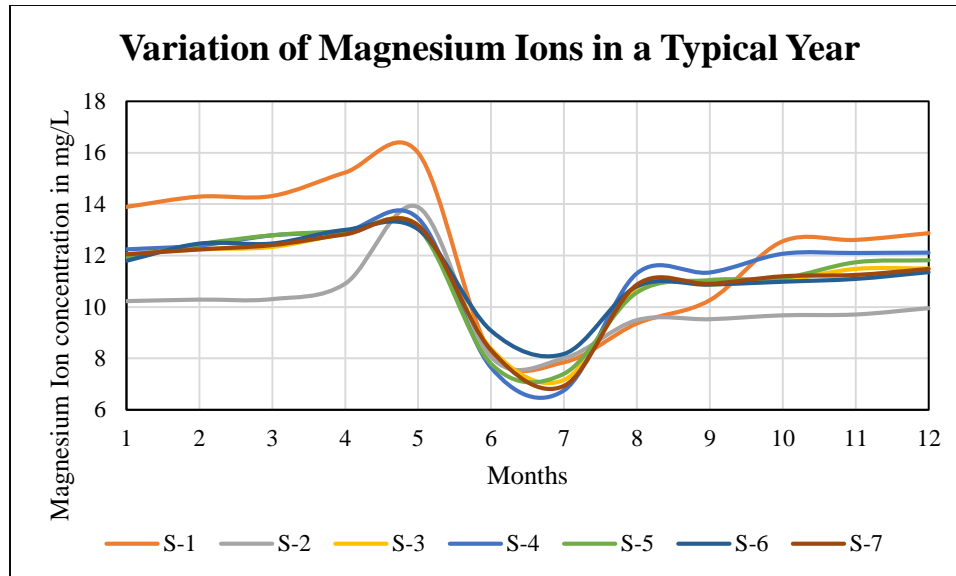


Fig 3.20: Variation of Magnesium Ion Concentration for all the sites in a typical year

3.2.19 Sodium Ion Concentration in mg/L:

This is a very common estimate present in water. Saline intrusion, effluents from sewage and mineral deposits can result contribute to the presence of sodium ions in water. Figure 3.21 shows the variation of Sodium Ion Concentrations across all the stations for a typical year.

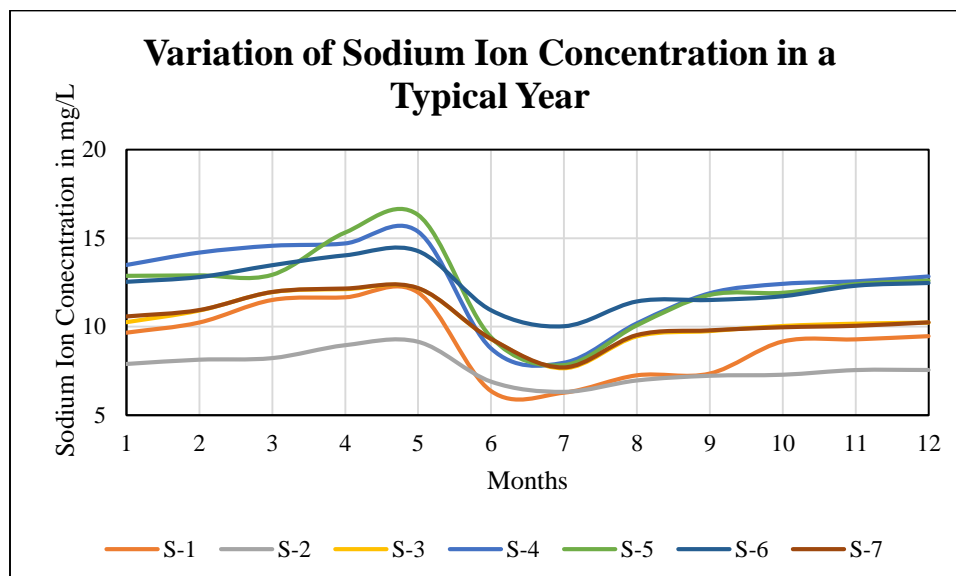


Fig 3.21: Variation of Sodium Ion Concentration for all the sites in a typical year

3.2.20 Potassium Ion Concentration in mg/L:

An essential constituent in human body, this parameter is usually within its permissible limits in water. Weathering of minerals and leaching of fertilizers result in an increased concentration of potassium ions in water. Figure 3.22 shows the variation of Potassium Ion Concentration across all the stations for a typical year.

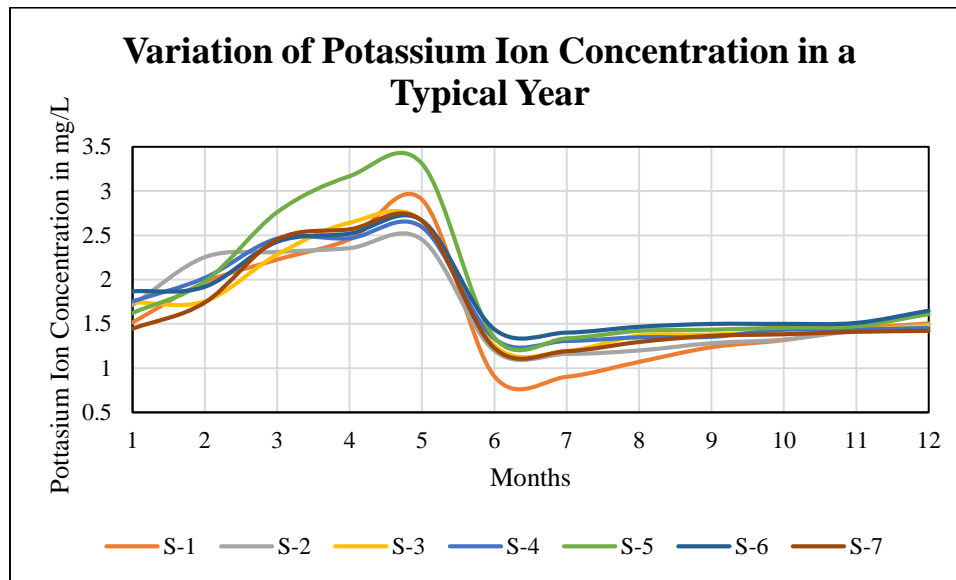


Fig 3.22: Variation of Potassium Ion Concentration for all the sites in a typical year

3.2.21 Chloride Ions Concentration in mg/L:

Leached from different rocks, chloride ions account for the salty taste to water when present above 250 mg/L. These ions may occur due to industrial and domestic wastes. Figure 3.23 shows the variation of Chloride Ions Concentration across all the stations for a typical year.

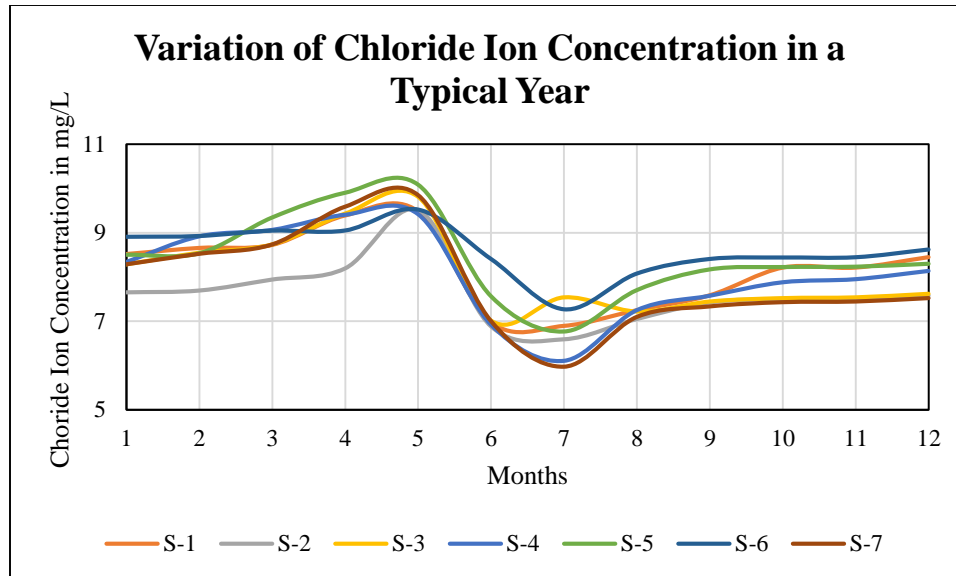


Fig 3.23: Variation of Chloride Ion Concentration for all the sites in a typical year

3.2.22 Sulphate Ions Concentration in mg/L:

This water quality variable originates from minerals like barite, epsomite and gypsum. The concentrations depend upon the cations associated with the sulphate ions. The wastewater from smelters and mines develop the concentration of this variable in water. Figure 3.24 shows the variation of Sulphate Ions Concentration across all the stations for a typical year.

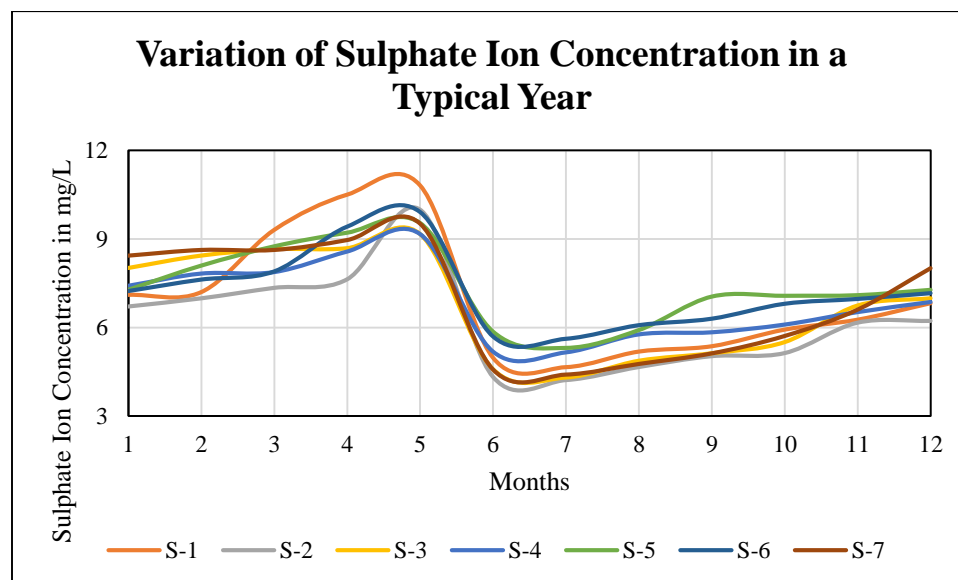


Fig 3.24: Variation of Sulphate Ion Concentration for all the sites in a typical year

3.2.23 Carbonate Ion Concentration in mg/L:

This estimate accounts for an increase in pH and alkalinity in a water sample. It is usually formed from the dissociation of carbon dioxide in water. Figure 3.25 shows the variation of Carbonate Ion Concentration across all the stations for a typical year.

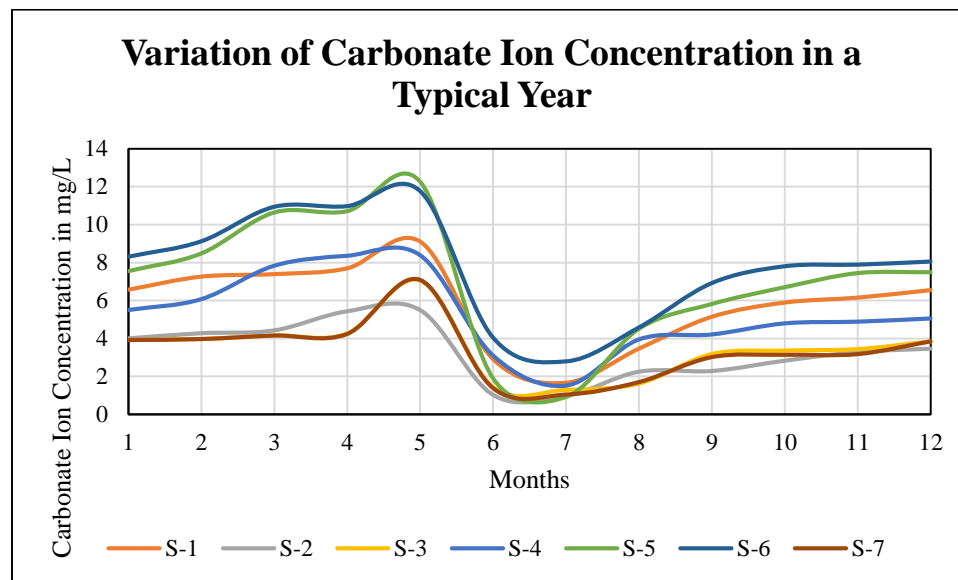


Fig 3.25: Variation of Carbonate Ion Concentration for all the sites in a typical year

3.2.24 Bicarbonate Ion Concentration in mg/L:

Resulting from the attempt to balance the carbonate equilibria, this parameter is abundantly found in running waters. The erosion of limestone results in the formation of this anion. Figure 3.26 shows the variation of Bicarbonate Ion Concentration across all the stations for a typical year.

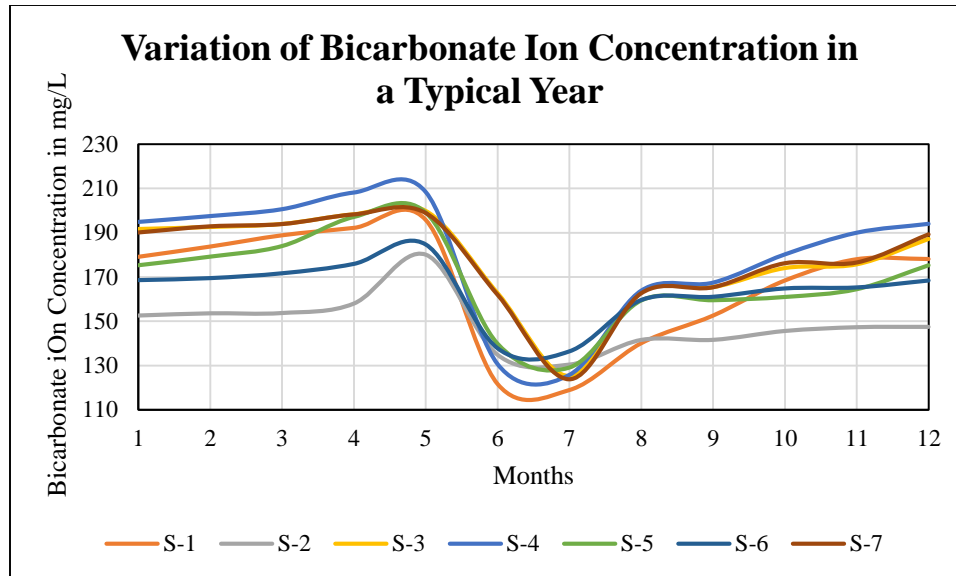


Fig 3.26: Variation of Bicarbonate Ion Concentration for all the sites in a typical year

3.2.25 Total Coliform Count in MPN/100 mL:

This variable is the collection of different bacteria in water. This may not be harmful but has the potential of being victimized by pathogens. Severe environmental pollution is responsible for this estimate. Figure 3.27 shows the variation of Total Coliform Count across all the stations for a typical year.

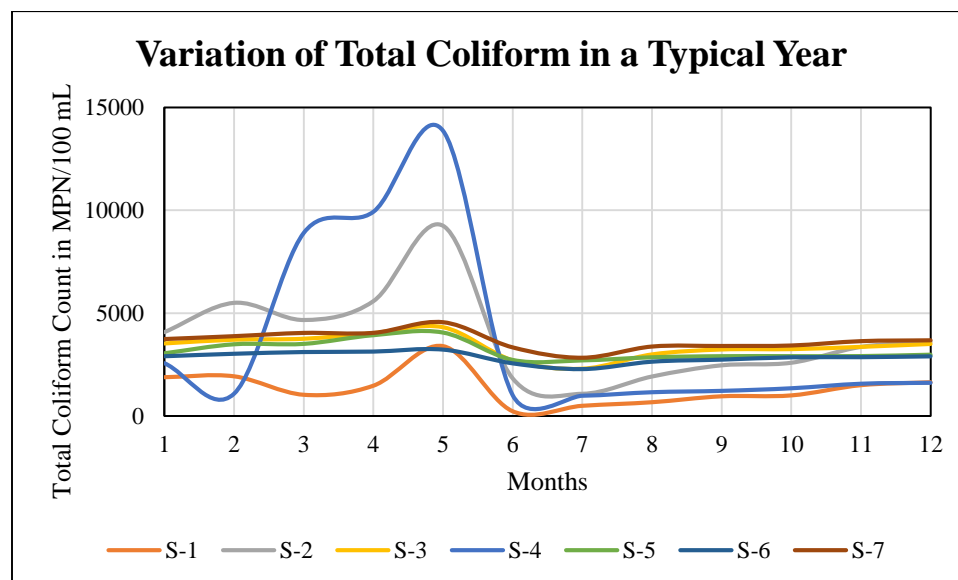


Fig 3.27: Variation of Total Coliform for all the sites in a typical year

3.2.26 Faecal Coliform Count in MPN/100 mL:

This variable results from anthropogenic pollution. The presence of pathogens is more prominent if a water sample contains Faecal Coliform.

Figure 3.28 shows the variation of Faecal Coliform Count across all the stations for a typical year.

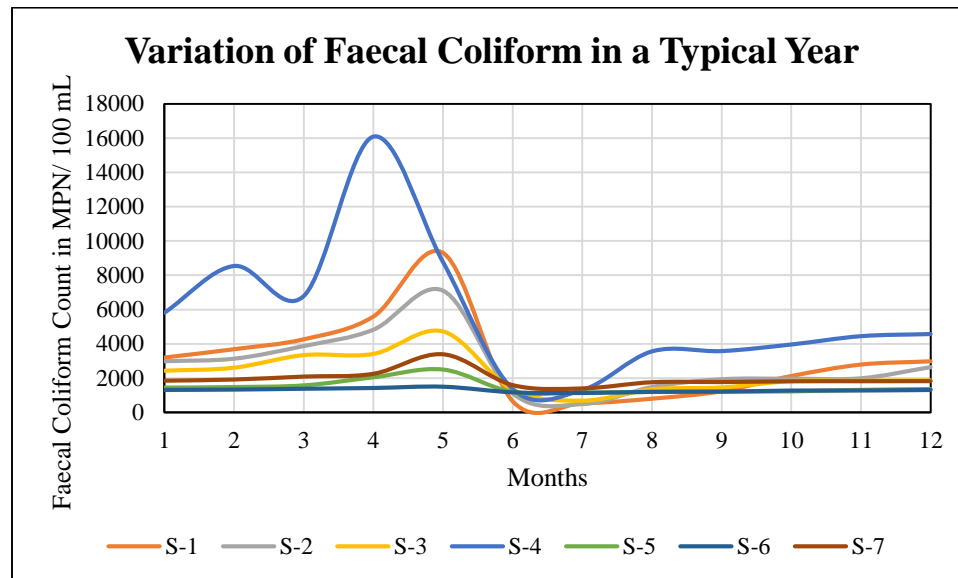


Fig 3.28: Variation of Faecal Coliform for all the sites in a typical year

CHAPTER IV **METHDOLOGY**

4.1 WATER QUALITY INDEX

Water quality indices can be devised by using increasing or decreasing scale indices. The method used in the current research work is of decreasing scale indices where the degree of pollution is inversely related to the index values. The formulation of water quality index involves four steps:

- i) Selection of Parameters:
- ii) Transformation of the selected parameters, having different units, into a common scale.
- iii) Assigning weights to the transformed parameters.
- iv) Aggregating the results obtained in step (iii) to determine a final index value.

Out of the above given steps, all are mandatory except step (iii). Water Quality Indices facilitate common people to understand the level of pollution of any water resource. However, formulation of Water Quality Index is not that simple. Various subjective opinions and references from experts are necessary and complete objectivity cannot be attained. The steps mentioned are explained below:

4.1.1 Parameter Selection:

A Water Quality Index would become clumsy if all the constituents present in a water sample would be included in the process. Therefore one needs to select a set of variables which as a whole reflect the overall quality of water for any given use. This is the place where subjectivity pops out. Different experts may have different opinions regarding the choice of sensitive parameters.

In the current work, a method of Student's t-test has been used to select three significant parameters from each chosen site.

4.1.1.1 Student's t-test:

Let us think of a small sample with size n from a normal population, having mean μ and standard deviation σ . Let \bar{y} be the mean and σ_s be the standard deviation of the sample drawn. Then, 't' statistic is given by equation 4.1 as

$$t = \frac{\bar{y} - \mu}{\sigma} \sqrt{n} \text{-----} (4.1)$$

where n is the number of observations. When the 't' statistic of each sample is calculated, we get a t distribution which is essentially symmetric and is normally distributed for large samples. The t-test was found by W.S Gosset in 1908 under the pen name of Student and hence the name 'Student's t-test'. This hypothesis test tells us whether means of two samples are statistically significantly different from each other, the null hypothesis being the means are the same. Thus, t-test belongs to the category of inferential statistics. If the t value is big, then the two samples can be considered significantly different or vice versa. How big the t value should be is better determined using the p-value which is the probability of the t value of lying with the region of significance. There are two levels of significance i.e. 5% and 1%. If for a t-test $p > 0.05$ (for 5% level of significance), the difference between \bar{y} and σ is not significant and if $p < 0.05$ the means are significantly different. In the current study 5% significance level has been chosen. Considering monthly discharge as an

independent parameter, the t-test is conducted between each water quality variable and discharge. Thus from this hypothesis testing we identify the variables which are independent of discharge. Further a two-tailed test is adopted thus assuming the relationship possibility to be in both the directions. The concept of a t-distribution has been elaborated in Figure 4.1.

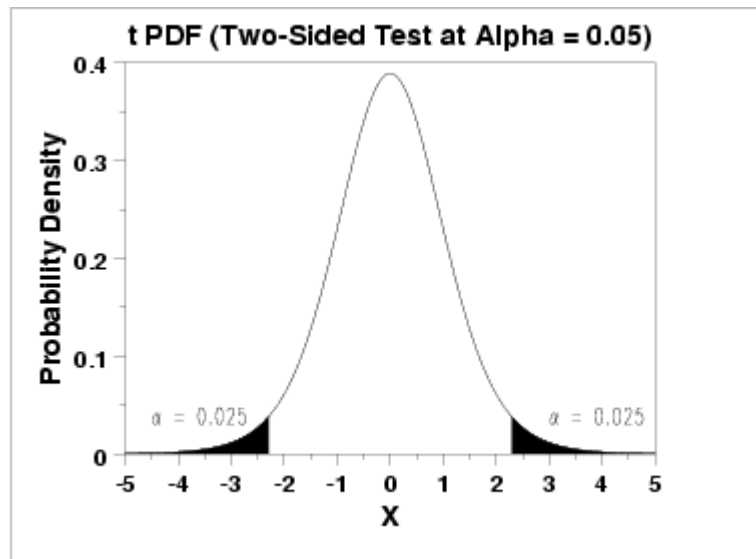


Fig 4.1: A typical two-sided t-distribution (PDF = probability density function)

4.1.2 Conversion of parameters of different units into a common scale:

Different parameters have different ranges of occurrence and different are measured in different units. pH has no unit and is limited to a value of 14. Dissolved Oxygen is measured in mg/L and is rarely beyond 12mg/L while sodium, being measured in the same unit may extend up to 1000mg/L. Toxic constituents like mercury occur mostly below 1mg/L. On the other hand, alkalinity and hardness always occur above 1mg/L. Still a water sample with a chloride content of 10mg/L is as healthy as one containing chloride fifteen times higher the former. But a water sample constituting of mercury of level 0.001mg/L is acceptable and one containing even twice of that is harmful.

Therefore transformation of the different parameters into a common scale is a necessity. The scale usually ranges from zero to one. It may also range from zero to hundred.

4.1.2.1 Development of Sub-indices:

Various mathematical functions are applied to formulate different pollutant variables, thus, generating sub-index functions. For n water quality variables x_1, x_2, x_3 and so on each variable is represented by a sub-index $S.I_i$ calculated using a function as per Equation 4.2:

$$S.I_i = f_i(x_i) \text{-----} (4.2)$$

After computation of the sub-indices, they are aggregated to form a final value in the next mathematical step as shown in Equation 4.3.

$$I = g(S.I_1, S.I_2, \dots S.I_n) \text{-----} (4.3)$$

The function may be a summation or a multiplicative operation or some other aggregation method. The overall procedure for development of sub-indices has been explained in the Figure 4.2:

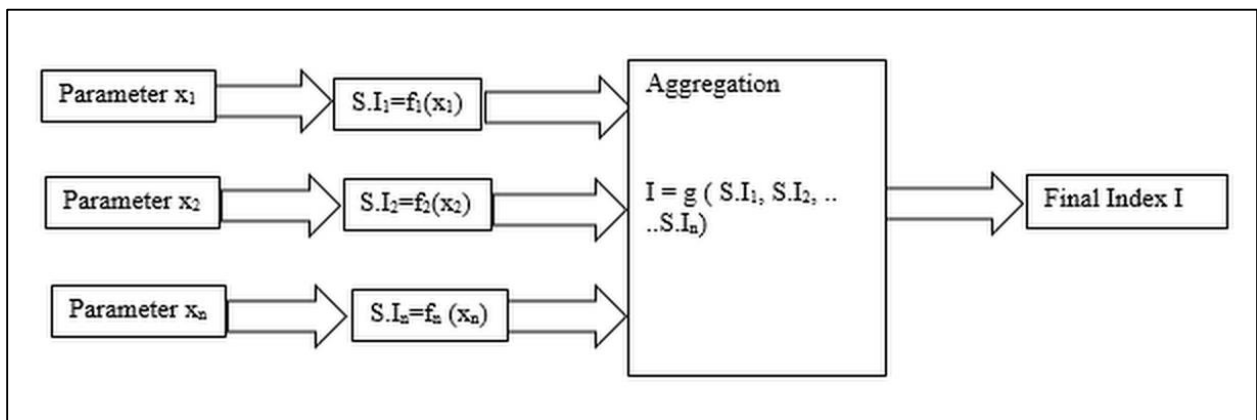


Fig 4.2: Mechanism of Water Quality Indexing

4.1.2.2 Various types of Sub-Indices:

The sub-indices can be of the following types:

- i) **Linear Function Sub-Indices:** This is the simplest function to compute sub-indices and is directly proportional to the pollutant concentration.
- ii) **Segmented Linear Function Sub-Indices:** Such functions constitute of two or more linear functions resulting in straight line segments which are connected at threshold levels. These functions are used in BIS, WHO and other administrative limits. The current research work also incorporates this method.
- iii) **Non-Linear Function Sub-Indices:** This function is used when the relationship between cause and effect is non-linear. It can either be plotted on a graph (implicit function) or can be defined by a mathematical equation (explicit function).
- iv) **Segmented Non-Linear Function Sub-Indices:** These are similar to segmented linear functions with at least one segment being non-linear.

4.1.3 Assigning Weights to the Sub-Indices:

Out of the numerous water quality variables, very few have to be shortlisted, thus keeping a balance between the index size, the authenticity of the water quality data and the potency of the indices. Again after shortlisting the variables, all of them may not be of equal importance. The assignment often becomes subjective. Hence proper care must be taken to assign the weightages judiciously. The following techniques are used for assigning weightages to the indices:

4.1.3.1 Information Entropy Method

Originating from thermodynamics, the concept of the theory of entropy was mingled by Shannon in 1948 with information theory. The degree of disorderliness of a system is explained by information theory. More the entropy value in an information system, more will be the randomness possessed by the system and hence less will be the information sought from the data (Abbasi and Abbasi, 2012; Zeleny, 1982). Thus entropy is an account of the uncertainties as well as the information retrieved from a system. Hence, it can be applied as an objective means of determining the weights of the variables based on the amount of useful information available in a dataset. More the entropy, more will be the uncertainty and less will be the useful information. Entropy and entropy weight are inversely related.

There are three significant variables ($m=3$) and 276 samples ($n=276$) collected month wise from 1991 to 2013. The raw values of each variable are normalized using the efficiency type construction function as shown in Equation 4.4.

$$Z_{ij} = \frac{x_{ij} - x_{ijmin}}{x_{ijmax} - x_{ijmin}} \text{-----} (4.4)$$

where $i=1,2,..m$; $j=1,2,3..n$ and $x_{ijmax} - x_{ijmin} \neq 0$.

Now the ratio of the value of the j^{th} index in the i^{th} sample is computed using Equation 4.5

$$R_{ij} = \frac{z_{ij}}{\sum_{i=1}^m z_{ij}} \text{-----} (4.5)$$

where $\sum_{i=1}^m z_{ij} \neq 0$.

The information entropy is obtained by Equation 4.6.

$$e^*_j = -\frac{1}{\ln m} \sum_{i=1}^m R_{ij} \ln R_{ij} \text{-----} (4.6)$$

where $\ln m \neq 0$.

The value of entropy is inversely proportional to the useful information of the water quality variable. The entropy weight is finally calculated as in Equation 4.7.

$$w^*_{ij} = \frac{1-e^*_j}{\sum_{j=1}^n 1-e^*_j} \text{-----} (4.7)$$

where $\sum_{j=1}^n (1 - e^*_j) \neq 0$.

The entropy weight of a variable is directly proportional to the useful information extracted from the dataset, and inversely proportional to the entropy.

4.1.3.2 Relative Pollution Degree

If entropy weight accounts for the deteriorating water quality, another weight can account for the improvement in water quality. This can be done by subtracting the calculated sub-index values from the ideal value (here, 100). This can be used as a measure for finding the rejuvenating capacity of the river. Equation 4.8 has been used to determine the monthly weights of all the significant parameters and later they are averaged to obtain the weights for a typical year.

$$W_{R.P} = \frac{100 - S.I}{\sum_{i=1}^n (100 - S.I)} \text{-----} (4.8)$$

where n is the number of observations and $\sum_{i=1}^n (100 - S.I) \neq 0$.

4.1.4 Sub-Indices Aggregated to Form the Final Index

The three most basic methods of aggregation of sub-indices are additive, multiplicative and logical. Horton's Indices (1965) and Indices of Brown (1970) are based on additive aggregation; Indices of Bhargava (1985) and Dinius (1987) are based on multiplicative aggregation; Smith Indices (1990) are based on logical aggregation. These operations have also been combinedly used by scientists like Inhaber (1975) and Dojlido (1994) to compute the final indices. Nowadays fuzzy techniques, principal component analysis, genetic algorithms, entropy and other concepts are being increasingly used to develop the final indices. The operators of Ordered Weighted Averaging (OWA) formulated by Yager in 1988 are also being used in water quality indexing.

4.1.4.1 Linear Sum Index:

This is worked out by adding unweighted sub-indices given by Equation 4.9:

$$I = \sum_{i=1}^n S.I_i \text{-----} (4.9)$$

where $S.I_i$ is the sub-index for the water quality variable i and n is the total number of variables. This method is simple yet is susceptible to the problem of ambiguity (when the final index violates the permissible limit though the variables are well within the limits).

4.1.4.2 Weighted Sum Index:

This index is defined by Equation 4.10

$$I = \sum_{i=1}^n w_i S_i I_i \text{-----} (4.10)$$

where S_i is the sub-index value for the i^{th} water quality variable and w_i is the corresponding weight. Here the problem of ambiguity is removed. But this method of aggregation is often affected by the problem of eclipsing, where the final index does not reflect the violation of the permissible limits by one or more pollutant variables. Weighted Sum Index has been used in the current research work.

Another problem associated with water quality index is rigidity. This problem comes into play when additional water quality variables are needed to define the water quality based on a new region or a new circumstance.

4.2 FUNCTIONAL DATA ANALYSIS

This is a new statistical technique which is used to assess a huge set of longitudinal data. The data set is represented in the form of curves and each curve becomes a single entity. The first step converts discrete data into smooth concentration curves. The question of smoothing arises if there is considerable measurement error. The combination of basis functions may be used as shown in Equation 4.11

$$y_i(t) = \sum_{m=1}^M \alpha_{im} \tau_{im}(t) \quad i = 1, 2 \dots L; t \in T \text{-----} (4.11)$$

where T is the time domain, $\tau_{im}(t)$ is the basis function (Fourier Series, B-splines, wavelets) and α_{im} are the coefficients found from the equation 4.12:

$$\min \left\{ \sum_{k=1}^{a_i} \left(\sum_{m=1}^M \alpha_{im} \tau_{im}(t) - y_i(t_{ik}) \right)^2 \right\} \text{-----} (4.12)$$

Fourier series are used when the data is periodic, B-splines used when non-periodic and wavelets used in case of severe fluctuations. In the current method smoothing spline is used

to cover the discrete sub-index values obtained from MATLAB Programming. The weights assigned to the sub-indices in this method are based on Entropy Weight method, Relative Pollution Degree and Conventional Water Quality Indexing. The following logic is used to compute the sub-indices.

For the variables where a larger value improves the water quality, Equation 4.13 defines the indices.

$$y_i(t) = \begin{cases} 100 & \text{when } x_i(t) > c_{i0} \\ \frac{100(5-j)}{5} + \frac{100}{5} \cdot \frac{x_i(t)-c_{ij}}{c_{i(j-1)}-c_{ij}} & \text{when } c_{ij} < x_i(t) \leq c_{i(j-1)} \text{-----} (4.13) \\ 0 & \text{when } x_i(t) \geq c_{i5} \end{cases}$$

And for the variables which deteriorate the water quality, the sub-indices are found out by Equation 4.14:

$$y_i(t) = \begin{cases} 100 & \text{when } x_i(t) < c_{i0} \\ \frac{100(5-j)}{5} + \frac{100}{5} \cdot \frac{c_{ij}-x_i(t)}{c_{ij}-c_{i(j-1)}} & \text{when } c_{i(j-1)} \leq x_i(t) < c_{ij} \text{-----} (4.14) \\ 0 & \text{when } x_i(t) \geq c_{i5} \end{cases}$$

where $y_i(t)$ is the sub-index curve for each indicator.

4.3 MULTIVARIATE STATISTICS

4.3.1 Principal Component Analysis

This is one of the multivariate techniques used for identifying the linear components belonging to a huge set of data. Multivariate analysis, on the other hand, refers to those where more than one outcome variables are expected. Principal Component Analysis thus gives a direction towards maximum variance. In this method we deal with a correlation matrix. This matrix defines the linear dependence between a pair of variables. From this matrix, variates

are calculated using the Eigen vectors from the correlation matrix and their number is same as the number of variables. Only large Eigen values are extracted thus accounting for maximum variance. Equation 4.15 is used to define a component.

$$Z_i = a_1X_{1i} + a_2X_{2i} + \dots + a_nX_{ni} \text{-----} (4.15)$$

where $a_1, a_2 \dots a_n$ represent the loadings which are the Eigen vectors and $X_{1i}, X_{2i} \dots X_{ni}$ are the normalized values of the variables.

For use in indexing the first component is extracted. Equation 4.16 is used to calculate each index.

$$WQI = \sum_{i=1}^n \left\{ \frac{I_i}{\sum_i^n (I_i \times L_i)} \times \frac{[C_i]}{[S_i]} \right\} \text{-----} (4.16)$$

where I is the amount of variance explained by a component i, L is the loading of the i^{th} component, C is the concentration of the water quality variable in the i^{th} component and S is the quality criterion of the water quality variable in the i^{th} component and $i=1,2,3..n$ and the denominator i.e. $\sum_i^n ((I_i \times L_i) \times [S_i]) \neq 0$. In the present study the only first component for each station is considered.

4.3.2 Principal Factor Analysis

Principal Factor Analysis is another multivariate technique used for identifying the probability of correlations between variables occurring due to one or more latent variables. Each latent variable occurs in the form of a linear model. The factors account for the underlying variables. In a correlation matrix, the large off-diagonal coefficients between variable subsets imply that those variables can be the measuring aspect of a similar underlying dimension. This underlying dimension is referred to as a factor. Thus data set is reduces into a few factors

attaining parsimony and also explaining maximum common variance. The mathematical representation of a factor is identical as that of a component i.e. Equation 4.15.

4.3.3 Cluster Analysis

This multivariate data mining technique classifies data into clusters. The concept of clustering adheres to homogeneity within then and heterogeneity across depending on the characteristics. There is no objective function and hence, no dependent variable. Since clustering depends on the values of input parameters, it is often known as subjective segmentation and hence, unsupervised classification. This method may be of various types: K-means, Functional Embedding (FEM), Hierarchical Clustering, Support Vector Machines, K-Nearest Neighbour and so on. The method used in the current work is hierarchical clustering.

Hierarchical Clustering starts from each variable as an independent cluster. At each step, when two clusters are closer to each other, they are combined as a single one. This process is continued till a single cluster is obtained. Thus sequential partitional clustering leads to hierarchical clustering. This method holds good for a smaller set of data. However, it can also be satisfactorily used for a bigger data set. The results are shown in the form of a dendrogram which displays the relationship between clusters and their sub-clusters and the order in which clusters are split or merged.

4.4 MISCELLANOUS METHODOGIES

Standard Deviation is a measure of data dispersion about the mean. A value close to 1 refers to a compact data set while a larger value refers to a dispersed data set thus showing more fluctuations and perhaps deterioration of water quality, in case of assessment of water quality variables. This statistic is computed using equation 4.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (\bar{y} - y_i)^2}{n - 1}}$$

where \bar{y} is the mean, y_i are the individual values and the denominator $\sqrt{n - 1} \neq 0$.

Skewness is another measure of variability of data accounting for the asymmetric distribution of data. A value near zero indicates a symmetric distribution. The data that are right-skewed have positive skewness and those which are left-skewed have negative skewness.

Kurtosis is a measure of the peakedness of a data set. A higher value represents more variance resulting from occasional deviations.

A scree plot is used to display the number of factors that should have been extracted. It is a plot between Eigen values (or the variance explained) and the components formed. In a typical graph, a sharp descent is followed by a gradual tailing off. The point of deflection or the elbow of the plot is used as the means for extraction.

Rotation of results improves the solution without changing it. The types of rotation may be orthogonal or oblique. The type of rotation used in the present work is varimax which belongs to orthogonal rotation. This method maximizes the dispersion of loadings with a factor or a component. Thus a lesser number of variables are highly loaded onto a component making the result more interpretable.

Kaiser-Meyer-Olkin Measure of sampling adequacy (KMO) represents the ratio of squared correlations to the squared partial correlations among variables. This value spans between 0 and 1. A value close to zero implies that there is more partial correlation thus more diffusion in the correlation patterns. In this case, factor analysis is not to be used. A value close to one indicates compact patterns of correlations and factor analysis is appropriate in this case. Values lying between 0.5 and 0.7 are mediocre and between 0.7 and 0.8 are good.

Bartlett's test of sphericity checks whether the correlation matrix is an identity matrix. The determinate of the correlation matrix is converted into a chi square value and further tested for significance. For a large sample, the chi square value is also large. The level of significance is kept at 5%.

4.5 PERMISSIBLE WATER QUALITY CRITERIA

On account of the methods mentioned above Table 4.1 briefly describes the acceptable and permissible limits of the water quality variables used in the present study.

Table 4.1 Permissible and acceptable limits of water quality parameters

Sl. No	Water Quality Parameter	Acceptable Limit	Permissible Limit (where alternative source is absent)	Undesirable Effect Beyond Permissible Limit
1	Temperature in °C	10-15	-	High water temperature can increase its solubility and decrease the amount of Dissolved Oxygen in it.
2	pH	6.5-8.5	No relaxation	pH affects the mucous system of the human body and the water supply system as well.
3	EC in $\mu\text{mho/cm}$ (max)	500	2000	Beyond this palatability decreases and may cause Gastro intestinal irritation
4	TDS in $\mu\text{mho/cm}$ (max)	500	2000	Beyond this palatability decreases

				and may cause Gastro intestinal irritation
5	NH ₃ ion concentration in mg/L (max)	0.5	No relaxation	High Ammonia concentration indicates faecal pollution and agricultural run-off. This often reduces the quantity of Dissolved Oxygen.
6	NO ₂ and NO ₃ ion concentration in mg/L (max)	45	No relaxation	Beyond this methaemoglobinemia takes place.
7	SiO ₂ ion concentration in mg/L	0	2	May disturb phosphorylation of human body and causes boiler scale in industry
8	DO in mg/L	15	4	Mortality rates of fishes will rise and the concentration of stable pollutant compounds will increase.
9	BOD in mg/L	0	No relaxation	This indicates reduced level of oxygen, hence reduced aquatic life and rejuvenating power of the stream.
10	COD in mg/L	0	No relaxation	This indicates reduced level of oxygen, hence reduced aquatic life and rejuvenating power of the stream.
11	Phenol ion concentration in mg/L (max)	0.001	0.002	Beyond this, it may cause objectionable taste and odour
12	Total Alkalinity in mg/L	200	600	Beyond this limit, taste becomes unpleasant
13	Total Hardness (as CaCO ₃) in mg/L	300	600	Encrustations in water supply structure and adverse effect on domestic use
14	Ca Hardness in mg/L (max)	200	600	Scaling of utensils and also of hot water systems; scums in soaps.
15	F ⁻ ion concentration in mg/L (max)	1.0	1.5	Fluoride may be kept as low as possible. High fluoride may cause fluorosis

16	Ca Major Ion concentration in mg/L (max)	75	200	Encrustations in water supply structure and adverse effect on domestic use
17	Mg ²⁺ ion concentration in mg/L (max)	30	100	Encrustations in water supply structure and adverse effect on domestic use
18	Na ⁺ ion concentration in mg/L	20-250	-	Non-volatile and found associated with particulate matter.
19	K ⁺ ion concentration in mg/L	8	400	Abnormal increase in K ⁺ ions may cause toxicity and breathing problems.
20	Cl ⁻ ion concentration in mg/L (max)	250	1000	Beyond this limit taste, corrosion and palatability are affected
21	SO ₄ ²⁻ ion concentration in mg/L (max)	200	400	Beyond this causes Gastro intestinal irritation when magnesium or sodium are present
22	CO ₃ ²⁻ ion concentration in mg/L	90	450	Increases the SAR index in irrigation water.
23	HCO ₃ ⁻ ion concentration in mg/L	90	450	Increases the SAR index in irrigation water.
24	TC in MPN/100mL	Shall not be detectable in any 100 ml sample	Shall not be detectable in any 100 ml sample	Presence of Coliform indicate presence of pathogenic bacteria
25	FC in MPN/100mL	Shall not be detectable in any 100 ml sample	Shall not be detectable in any 100 ml sample	Presence of Coliform indicate presence of pathogenic bacteria

CHAPTER V

RESULTS AND DISCUSSIONS

5.1 PRELIMINARY ANALYSIS

5.1.1 Descriptive Statistics:

Descriptive statistical analysis is carried out on the water quality variables from 2009-2013 to know about the recent water quality status and are displayed in Table 4.1. Many of the variables deviate from the recommended water quality criteria as mentioned in Chapter IV. Important variables like Biochemical Oxygen Demand, Chemical Oxygen Demand, Coliform Count and Faecal Coliform show unusual variations in terms of their maximum values, standard deviation, skewness and kurtosis as mentioned in Table 5.1.

Table 5.1- Descriptive Statistics for the span of 2009-2013

Water Quality Variables	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Discharge in cumecs	504	0.7	4407.9	365.4	6.6	3.1	0.1	13.7	0.2
Temperature in °C	504	7.6	35	28.4	2.5	22.4	0.1	1.8	0.2
pH	504	6.8	8.9	8.2	0.2	-1.3	0.1	4.2	0.2
EC in $\mu\text{mho/cm}$	504	108.0	759.0	270.3	9.6	1.5	0.1	9.9	0.2
TDS in $\mu\text{mho/cm}$	504	96.0	713.0	184.0	6.3	1.6	0.1	4.5	0.2
NH ₃ ion concentration in mg/L	504	0.0	604.0	24.7	6.9	3.1	0.1	3.5	0.2
NO ₂ and NO ₃ ion concentration in mg/L	504	0.0	54.0	94.3	7.1	21.0	0.1	5.2	0.2
SiO ₂ ion concentration in mg/L	504	0.0	64.4	23.5	1.3	-0.4	0.1	1.3	0.2
DO in mg/L	504	0.0	15.9	9.4	8.9	3.7	0.1	6.5	0.2
BOD in mg/L	504	0.0	10.2	5.1	0.6	0.6	0.1	0.3	0.2
COD in mg/L	504	7.0	30.7	23.8	2.2	1.5	0.1	7.3	0.2
Phenol ion concentration in mg/L	504	0.0	20.6	3.3	4.7	1.2	0.1	0.2	0.2
Total Alkalinity in mg/L	504	57.1	410.2	156.6	8.8	1.9	0.1	6.2	0.2
Total Hardness in mg/L	504	53.2	286.4	115.7	4.5	1.0	0.1	4.3	0.2
Ca Hardness in mg/L	504	41.0	210.8	69.9	2.1	0.7	0.1	0.7	0.2
F ⁻ ion concentration in mg/L	504	0.0	2.8	2.1	0.2	0.8	0.1	0.9	0.2
Ca Major Ion concentration	504	16.4	44.3	28.0	4.8	0.7	0.1	0.7	0.2
Mg ²⁺ ion concentration in mg/L	504	1.0	50.5	11.0	4.3	1.9	0.1	3.9	0.2
Na ⁺ ion concentration in mg/L	504	1.8	33.5	10.2	4.1	1.4	0.1	3.8	0.2

K ⁺ ion concentration in mg/L	504	0.1	6.8	1.8	0.9	2.0	0.1	5.2	0.2
Cl ⁻ ion concentration in mg/L	504	3.9	33.2	9.2	3.2	1.9	0.1	7.8	0.2
SO ₄ ²⁻ ion concentration in mg/L	504	0.1	48.4	9.1	9.0	2.3	0.1	5.0	0.2
CO ₃ ²⁻ ion concentration in mg/L	504	60.0	325.2	124.2	5.8	1.1	0.1	0.2	0.2
HCO ₃ ⁻ ion concentration in mg/L	504	71.0	491.0	177.1	4.9	1.7	0.1	7.4	0.2
TC in MPN/100mL	504	95	21500	1439.0	22.5	5.1	0.1	9.8	0.2
FC in MPN/100mL	504	36	9460	504.0	25.5	5.4	0.1	5.5	0.2
Valid N (list wise)	504								

5.1.2 PRINCIPAL COMPONENT ANALYSIS AND PRINCIPAL FACTOR ANALYSIS

5.1.2.1 Correlation Matrix

The correlation matrix accounts for the dependencies between two variables. The Table 5.2 represents the correlation matrix of all the variables accounted for. It can be seen that Electrical Conductivity has a significant positive relationship with Total Dissolved Solids ($R^2=0.87$), Calcium Hardness ($R^2=0.57$), Calcium Major Ions ($R^2=0.57$) and pH ($R^2=0.87$). Dissolved Oxygen bears a negative correlation with temperature ($R^2=-0.89$), Biochemical Oxygen Demand ($R^2=-0.87$), Chemical Oxygen Demand ($R^2=-0.85$). Fluoride is found to be positively correlated with Electrical Conductivity ($R^2=0.57$) and Total Dissolved Solids ($R^2=0.77$). Calcium Major Ions bears an R^2 value of 0.57 with EC and 0.77 with Total Hardness. Magnesium Ions bear a positive correlation of 0.84 with Total Hardness and 0.53 with Electrical Conductivity. Sodium Ion has an R^2 value of 0.51 and 0.46 with Electrical Conductivity and Total Dissolved Solids respectively. The carbonate ion bears a correlation coefficient of 0.96 with phenol and -0.61 with pH. The bicarbonate ion bears correlation coefficients of 0.52 with Electrical Conductivity, 0.72 with Total Alkalinity, 0.61 with Total Hardness and 0.55 with Calcium Major Ions. Coliform Count and Faecal Coliform are again positively correlated with an R^2 value of 0.96.

Calcium Major and Calcium Hardness have a correlation coefficient of 1 and hence treated as one variable. Positive correlations imply that the increase or decrease in one variable is directly proportional to the other. Negative Correlations mean the opposite.

TDS	NH ₃	Nitrate Nitrite	SiO ₂	DO	BOD	COD	Phenol	Total Alkalinity	Total Hardness	Ca Hardness
1.00										
0.61	1.00									
0.19	0.23	1.00								
-0.37	-0.65	-0.14	1.00							
0.52	0.76	0.22	-0.58	1.00						
-0.02	0.00	0.00	0.03	-0.87	1.00					
-0.04	-0.03	0.02	0.00	-0.85	0.83	1.00				
0.21	0.15	0.00	-0.07	0.14	0.00	-0.10	1.00			
0.27	0.03	0.04	0.11	0.05	0.02	0.14	0.08	1.00		
0.48	0.03	0.00	0.14	0.03	-0.01	-0.04	0.13	0.43	1.00	
0.39	0.01	0.01	0.10	0.02	0.01	0.03	0.05	0.40	0.77	1.00
0.85	0.01	0.03	-0.07	0.02	-0.02	0.17	-0.12	0.08	-0.09	-0.05
0.39	0.01	0.01	0.10	0.02	0.01	0.03	0.05	0.40	0.77	1.00
0.38	0.04	0.00	0.12	0.03	-0.03	-0.09	0.16	0.30	0.84	0.31
0.46	0.16	0.01	-0.04	0.16	0.01	-0.05	0.17	0.21	0.40	0.37
-0.07	0.03	-0.01	-0.13	-0.01	0.02	0.14	-0.11	-0.03	-0.13	-0.10
0.12	0.05	0.05	0.06	0.07	-0.02	0.03	0.01	0.32	0.21	0.18
-0.03	0.03	0.02	0.05	0.06	0.02	0.18	-0.15	0.00	-0.01	0.01
0.20	0.14	0.00	-0.06	0.13	0.02	-0.10	0.96	0.08	0.12	0.04
0.39	0.03	0.01	0.13	0.03	0.00	0.11	-0.10	0.72	0.61	0.54
-0.05	-0.05	-0.01	0.01	-0.77	0.82	0.78	-0.06	0.00	-0.03	0.01
-0.07	-0.07	-0.02	0.03	-0.78	0.81	0.84	-0.06	-0.01	-0.04	0.00

Variables	Temp	pH	EC
Temp	1.00		
pH	-0.12	1.00	
EC	-0.09	0.87	1.00
TDS	0.02	0.16	0.69
NH ₃	0.15	0.13	0.10
Nitrate Nitrite	-0.01	0.00	0.08
SiO ₂	-0.15	-0.06	0.09
DO	-0.89	0.10	0.48
BOD	0.85	-0.02	-0.02
COD	0.88	-0.08	-0.03
Phenol	-0.05	0.62	0.18
Total Alkalinity	-0.06	0.08	0.40
Total Hardness	-0.09	0.14	0.68
Ca Hardness	-0.04	0.06	0.57
F	0.09	-0.15	0.78
Ca Major	-0.04	0.06	0.57
Mg	-0.10	0.16	0.53
Na	-0.06	0.15	0.51
K	0.06	-0.12	-0.11
Cl	-0.04	0.04	0.23
SO ₄	0.04	-0.19	-0.02
CO ₃	-0.06	-0.61	0.17
HCO ₃	-0.08	-0.45	0.56
CC	0.91	-0.02	-0.04
FC	0.89	-0.03	-0.04

5.1.2.2 KMO-Sampling Adequacy

Table 5.3 gives us a measure of KMO sampling adequacy and Bartlett's Test of Sphericity. Kaiser-Meyer-Olkin Measure of Sampling Adequacy ranges between 0 and 1. With a minimum value of 0.6, the more the value, the better the measure.

In Bartlett's Test of Sphericity, the null hypothesis that the correlation matrix is an identity matrix, is checked and it is preferable to reject the null hypothesis. The present results also reject the null hypothesis. This is followed by an approximate chi-square distribution whose value should be large enough for the correlation matrix to be identical. Here the value of chi-square is 45280.53 which is large enough with enough significance.

From these results, we can conclude that our dataset is fit for principal factor analysis or principal component analysis.

Table 5.3: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.717
Bartlett's Test of Sphericity	Approx. Chi-Square	45280.53
	df	300
	Sig.	.000

5.1.2.3 Total Variance Explained

Table 5.4 illustrates about the variances explained by the principal components extracted from the data set. Seven components are extracted and they explain

68.57% variance of the system. The decrease in variance of the system is because of the diversified data set. The first component accounts for 22.9% of variance which is around half of the total variance explained.

Table 5.4: Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.508	22.949	22.949	5.508	22.949	22.949
2	2.640	10.998	33.947	2.640	10.998	33.947
3	2.318	9.659	43.606	2.318	9.659	43.606
4	1.915	7.979	51.586	1.915	7.979	51.586
5	1.764	7.348	58.934	1.764	7.348	58.934
6	1.208	5.032	63.966	1.208	5.032	63.966
7	1.106	4.607	68.573	1.106	4.607	68.573
8	.997	4.156	72.729			
9	.922	3.841	76.570			
10	.854	3.558	80.128			
11	.700	2.916	83.044			
12	.606	2.525	85.569			
13	.558	2.324	87.892			
14	.532	2.219	90.111			
15	.478	1.991	92.102			
16	.438	1.823	93.926			
17	.399	1.664	95.589			
18	.325	1.353	96.942			
19	.284	1.182	98.124			
20	.213	.889	99.013			
21	.122	.510	99.522			
22	.096	.399	99.921			
23	.019	.077	99.999			
24	.000	.001	100.000			

Extraction Method: Principal Component Analysis.

5.1.2.4 Principal Component Matrix

From Table 5.2 shown below, it is found that seven components have been extracted. The first component is strongly correlated with Dissolved Oxygen (0.854), Biochemical Oxygen Demand (-0.815), Chemical Oxygen Demand (-0.847), Total Hardness (0.711), Calcium Hardness (0.865), Calcium Major Ions (0.864), Electrical Conductivity (0.796), Coliform Count (-0.871) and Faecal Coliform (-0.892). A positive value implies a

direct relationship while a negative value states an inverse relationship. Thus, as Dissolved Oxygen decreases, Biochemical Oxygen Demand, Chemical Oxygen Demand, Coliform Count and Faecal Coliform Increase and the other variables decrease. The results obtained can be satisfactorily simulated with the current status report of River Narmada. The other components can be similarly interpreted.

Table 5.5: Principal Component Matrix

Water Quality Variables	Component						
	1	2	3	4	5	6	7
Temp	-.013	.198	-.104	-.048	.132	-.273	.426
pH	.072	.055	.771	.017	-.167	.053	-.058
EC	.796	.094	.129	-.024	-.060	.156	-.155
TDS	.594	.642	.126	-.025	-.042	.061	-.102
NH3	.067	.905	.091	-.031	.028	.017	.040
Nitrate and Nitrite	-.069	.375	-.064	.016	-.095	.339	-.050
SiO ₂	.103	-.796	-.030	-.012	-.080	.146	-.074
DO	.854	.858	.075	-.053	.013	.058	.021
BOD	-.815	-.034	.050	-.025	-.064	.085	.673
COD	-.847	-.052	-.035	.222	.394	.339	.249
Phenol	.078	.071	.947	-.033	-.077	-.023	.009
Total Alkalinity	.459	-.034	.065	-.017	.112	.662	.085
Total Hardness	.711	-.053	.062	-.029	-.086	.109	-.112
Calcium Hardness	.865	-.060	-.038	.043	-.062	.057	.269
F	-.063	.020	-.050	-.003	.707	.082	.020
Calcium Major	.864	-.059	-.038	.042	-.061	.056	.269
Mg	.628	-.027	.124	-.085	-.078	.114	-.400
Na	.588	.166	.170	-.008	.076	-.102	-.101
K	-.067	.032	-.036	-.004	.708	-.150	-.007
Cl	.139	.033	.018	-.036	.001	.649	-.057
SO ₄	.041	.007	-.140	.053	.592	.043	-.036
CO ₃	.070	.063	.948	-.034	-.052	-.019	.016
HCO ₃	.687	-.025	-.160	-.014	.089	.472	-.020
TC	-.871	-.020	-.019	-.053	.065	-.003	-.023
FC	-.892	-.041	-.022	-.078	.041	-.015	-.025
Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.							
a. Rotation converged in 6 iterations.							

5.1.2.5 Scree plot

Figure 5.1 shows the plot of eigen values and the twenty four variables considered. The plot shows that seven components show maximum variability. The other components do not account much for the variance and the line straightens after the seventh one.

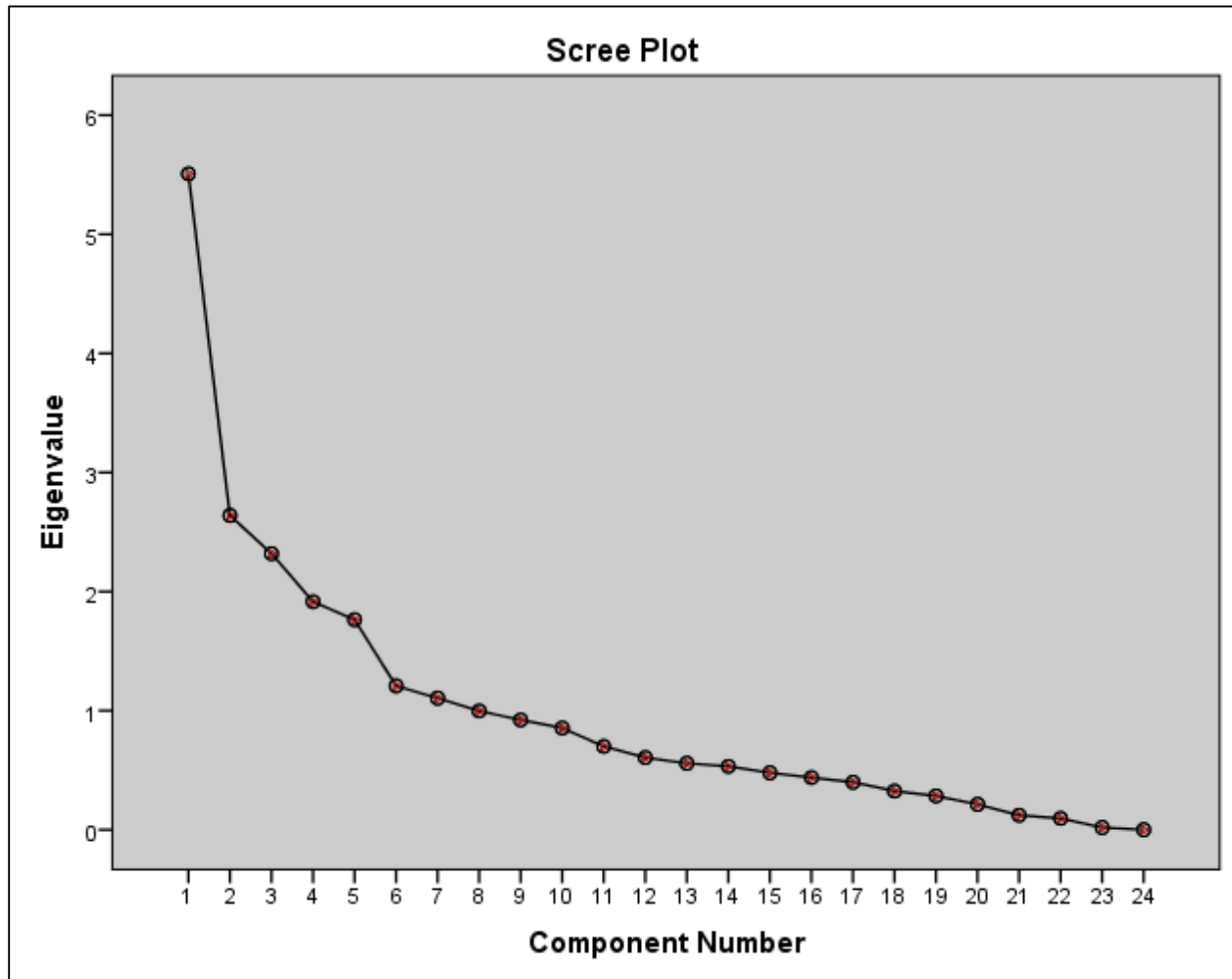


Fig 5.1: A Scree Plot of Eigen values vs Components

5.1.2.6 Rotated Factor Matrix

As mentioned in Table 5.3, Factor Analysis also yields 7 factors. The first factor is strongly correlated with Calcium Hardness (0.967), Calcium Major Ions (0.966), Dissolved Oxygen (0.822), Biochemical Oxygen Demand (-0.805), Chemical Oxygen Demand (-0.831), Total Hardness (0.709) and

Electrical Conductivity (0.820). The growing agricultural runoff can be made responsible for the increasing alkalinity of water. It can also be stated that an alkaline environment can help in the rejuvenation of water. The other strong correlations are boldfaced.

Table 5.6: Rotated Factor Matrix

Water Quality Variables	Factor						
	1	2	3	4	5	6	7
Temp	-.010	.153	-.066	-.113	-.027	-.057	.119
pH	.041	.059	.608	.048	.011	.049	-.218
EC	.820	.114	.136	.422	-.020	.369	-.118
TDS	.378	.644	.138	.277	-.019	.310	-.099
NH ₃	.025	.909	.098	.010	-.024	.047	.020
Nitrate and Nitrite	-.002	.248	-.016	.054	.000	.004	.000
SiO ₂	.051	-.710	-.043	.161	-.002	.051	-.079
DO	.822	.802	.088	.030	-.043	.033	.011
BOD	-.805	-.010	.008	.008	-.013	-.039	.026
COD	-.831	-.027	-.058	.142	.142	-.108	.300
Phenol	.046	.058	.971	.004	-.030	.020	-.112
Total Alkalinity	.215	.000	.073	.801	-.002	-.005	.090
Total Hardness	.709	-.048	.082	.318	-.041	.586	-.053
Calcium Hardness	.967	-.021	-.002	.233	.012	-.017	-.034
F	-.039	.033	-.052	.032	.007	-.061	.542
Calcium Major Ions	.966	-.020	-.002	.233	.011	-.017	-.034
Mg	.253	-.050	.124	.278	-.075	.857	-.058
Na	.359	.170	.150	.209	-.010	.206	-.027
K	-.064	.047	-.051	-.065	.008	-.037	.488
Cl	.105	.026	.017	.327	-.013	.042	.024
SO ₄	.017	.020	-.114	.038	.043	.019	.412
CO ₃	.044	.050	.971	.002	-.032	.012	-.079
HCO ₃	.385	.013	-.131	.769	-.005	.177	.061
TC	.736	-.030	-.013	-.014	.977	.017	.105
FC	.710	-.053	-.018	-.020	.966	.021	.076
Extraction Method: Principal Axis Factoring. Rotation Method: Varimax with Kaiser Normalization.							
a. Rotation converged in 7 iterations.							

5.1.2.7 CLUSTER ANALYSIS

In figure 5.1, the results of hierarchical clustering are shown. The Y displayed on the left hand side denotes the closeness between individual clusters. The rescaled distance cluster combine gives an account of the distances on the basis of which the clusters are combined. It may be noted that the ratio of rescaled distances in the dendrogram and those of the original distances are the same. Four distinct clusters containing Biochemical Oxygen Demand, Chemical Oxygen Demand, Coliform Count, Faecal Coliform and the remaining variables are found. Thus the clusters constituting the individual variables imply that the surface water quality of River Narmada is primarily governed by the four variables mentioned.

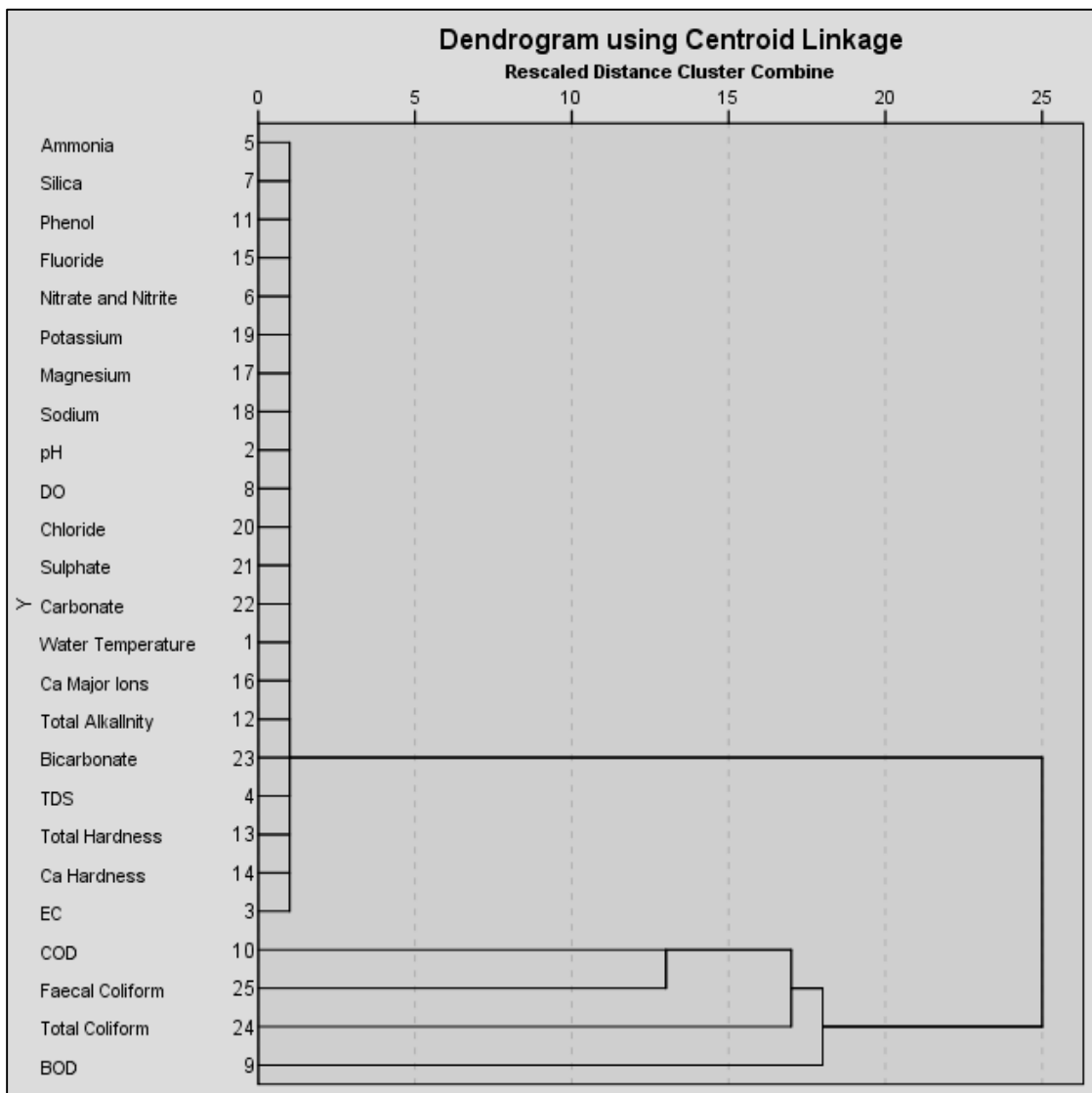


Fig 5.2: Dendrogram obtained using Hierarchical Clustering

5.2 Student's t-test

Out of the twenty five water quality variables of each of the seven stations, three most significant variables are determined using Student's t-test. The following table depicts the p-values of the t-test for the three most significant parameters of each station.

Table 5.7: p-Values for Student t-tests applied all the water quality variables

Sl.No.	Parameter 1	p-value	Parameter 2	p-value	Parameter 3	p-value
S1	Hardness	0.002	pH	0.0032	Phosphate Ions	0.004
S2	BOD	0.0012	TC	0.0043	TDS	0.006
S3	Ammonia	0.0047	pH	0.0052	Fluoride	0.0071
S4	BOD	0.0001	TC	0.0009	COD	0.0073
S5	TDS	0.0044	BOD	0.0061	Chloride	0.0079
S6	pH	0.0027	BOD	0.0042	EC	0.0067
S7	Nitrate	0.001	TDS	0.0048	pH	0.0078

All the water quality variables (N=277) are associated with the corresponding discharge of each station. ($\bar{y}_1 = 38.633$ cumecs and $\sigma = 4.45$, $\bar{y}_2 = 331$ cumecs and $\sigma = 4.7$, $\bar{y}_3 = 442$ cumecs and $\sigma = 8.5$, $\bar{y}_4 = 597$ cumecs and $\sigma = 2.7$, $\bar{y}_5 = 724.3$ cumecs and $\sigma = 6.3$, $\bar{y}_6 = 866.15$ cumecs and $\sigma = 5.3$, $\bar{y}_7 = 441.1$ cumecs and $\sigma = 8.2$). To test the hypothesis that all the variables for each station are associated with statistically significantly different mean from discharge, a two-tailed independent heteroscedastic t-test is conducted. The distributions of all the variables are sufficiently normal for the purpose of conducting a t-test (i.e. skewness $< |2.0|$ and kurtosis $< |9.0|$; Schmider, Ziegler, Danay, Beyer and Buhner, 2010). The three most significant variables from each station are chosen for further analysis.

5.3 Conventional Water Quality Indexing:

Water Quality Indices were calculated using the conventional method mentioned in Chapter IV. The following table depicts the water quality of the sites considered for a typical year by taking the average of the indices for each month from 1991 to 2013. The symbols mentioned in the Index table have been deciphered below it. The water quality varies from Class A (Ex) to Class C (F) for the river basin. According to the water quality criteria followed, the level of quality should always be excellent. However it is seen to be degrading in sites 2, 3 4 and 7. This can be accounted for the anomalous values of the significant variables deduced from the T-Test.

Table 5.8: Conventional Indices for a Typical Year

Sl.No./Months	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
S-1	Ex	Ex	Ex	G	G	Ex	Ex	Ex	Ex	Ex	Ex	Ex
S-2	G	Fair	G	G	G	Ex	Ex	Ex	Ex	Ex	Ex	Ex
S-3	G	G	Ex	G	G	Ex	Ex	Ex	Ex	Ex	G	G
S-4	G	Fair	G	G	Fair	G	Ex	Ex	G	G	G	G
S-5	Ex	G	Ex	G	G	G	Ex	Ex	Ex	Ex	Ex	Ex
S-6	Ex	Ex	G	G	G	Ex	Ex	Ex	Ex	Ex	Ex	Ex
S-7	Ex	Ex	G	G	G	Ex	Ex	Ex	G	G	Ex	Ex

The indices can be deciphered from Table 5.8:

Table 5.9: Different Classes of water with their Notations

Ex=Excellent	$80 < \text{Final Indices} \leq 100$	Class A	Conventional treatment not required; Only disinfection necessary
G=Good	$60 < \text{Final Indices} \leq 80$	Class B	Suitable for Outdoor Bathing
F=Fair	$40 < \text{Final Indices} \leq 60$	Class C	Conventional Treatment and Disinfection mandatory
Poor	$20 < \text{Final Indices} \leq 40$	Class D	Wildlife and Fisheries may be propagated
Bad	$0 < \text{Final Indices} \leq 20$	Class E	Fit for irrigation and coolant purposes

5.4 Water Quality Indexing Based on Functional Data Analysis:

The missing data found are interpolated and retrieved. The measured values of the significant variables are converted into sub-indices using simple MATLAB programs. The conventional sub-indices for a typical year averaged from 1991-2013 for each station are plotted along with the sub-index curves generated using smoothing spline function in MATLAB. For better visualization, the final graphs are plotted in Microsoft Excel. The variables can be visualized with the threshold lines of 20, 40, 60 and 80. If a value lies between 80 and 100, the state can be considered as Excellent (Level I), if between 60 and 80, Good (Level II), between 40 and 60, Fair (III) and below 40, Poor (IV).

5.4.1 Development of Sub-Index Curves

From figure 5.3 shown below, it is seen that, the sub-index values of Phosphate ions and pH remain in Level II throughout the year except in May. The sub-index values of Hardness remain in Level I and deteriorate in May. The smoothing spline fits well with all the sub-indices thus being robust.

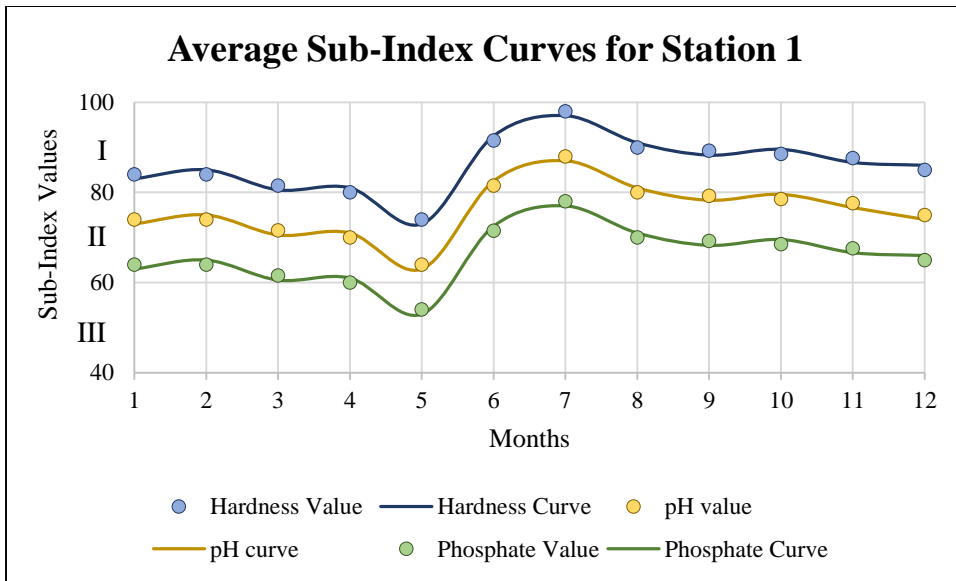


Fig 5.3: Average sub-index curve of Station 1

From figure 5.4 shown below, it is seen that the sub-index values of Coliform Count almost lie in Level III throughout the year. The values of BOD and Nitrate and Total Dissolved Solids lie mostly in zones II and I respectively.

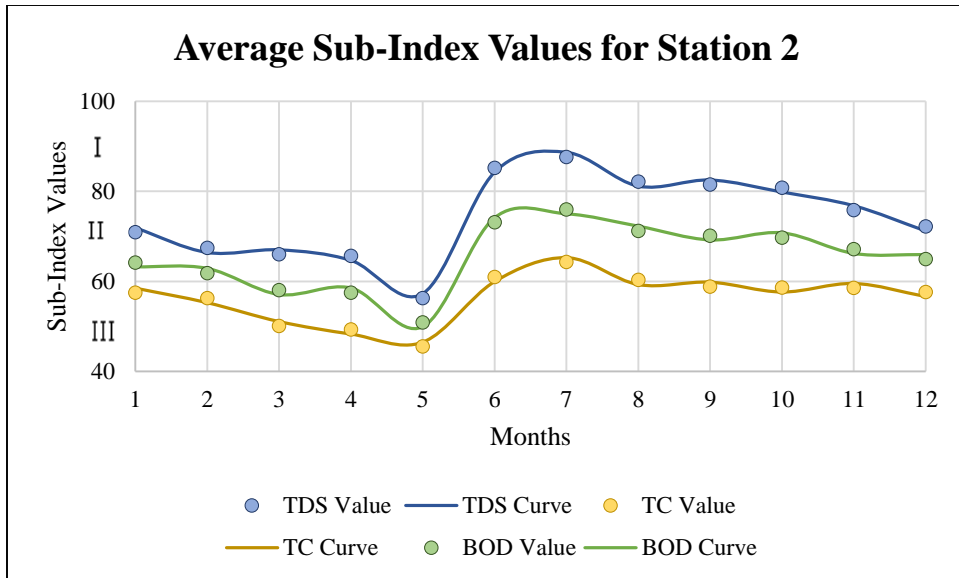


Fig 5.4: Average sub-index curves of Station 2

From figure 5.5 shown below, it is found that the sub-index values of Fluoride Ions and pH fluctuate between Levels II and III. The sub-index values of Ammonia Ions vary from Level I to III throughout.

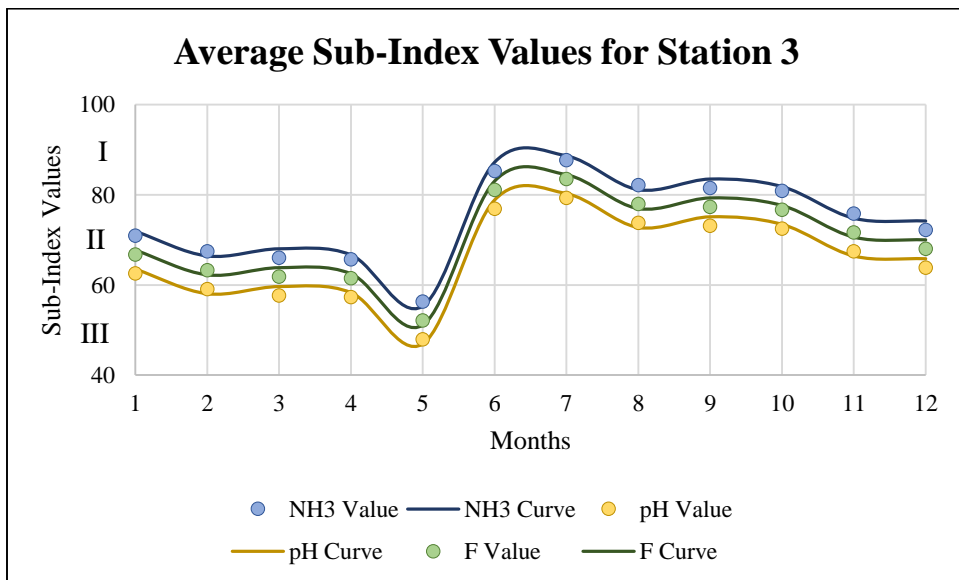


Fig 5.5: Average sub-index curves of Station 3

From the figure 5.6 shown below, it is found that the sub-index values of Coliform Count fall to Level IV from February to May. The sub-index values of BOD and COD vary between zones of II and III. None of the variables are found to be in Level I throughout the year.

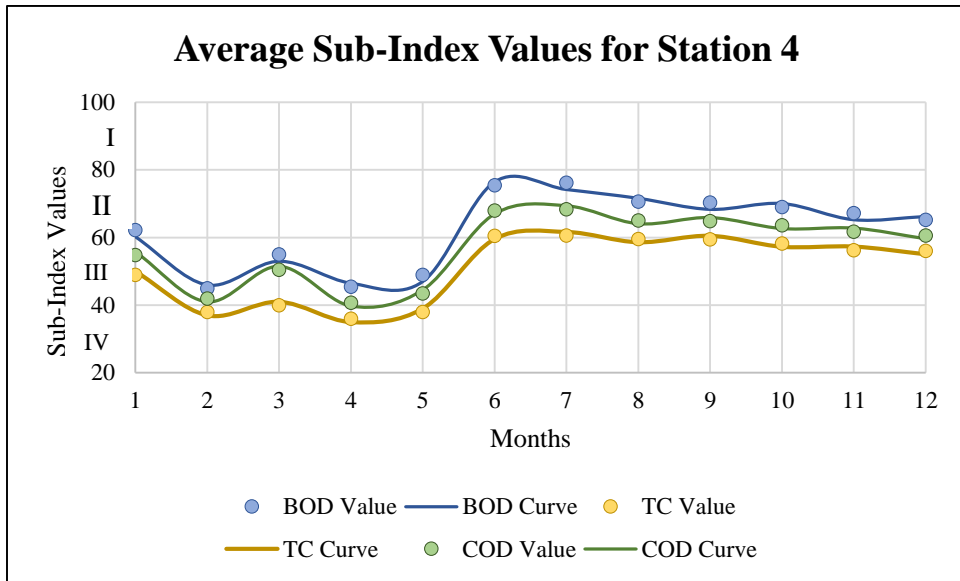


Fig 5.6: Average sub-index curves of Station 4

From figure 5.7 shown below, it is found that, sub-index values Total Dissolved Solids lie in Level II, those of BOD lie in Level III and those of Chloride Ions lie in Level I.

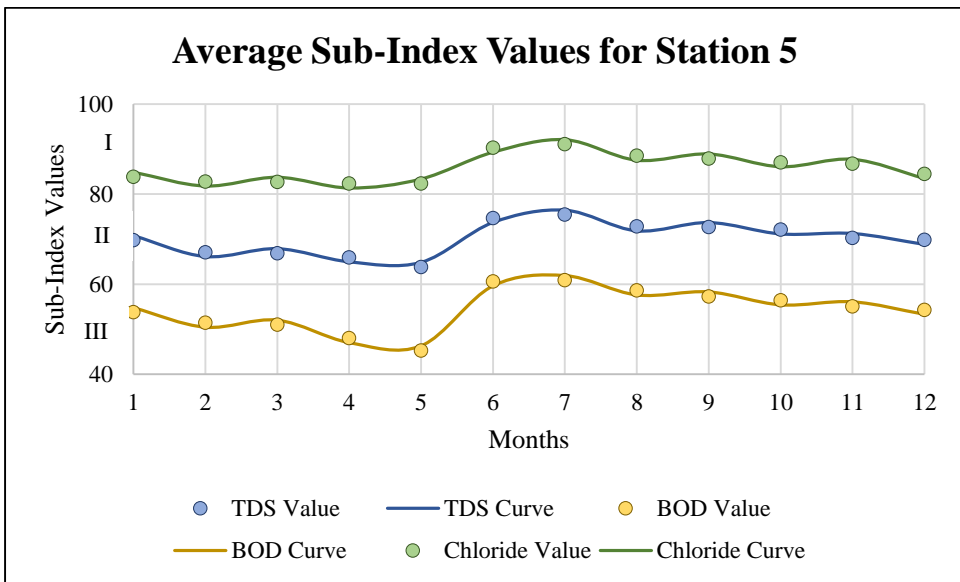


Fig 5.7: Average sub-index curves of Station 5

From figure 5.8 shown above, it is found that, the values of Calcium Major Ions lie almost in Level I, those of pH lie in Level II and those of Fluoride lie in zone I throughout the year.

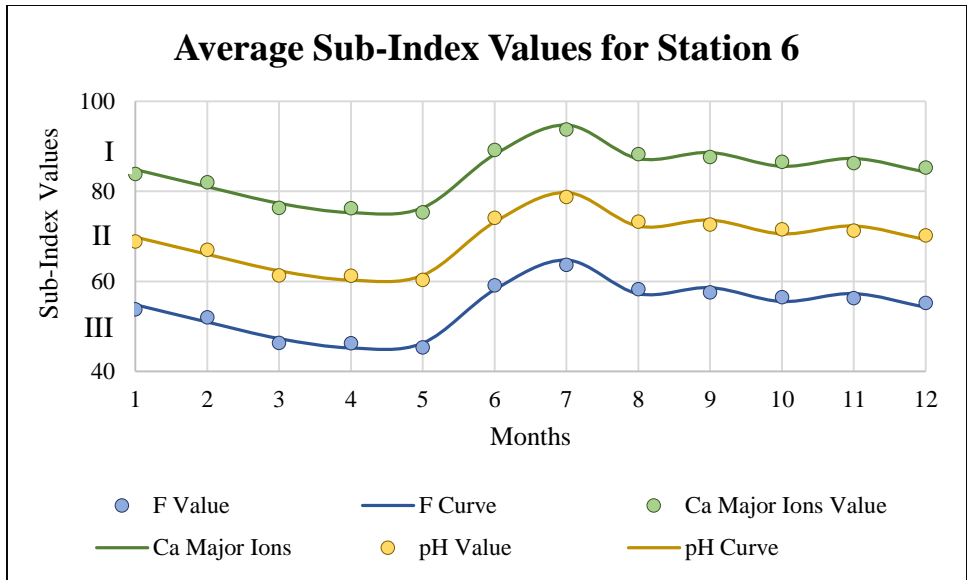


Fig 5.8: Average sub-index curves of Station 6

From the figure 5.9 shown below, it is found that, the sub-index values of BOD lie in zone I and those of Electrical Conductivity lie in level II almost throughout the year. Further the values of Fluoride lie in Level III.

If the sub-index value of a water quality variable is near 80, it can be potentially improved to Level I by subsequent treatment. A similar approach applies to the indices located at the lower levels.

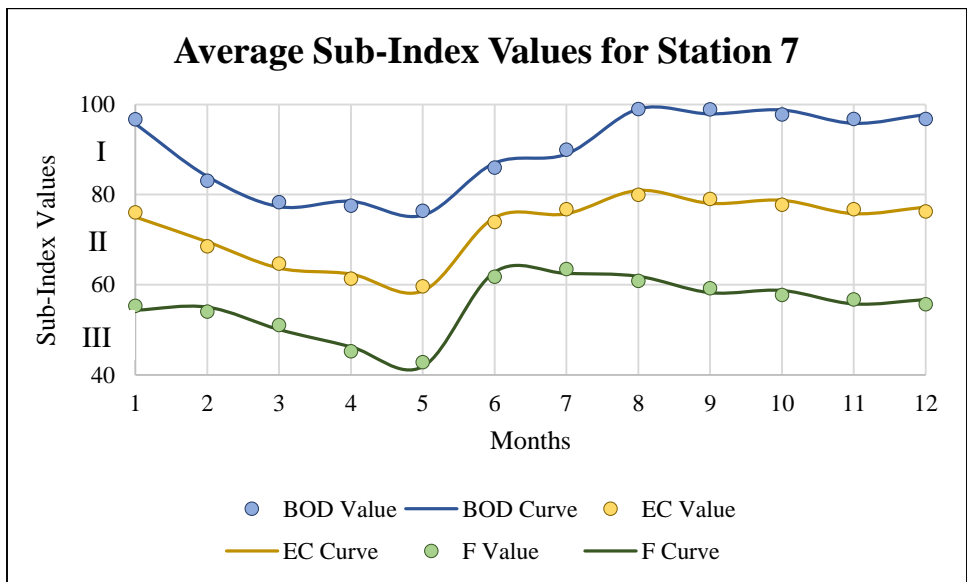


Fig 5.9: Average sub-index curves of Station 7

5.4.2 Assignment of Weights

The weights corresponding to each of the significant variable are determined using three techniques namely, conventional, information entropy method and relative pollution degree. These weights are shown in the form of curves (where applicable).

5.4.2.1 Conventional Weights

Following are the weights assigned to the variables corresponding to all the 7 stations.

Table 5.10: Stationary Weights of Significant Variables for all the Stations

Stations	Parameter 1	Weight	Parameter 2	Weight	Parameter 3	Weight
1	pH	0.36	Hardness	0.40	Phosphate Ions	0.25
2	BOD	0.32	TC	0.24	TDS	0.43
3	Fluoride	0.32	pH	0.24	Ammonia Ions	0.43
4	COD	0.35	BOD	0.37	Coliform Count	0.28
5	Chloride Ions	0.43	TDS	0.29	BOD	0.29
6	F	0.22	Ca-Major Ions	0.40	pH	0.38
7	BOD	0.39	EC	0.37	F	0.24

The variables bearing a smaller weight contribute less to the total water quality index and hence account for more contribution to pollution.

5.4.2.2 Assignment of Weights by Information Entropy Method:

The weights assigned by information entropy method are depicted in the following graphs. As we have seen, the entropy weight is directly proportional to the extent of useful information obtained from a system. The following conclusions are deduced from the entropy weight curves for a typical year.

That, the weights of pH lie below those of Hardness and Phosphate, state that pH accounts for more entropy or more uncertainties as shown in Figure 5.10. The decrease of the weights during the month of May indicate that the quality is more prone to disturbances during this month. The reasons may be natural or anthropogenic or measurement uncertainties.

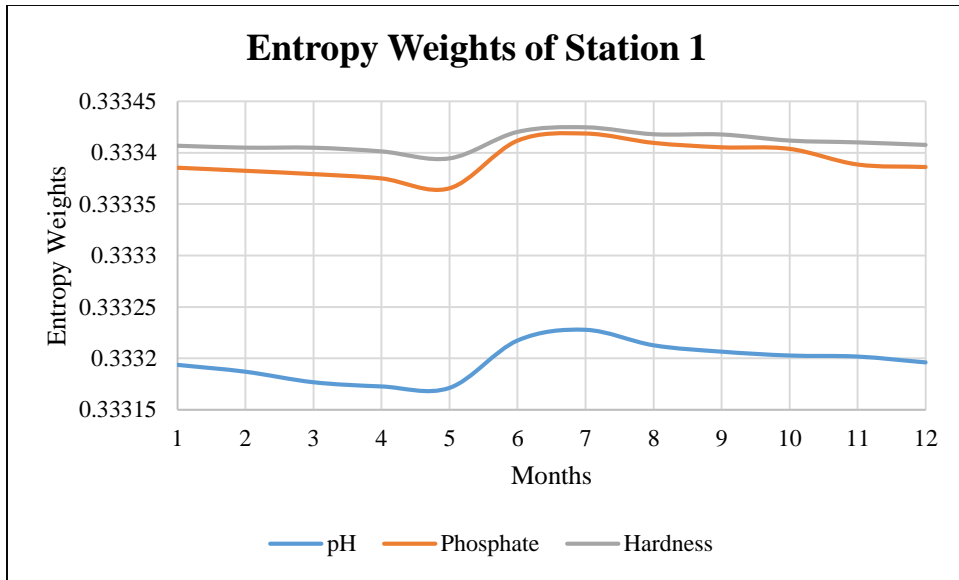


Fig 5.10: Entropy Weight Curves for Station 1

From Figure 5.11, it is seen that, Total Dissolved Solids account for more uncertainties than the other two variables. Further it has more uncertainty during the February and May, perhaps because of anthropogenic and natural conditions respectively. The same reason can be considered valid for the deviation of the weights for the other two variables.

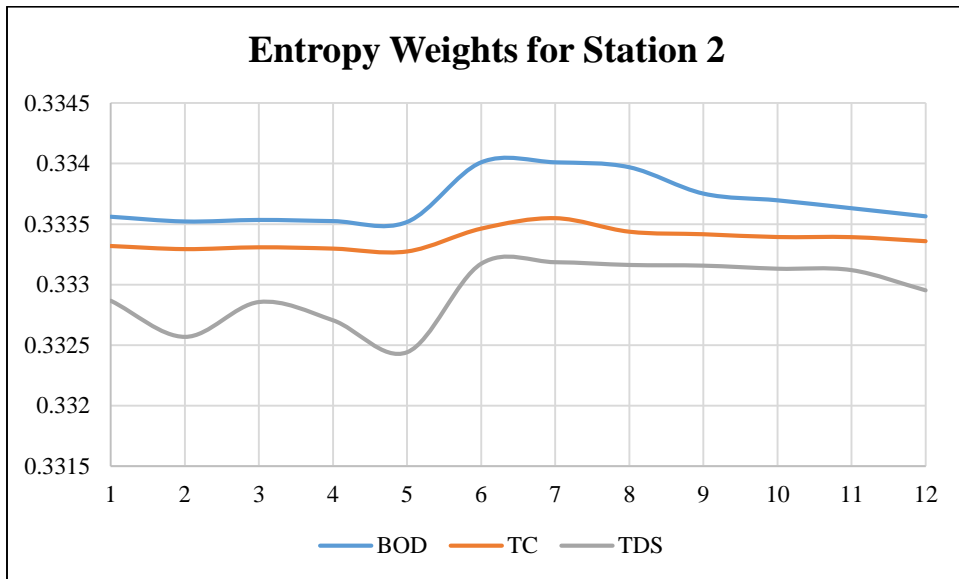


Fig 5.11: Entropy Weight Curves for Station 2

Out of the three variables considered for Station 3, from figure 5.12, we find that, pH is most vulnerable to uncertainties. A decrease in the weights of the variables during May is can be because of the natural causes.

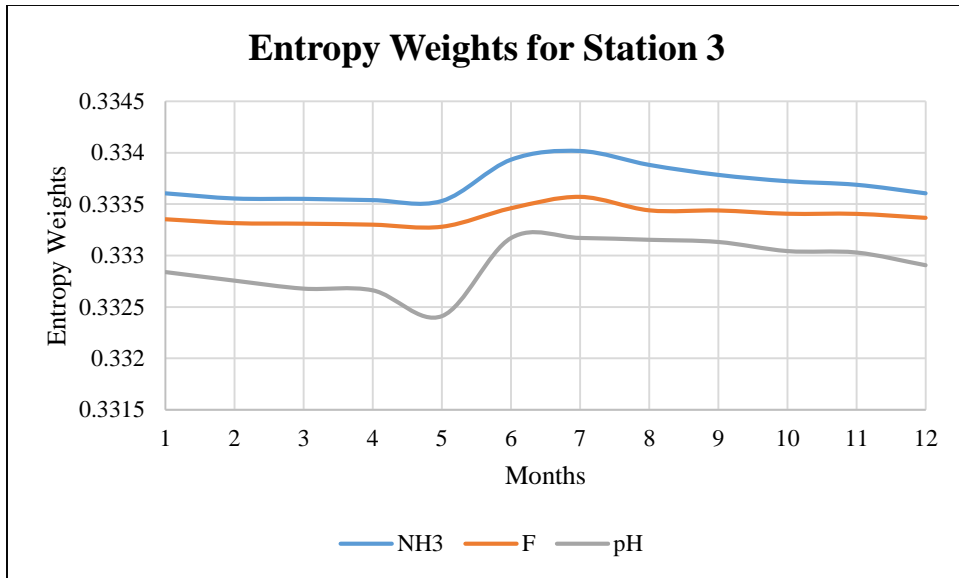


Fig 5.12: Entropy Weight Curves for Station 3

From figure 5.13, it is found that, out of the three variables, BOD shows maximum entropy. There is a kink in the curves during the months of February and March may be, because of anthropogenic and natural causes respectively.

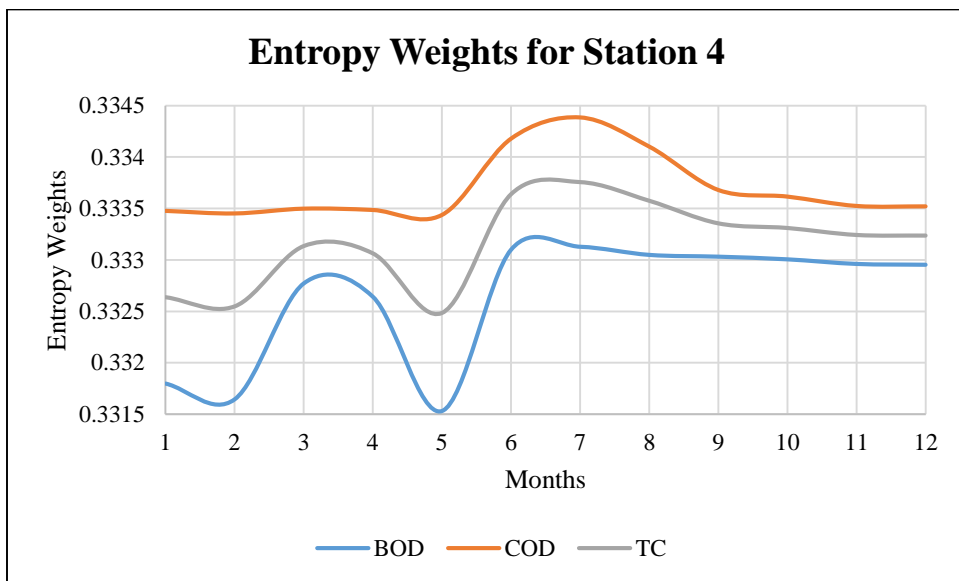


Fig 5.13: Entropy Weight Curves for Station 4

From figure 5.14, it is deduced that BOD shows maximum uncertainties of the three variables. The Chloride Ions and BOD show a decrease in weights during May due to the above mentioned reason.

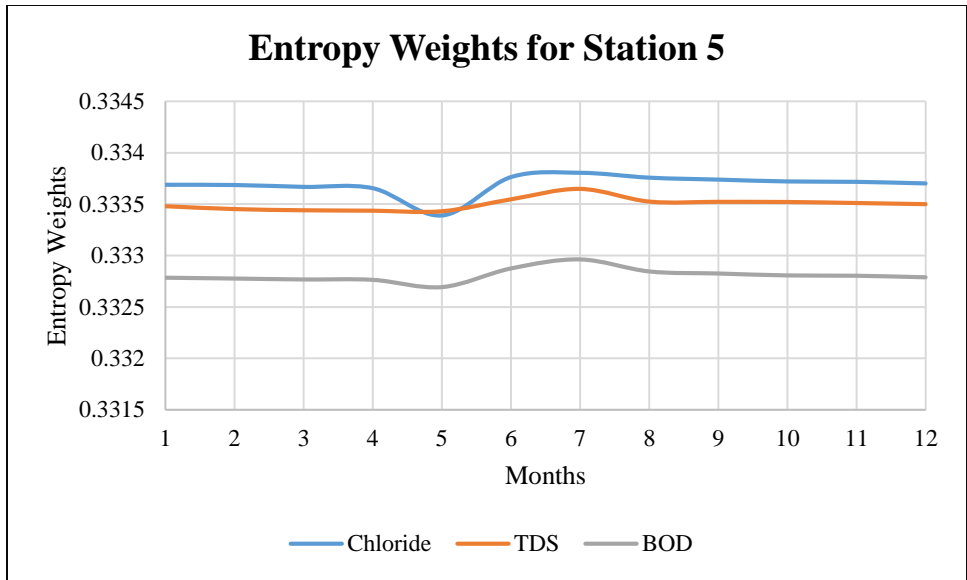


Fig 5.14: Entropy Weight Curves for Station 5

From figure 5.15, we see that, the weights of the variables of this station remain almost constant throughout the year, thus reducing the vulnerability of the water quality to the uncertainties. This reduction can be accounted for the presence of a dam near the site which changes the water quality status. pH is most vulnerable out of the three.

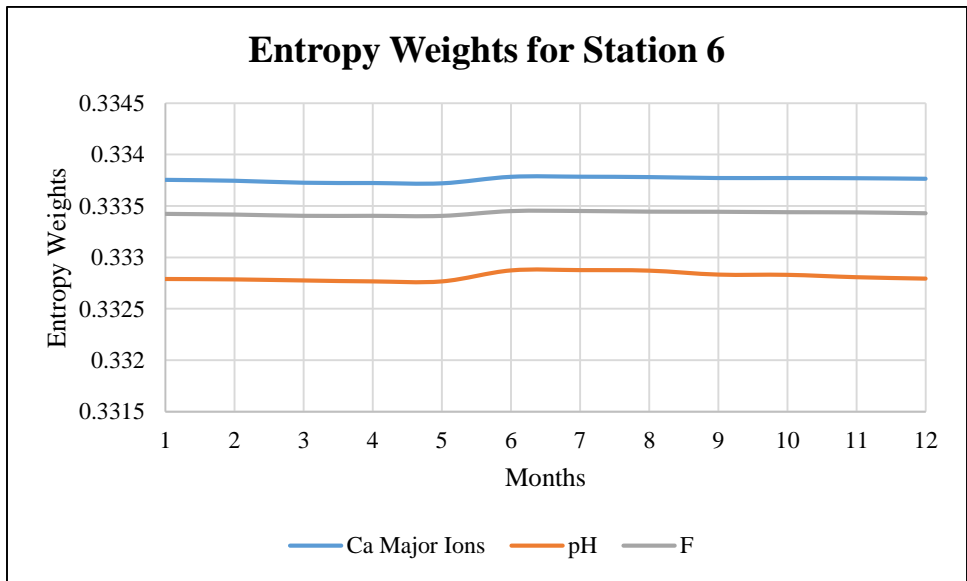


Fig 5.15: Entropy Weight Curves for Station 6

From figure 5.16, we find that, in station 7 also, the variables remain almost constant, EC being the most vulnerable one. The decrease of weight in May again can be accounted because of natural causes.

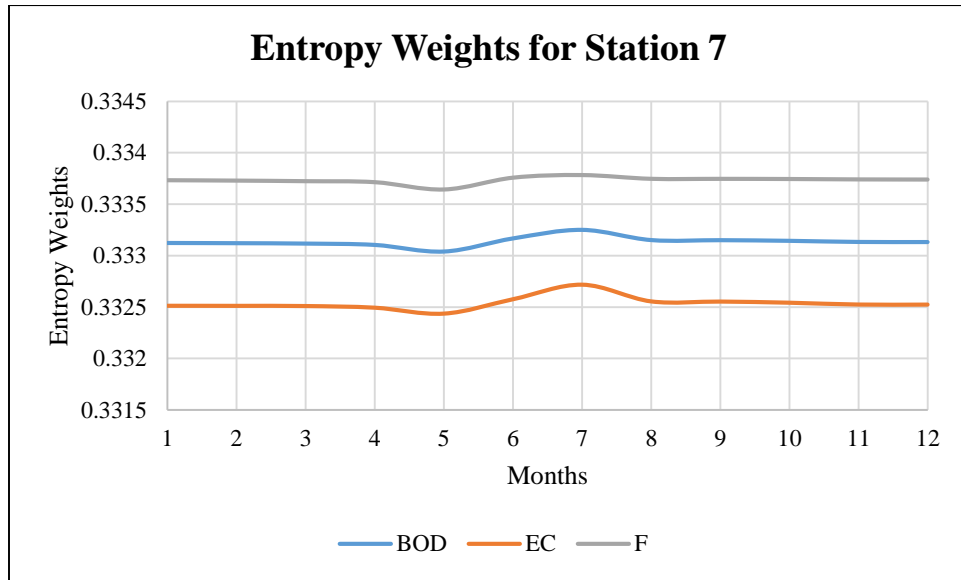


Fig 5.16: Entropy Weight Curves for Station 7

5.4.2.3 Assignment of Weights by Relative Pollution Degree

The weights obtained from the method of Relative Pollution Degree are averaged and plotted for a typical year. Lesser the relative pollution weight, less is the contribution of the corresponding variable to the pollution. All the results tally with the sub-index curves.

The relative pollution degree curves for Station 1 are shown in figure 5.17. All the variables account for minimal pollution during the month of May. Out of the three variables, Phosphate contributes the least to the pollution.

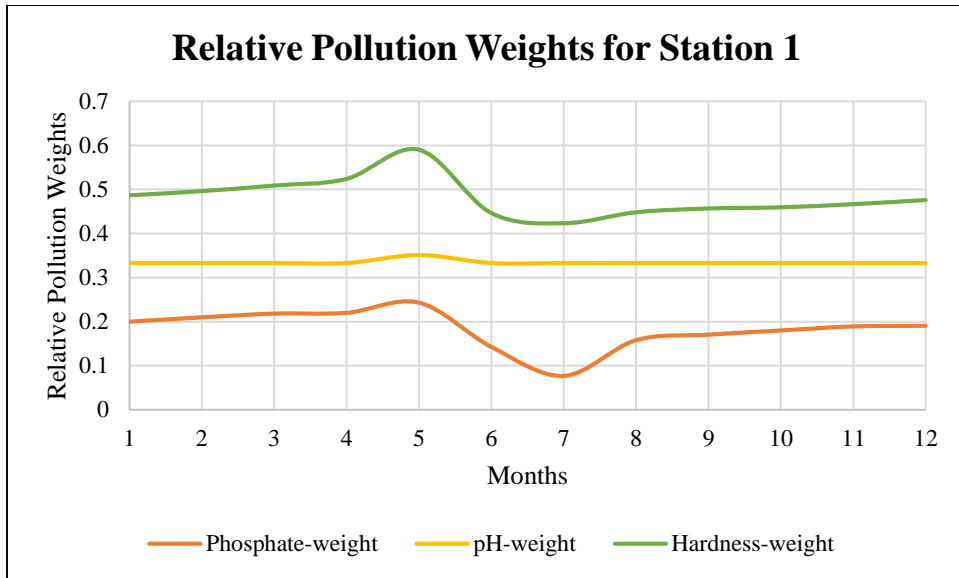


Fig 5.17: Relative Pollution Weight Curves for Station 1

From the figure 5.18 shown below, TDS accounts for minimum pollution and all the variables contribute less to pollution during May.

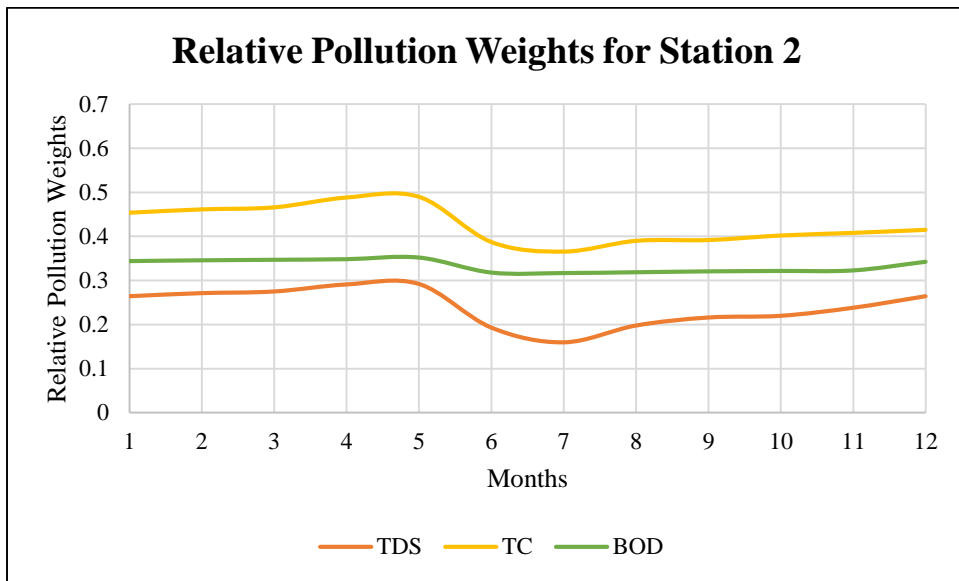


Fig 5.18: Relative Pollution Weight Curves for Station 2

From the figure 5.19 shown below, we find that the curve for Ammonia lies below the curves of the other two. Thus pH accounts for more pollution than Ammonia and Fluoride. The weights of Fluoride show that its concentration remains almost constant throughout the year.

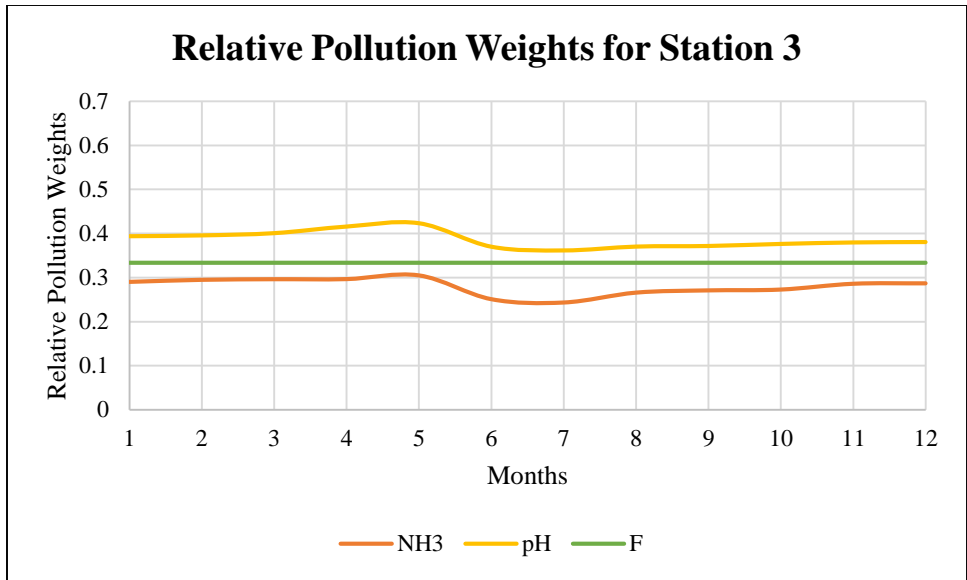


Fig 5.19: Relative Pollution Weight Curves for Station 3

The Relative Pollution Degree curves for Station 4 are shown in figure 5.20. Out of the three variables, COD accounts least for pollution. The weight of TC remains almost constant throughout the year.

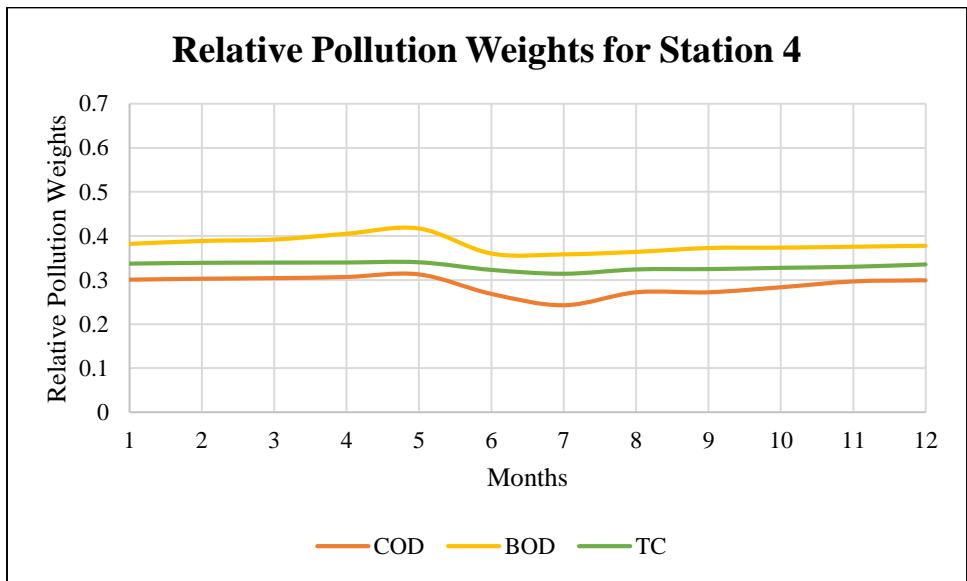


Fig 5.20: Relative Pollution Weight Curves for Station 4

From the figure 5.21 shown below, it is found that, the curve for TDS lies continuously below the other two variables throughout the year thus indicating less contribution to pollution.

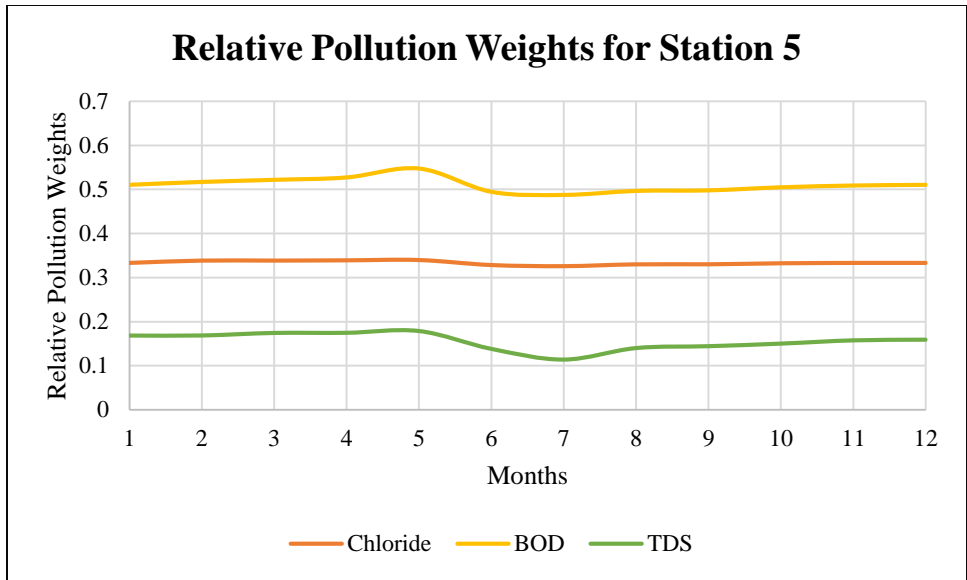


Fig 5.21: Relative Pollution Weight Curves for Station 5

The Relative Pollution Degree curves for Station 6 are shown in figure 5.22. The weight of Calcium Major Ions remains constant throughout the year. All the variables show a decrease in pollution during the month of July.

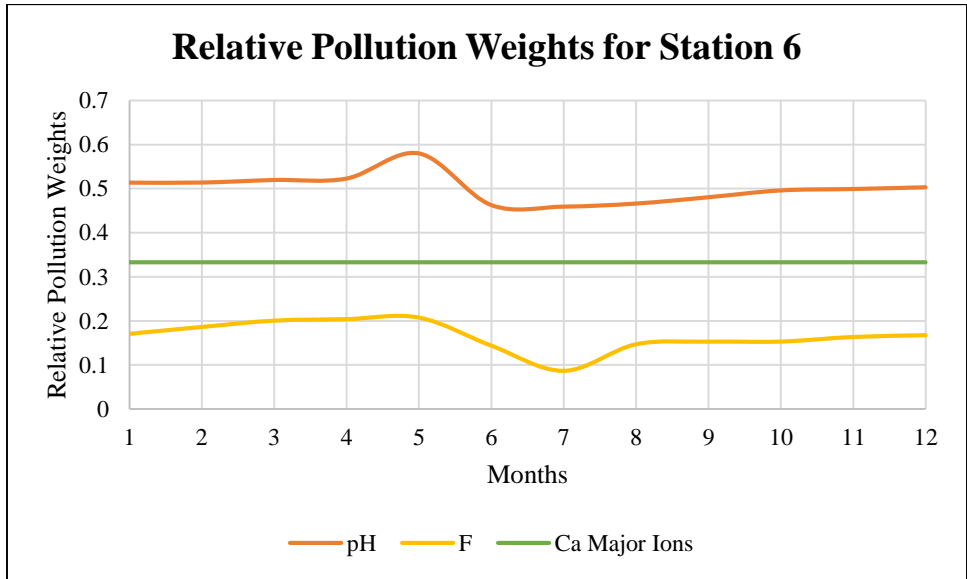


Fig 5.22: Relative Pollution Weight Curves for Station 6

The Relative Pollution Degree curves for Station 6 are shown in figure 5.23. BOD is accounts least for pollution while Fluoride accounts for most. The weight of EC remains constant throughout the year.

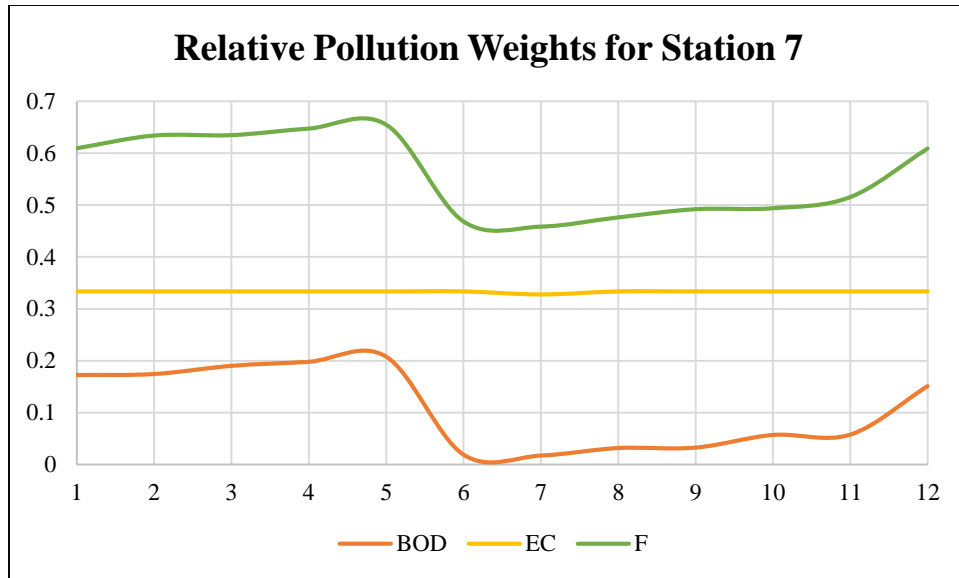


Fig 5.23: Relative Pollution Weight Curves for Station 7

5.4.2.4 Development of Water Quality Index Curves:

The final water quality index curves are generated using weighted sum model using conventional, entropy and relative pollution weights. In the following figures, W.S stands for Weighted Sum, C stands for Conventional Weighting, E stands for Entropy Weighting and R.P stands for Relative Pollution Weighting. The levels I, II, III and IV shown alongside the graphs depict the quality of water as per the sub-index curves. An overall visualization tells us that relative Pollution overestimates and Entropy Weight Method underestimates the water quality as compared to the conventional method.

The Water Quality Index curves for Station 1 are shown in figure 5.24. The curves show a decreasing trend from January to May and September to December. The quality develops in the month of June and July. The decreasing trend may be because of the pH according to the conventional and entropy weight approaches. The recovery of the trend may be accounted for because of phosphate with tends to balance the pH. The water quality at Station 1 remains in Level 1 almost every time throughout the year. Hence sophisticated water quality treatment may not be needed. The water quality index curve due to relative pollution is higher because the weights assigned to the indicators are higher.

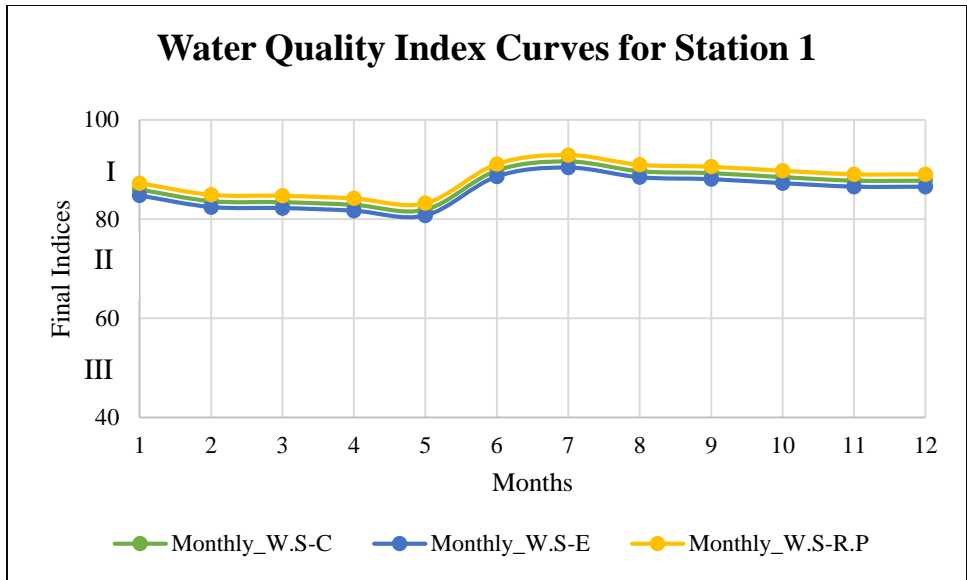


Fig 5.24: Water Quality Index Curves for Station 1

The Water Quality Index curves for Station 2 are shown in figure 5.25. Here also, the final indices decrease during the month of May. The decreasing trend may be due to increase in Coliform Count and BOD according to conventional and entropy weights. The curves revive during the months of June and July because of decrease in TC and BOD according to relative pollution weights. The level of water in this station can be classified into Level II, taking into consideration all the three approaches. Thus proper water treatment is required before its consumption at this station.

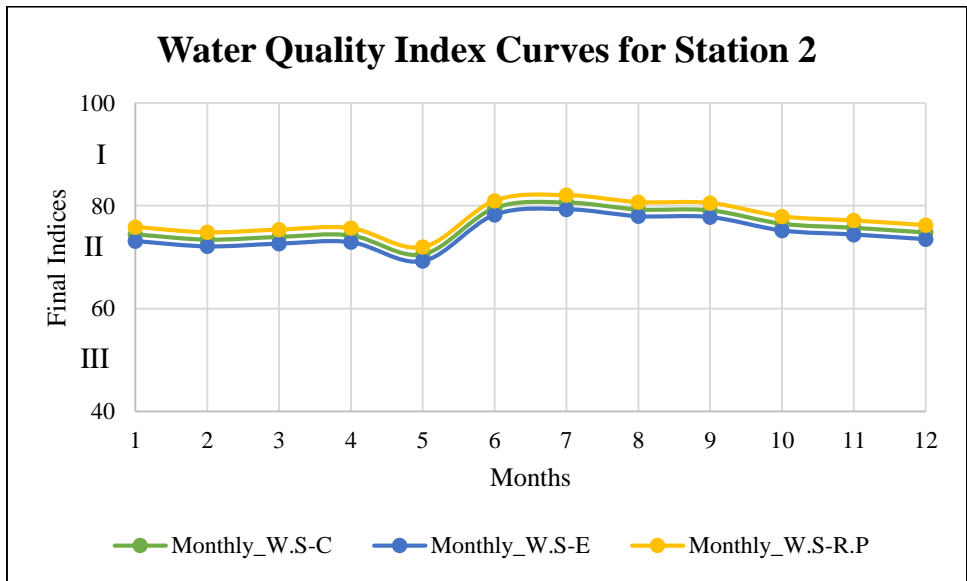


Fig 5.25: Water Quality Index Curves for Station 2

The Water Quality Index curves for Station 3 are shown in figure 5.26. The overall trend of water quality decreases from March to May. This decrease may be because of pH which bears a lower weightage in Conventional and Entropy weight approaches. The water quality improves during the months of June, July and August because of natural conditions and also perhaps because of a balance of pH during these months. The overall water quality lies in Level I and Level II for half of a typical year. Hence, treatment is necessary before consumption.

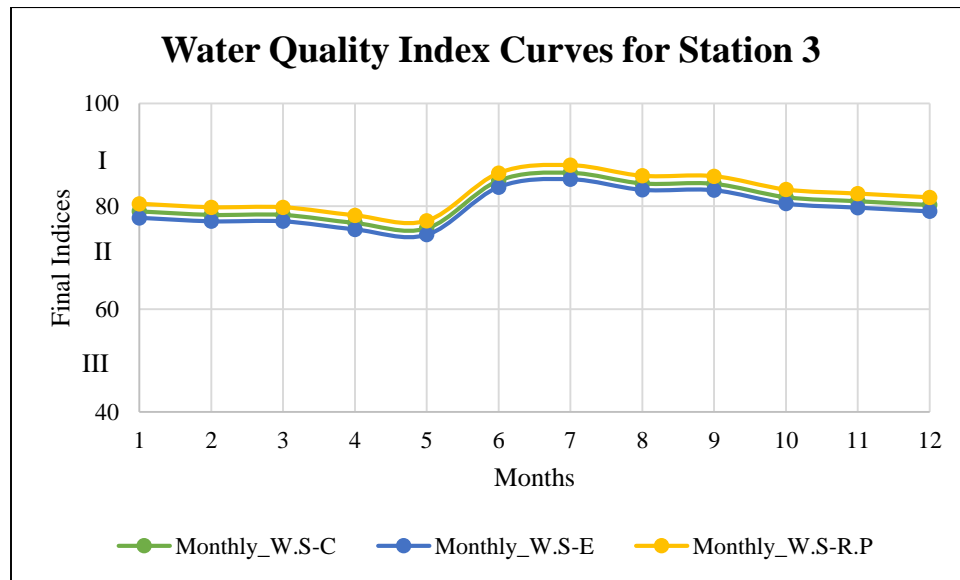


Fig 5.26: Water Quality Index Curves for Station 3

In station 4, as shown in Figure 5.27, the decrease in trends of water quality during the month of February and May, may be because of TC and BOD according to conventional and entropy weight approaches. The water quality revives because of the decrease in BOD and TC count. The water level at this station can be considered to be in Level II throughout the year and hence proper sophisticated treatment is mandatory.

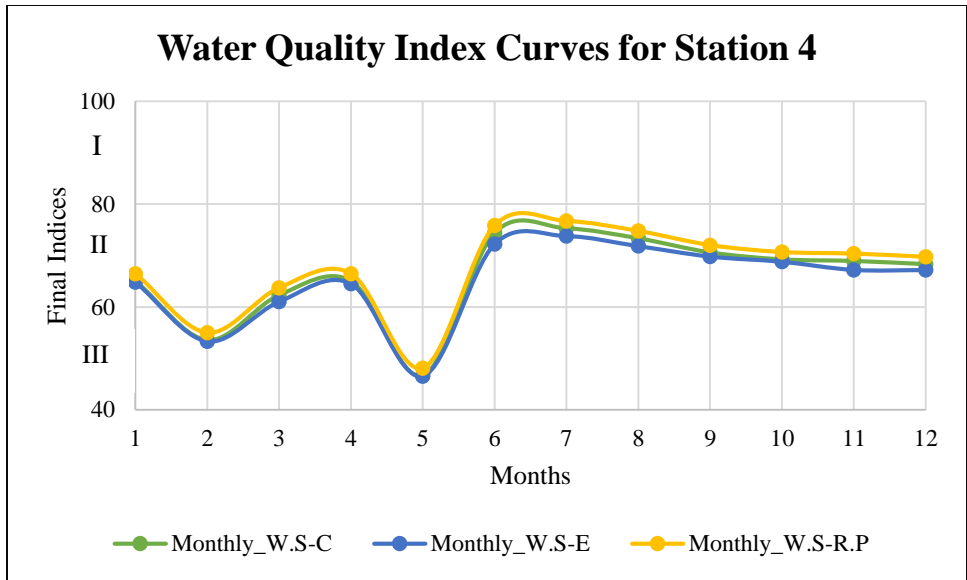


Fig 5.27: Water Quality Index Curves for Station 4

The Water Quality Index curves for Station 5 are shown in figure 5.28. The water quality index almost remains constant at this site, thus showing negligible trend throughout the year. The routinely decreasing trend during the month of May is pronounced. This may be due to an increase in BOD during this time as per the conventional and entropy weight approaches. The water quality revives during the month of June and July perhaps because of the decrease in BOD according to the relative pollution weight system. The overall water quality of the station can be categorized as being under Level I. However, primary treatment may be essential during non-monsoon seasons.

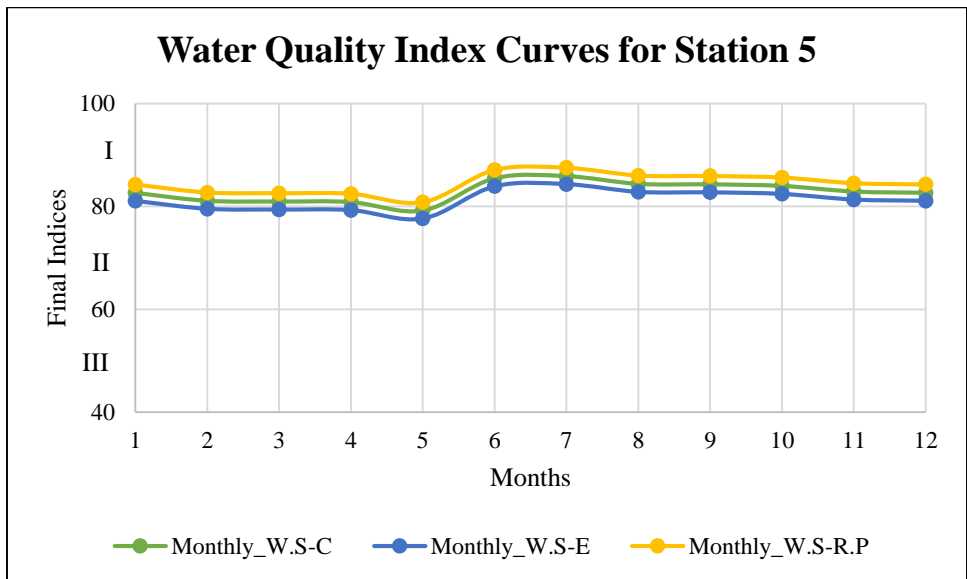


Fig 5.28: Water Quality Index Curves for Station 5

The Water Quality Index curves for Station 6 are shown in figure 5.29. Here, the decreasing trend from January to May can be due to low entropy and conventional weights of pH. The water quality improves during the months of June and July majorly because of adjustment of pH according to the relative pollution weight system. The overall water quality of this station can be categorized under Level I indicating that the water remains slightly polluted here. Hence just primary treatment is required. Other treatment processes may be introduced as per the requirement.

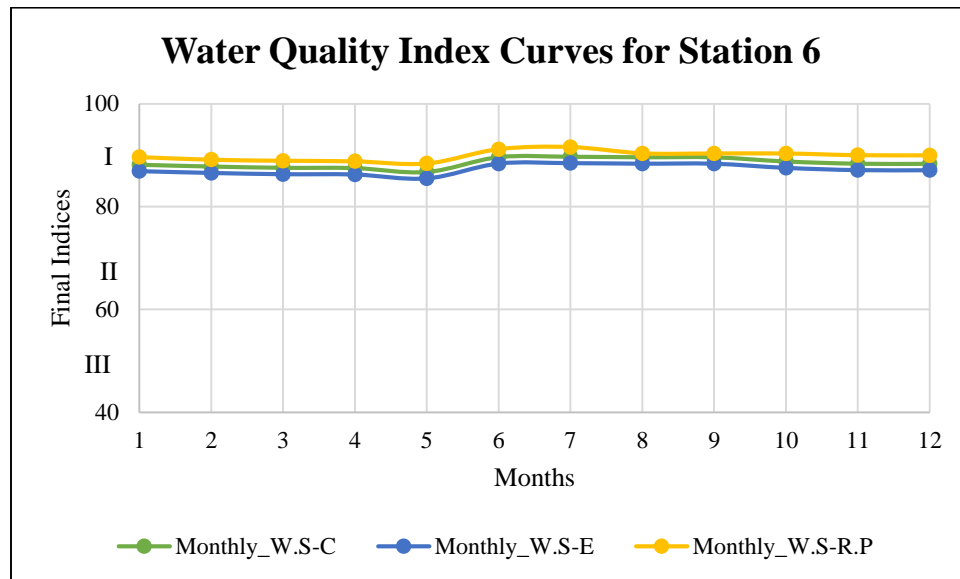


Fig 5.29: Water Quality Index Curves for Station 6

The Water Quality Index curves for Station 7 are shown in figure 5.30. The water quality decreases during the May because of a lower weight of EC as per conventional and entropy weight systems. The water quality revives due to balancing of Fluoride Ions and EC as per the relative pollution system. The water quality lies in Levels I and II indicating that proper treatment is necessary before its use.

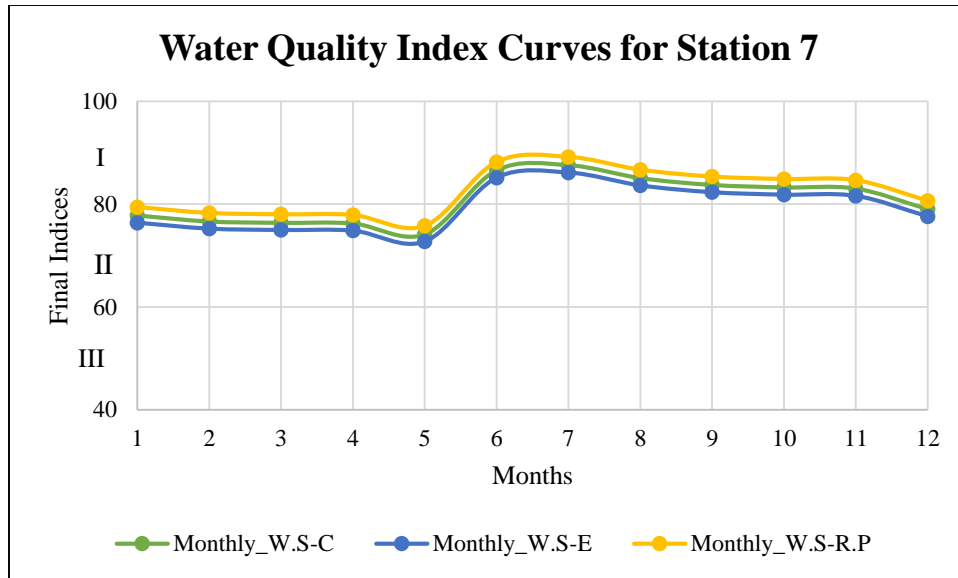


Fig 5.30: Water Quality Index Curves for Station 7

5.5 Water Quality Index Based on Principal Component Analysis

Principal Component is carried out in each station and the results resemble those obtained from t-test as mentioned before. The variable loadings of the first component and variance are put in the formula mentioned in Chapter 5 and the following graphs of Water Quality Indices are formulated. The results from incorporation of Principal Component Analysis to determine Water Quality Indices are almost the same as those obtained from conventional method, thus proving the robustness of the method.

The Water Quality Index curves derived from Conventional method and Principal Component Analysis for Station 1 are shown in figure 5.31. The results from Principal Component Analysis overestimate the water quality as compared to that computed using conventional indexing. In both the cases the high loading of hardness increases the water quality. In conventional indexing, phosphate bears a lesser weight while in the other method, pH bears a lower loading. The level of water quality can be considered to lie in Level I throughout the year. Hence preliminary treatment is required before its consumption.

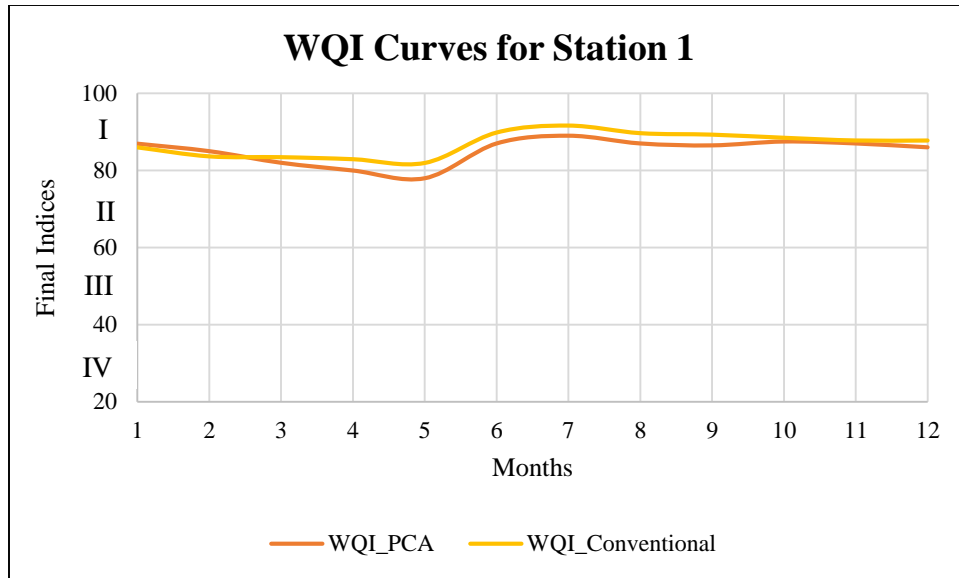


Fig 5.31: Water Quality Index Curves for Station 1 using PCA

The Water Quality Index curves derived from Conventional method and Principal Component Analysis for Station 2 are shown in figure 5.32. The results obtained from Principal Component Analysis overestimate the water quality from March to July, otherwise they entwine around the results from conventional indexing. According to Conventional Indexing, Coliform Count accounts for a lesser weight and hence decreases the water quality during the month of May. However TDS bears a lower loading in the first component. The level of water quality in this station fluctuates between Level I and II. Hence, proper treatment of water is important before consumption.

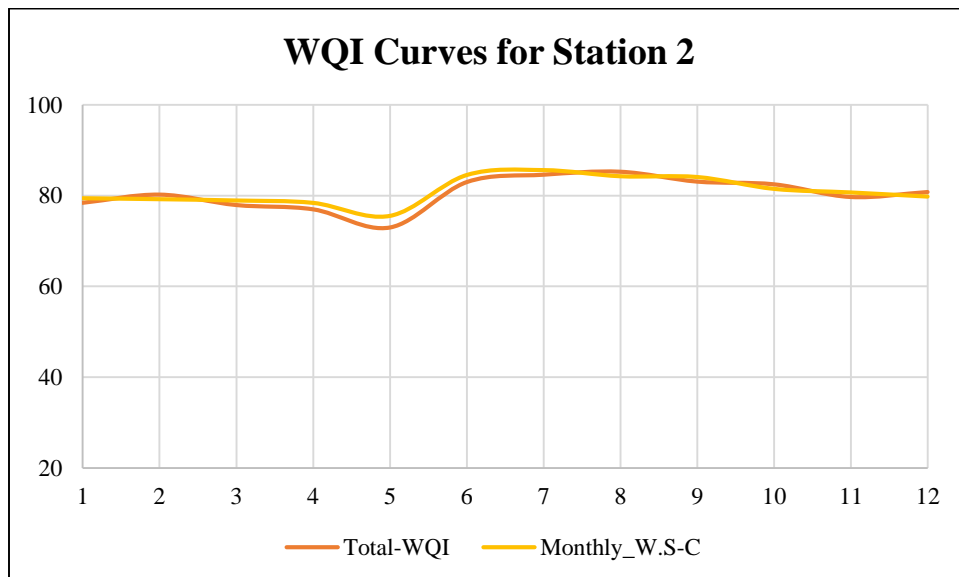


Fig 5.32: Water Quality Index Curves for Station 2 using PCA

The Water Quality Index curves derived from Conventional method and Principal Component Analysis for Station 3 are shown in figure 5.33. The results from Principal Component Analysis over-estimate and under-estimate the water quality as compared to conventional indices. The decrease in water quality is because of pH according to both the methods. The water quality level fluctuates between I and II in this station. Hence water treatment is necessary before consumption.

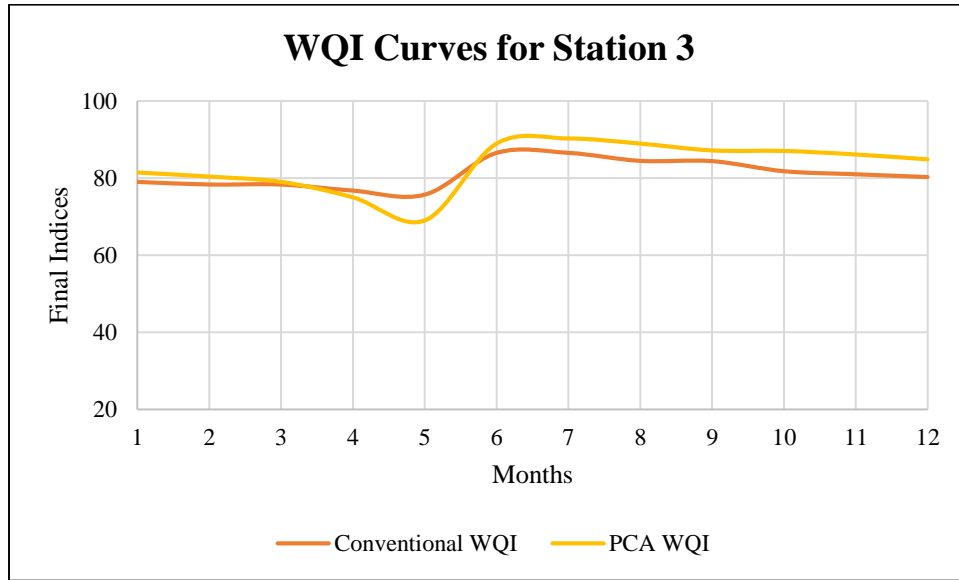


Fig 5.33: Water Quality Index Curves for Station 3 using PCA

The Water Quality Index curves derived from Conventional method and Principal Component Analysis for Station 4 are shown in figure 5.34. The water quality deteriorates due to Coliform Count as per conventional indexing and Principal Component Analysis. The results from Principal Component Analysis sometimes overestimate and sometimes underestimate those obtained from conventional indexing. The small kink in the curves during the month of February indicate the decline of water quality which becomes more pronounced during the dry season. The level of water quality drops to III during May. The water at this station needs sophisticated treatment before consumption.

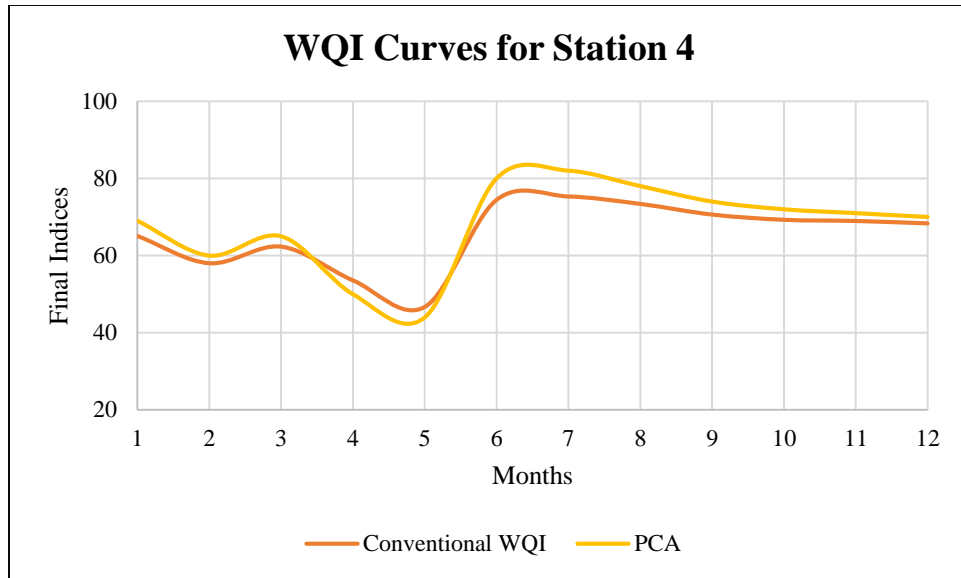


Fig 5.34: Water Quality Index Curves for Station 4 using PCA

The Water Quality Index curves derived from Conventional method and Principal Component Analysis for Station 5 are shown in figure 5.35. The results from Principal Component Analysis overestimate the water quality from January to mid-March and July to September and underestimate the quality from mid-March to July and from October and December. The decrease in water quality is because of BOD in conventional indexing while it is because of TDS according to indexing by Principal Component Analysis. The level of water is almost categorized under Level I and hence needs primary treatment before consumption.

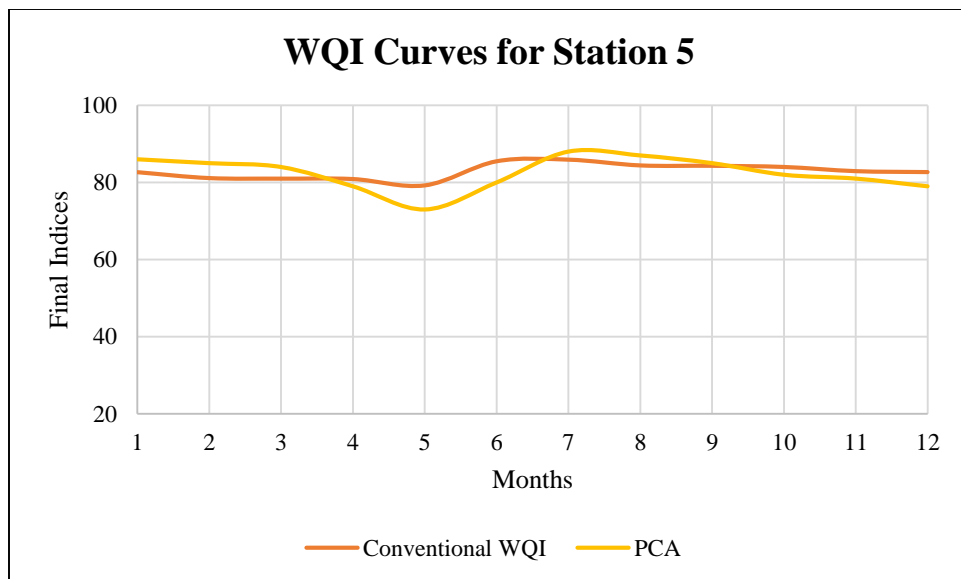


Fig 5.35: Water Quality Index Curves for Station 5 using PCA

The Water Quality Index curves derived from Conventional method and Principal Component Analysis for Station 6 are shown in figure 5.36. The results from Principal Component Analysis underestimate the water quality from January to mid-May and overestimate the quality for the rest of the year. The decline in water quality is because of lower weightage of pH in both conventional indexing and Principal Component Analysis. The level of water quality remains in Level I throughout the year. Hence, primary treatment may be done for water purification.

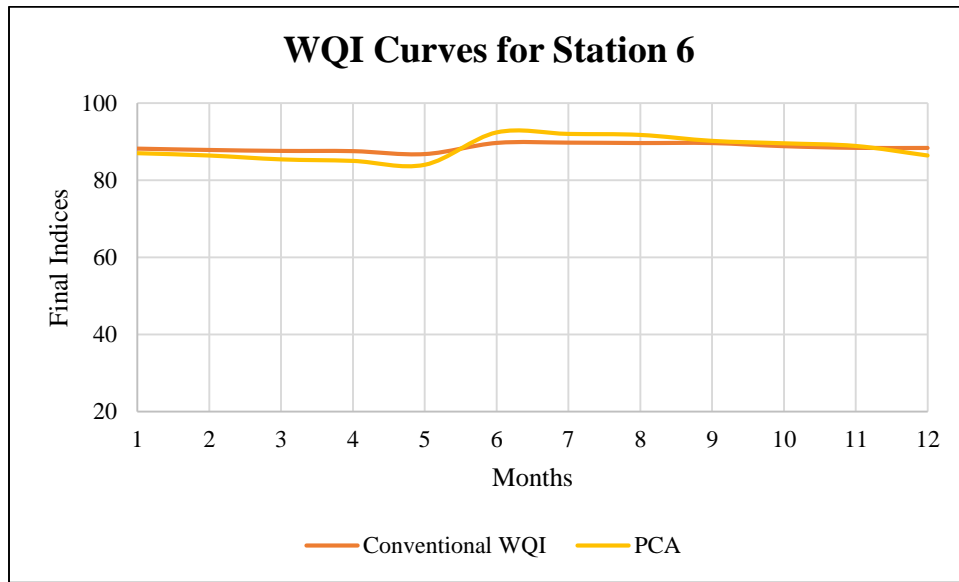


Fig 5.36: Water Quality Index Curves for Station 6 using PCA

The Water Quality Index curves derived from Conventional method and Principal Component Analysis for Station 7 are shown in figure 5.37. The results obtained from Principal Component Analysis underestimate water quality almost throughout the year as compared to those from conventional indexing. The decline in water quality is due to the concentration of Fluoride Ions according to conventional indexing and due to Electrical Conductivity according to Principal Component Analysis. The level of water remains in Level I throughout the year and hence preliminary treatment is required before its consumption.

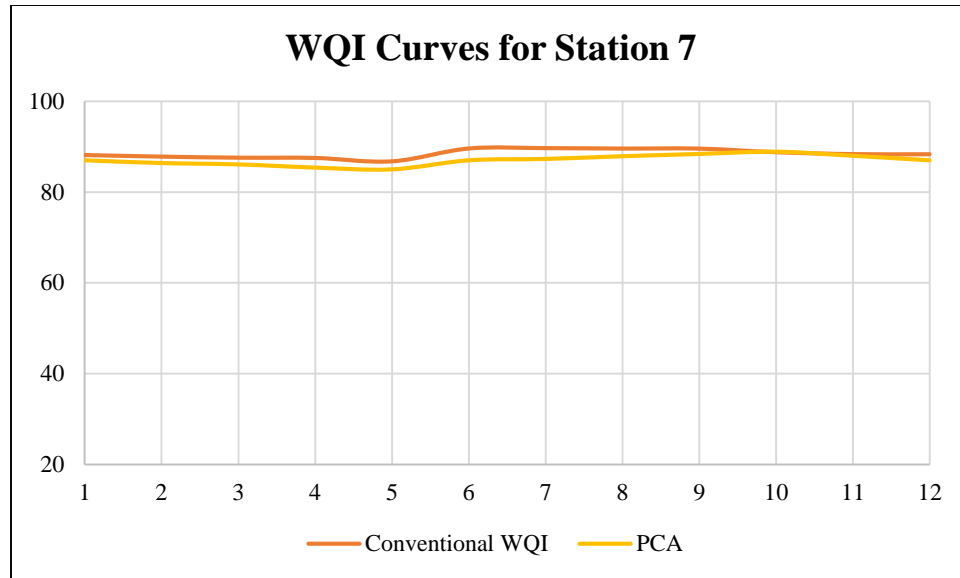


Fig 5.37: Water Quality Index Curves for Station 7 using PCA

5.6 Error Analysis

The error analysis of the Water Quality Index Models is done using Root Mean Square Error, Mean Absolute Error and Mean Absolute Percentage Error and the best model is selected. Table 5.10 shows the calculated errors. It is found that entropy weight method is closest to the conventional indexing for all the stations. Hence, in the final indexing entropy weight method is adopted.

Table 5.11: Error Analysis of the WQI Methods

Station	WQI using Entropy Weight		WQI using Relative Pollution		WQI using Principal Components	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
1	0.093	0.012	0.115	0.013	0.101	0.014
		14%	15%		17%	
2	0.085	0.013	0.118	0.014	0.217	0.045
		17%	15%		17%	
3	0.100	0.013	0.133	0.015	0.225	0.036
		16%	15%		17%	
4	0.203	0.010	0.248	0.014	0.337	0.012
		14%	15%		17%	
5	0.065	0.016	0.076	0.016	0.118	0.002

	MAPE	19%		MAPE	15%		MAPE	17%
6	RMSE	0.037		RMSE	0.060		RMSE	0.088
	MAE	0.013		MAE	0.014		MAE	0.002
	MAPE	14%		MAPE	15%		MAPE	17%
7	RMSE	0.141		RMSE	0.172		RMSE	0.320
	MAE	0.014		MAE	0.016		MAE	0.063
	MAPE	18%		MAPE	15%		MAPE	17%

5.7 Representation of Final Water Quality Indexing

The minimum water quality indices based on entropy weighting method are illustrated in Fig. 5.38. The figure shows that at Station 4, the water quality is the lowest and it tends to rejuvenate as the river approaches its mouth.

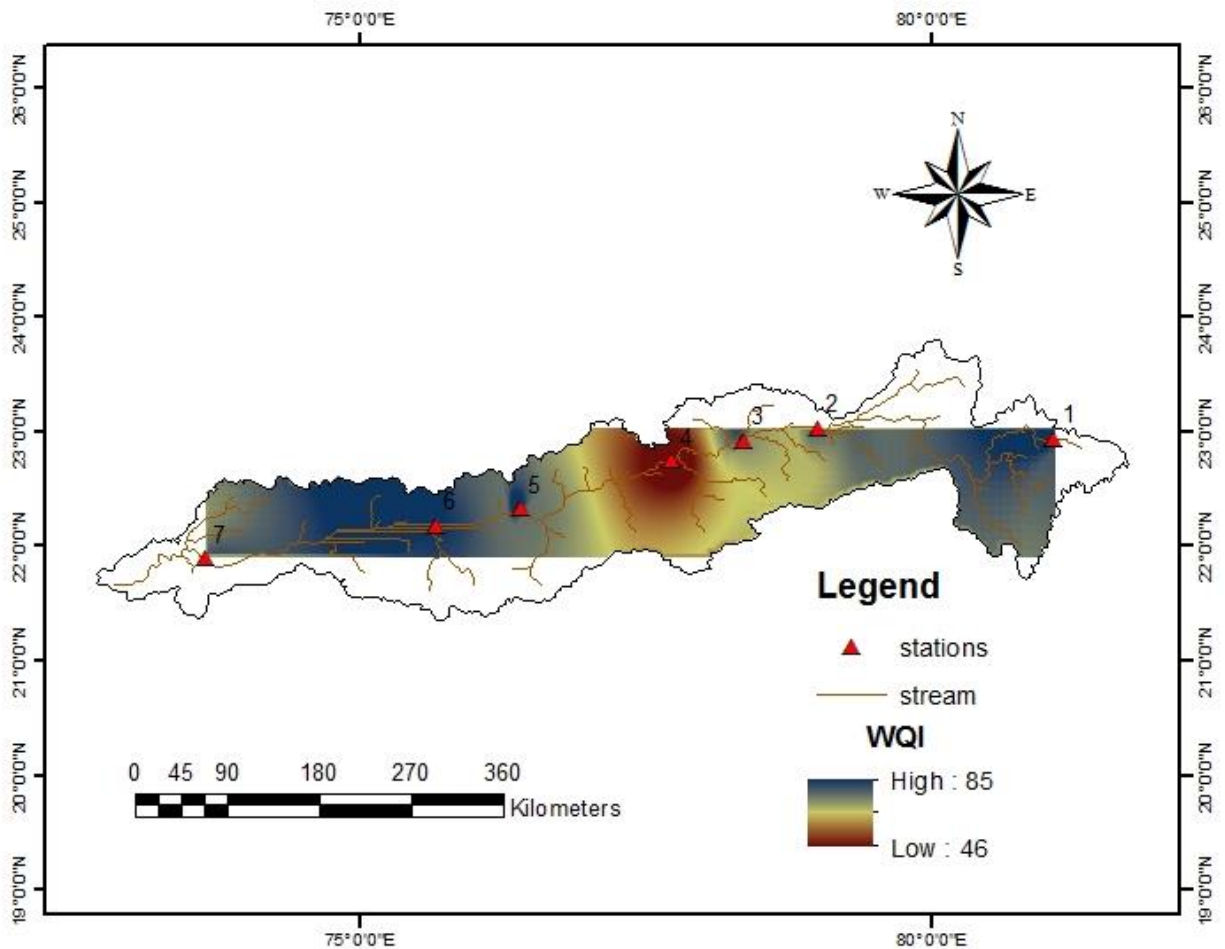


Fig 5.38 Minimum Water Quality Indices for Narmada River Basin for a typical year

CHAPTER VI

CONCLUSIONS

Thus the water quality analysis of River Narmada for a typical year is done using some preliminary methods, Functional Data Analysis and Principal Component Analysis and error analysis of the methods used is computed. In the preliminary analysis, descriptive statistics, hierarchical clustering, principal component analysis and principal factor analysis are done. Water Quality Indexing based on Functional Data Analysis is done by employing entropy and relative pollution degree weights. In case of Principal Component Analysis weights are calculated using the loadings of the first component. The weights are aggregated into Final Indices using Weighted Sum Model. The error analysis is done using Root Mean Square Error, Mean Absolute Error and Mean Absolute Percentage Error and the best method is selected. The best method is found to be the one using Functional Data Analysis by Entropy Weight Method for all the stations. The following conclusions are drawn on the basis of the analysis and Water Quality Status Report (2012), by Central Pollution Control Board:

- The preliminary analysis show that Biochemical Oxygen Demand, Dissolved Oxygen, Chemical Oxygen Demand and Total and Faecal Coliform are the most crucial water quality variables along the mainstream of Narmada River Basin. The correlation matrix further proves the inter-dependencies of Dissolved Oxygen with Biochemical Oxygen Demand ($R^2=-0.87$) and Chemical Oxygen Demand ($R^2=-0.85$). Coliform Count and Faecal Coliform are positively correlated with an R^2 value of 0.96.
- The KMO measure of sampling adequacy is 0.717 which implies that the degrees of common variances among the twenty five variables is “middling”. The factors extracted shall thus account for a fair amount of variance. Both Principal Component Analysis and Factor Analysis yield 7 components. This is further supplemented by the scree plot. The components/factors account for 68.573% of total variance. Using Principal Component Analysis, the first component is found to be strongly correlated with Dissolved Oxygen (0.854), Biochemical Oxygen Demand (-0.815), Chemical Oxygen Demand (-0.847), Total Hardness (0.711), Calcium Hardness (0.865), Calcium Major Ions (0.864), Electrical Conductivity (0.796), Coliform Count (-0.871) and Faecal Coliform (-0.892). The first component itself accounts for about 50% of the total variance explained. The first factor from Factor Analysis has a strong and positive correlation with Calcium Hardness, Dissolved Oxygen and Electrical Conductivity. It is negatively correlated with Biochemical Oxygen Demand and Chemical Oxygen Demand. Thus agricultural runoff and industrial wastes account for an increase in alkalinity in the river water. This reflects the factual information about River Narmada. It is also concluded that the deterioration in Dissolved Oxygen is being rejuvenated by variables causing alkalinity in water.
- The student's t-test conducted gives a proper account of the significant variables of each of the seven stations. For example for Station 1, Hardness, pH and phosphate ions are considered important. This is relevant because Station 1 is basically a village with adequate mineral resources. Similarly, for Station 4 Biochemical Oxygen Demand, Chemical

Oxygen Demand and Coliform Count are the three most significant parameters. Station 4 is a religious hub and witnesses numerous rituals throughout the year. Hence it is subjected to continuous anthropogenic activities. The increase in inorganic pollution is because of the paper industries located at the downstream region of the station.

- Conventional Water Quality Indexing done along the seven sites shows that for Station 1,2,5,6 and 7, water quality remains excellent almost throughout the year. In Stations 2 and 4 it drops down to Fair during the months of February and May. The decrease in water quality is due to the effect of mass bathing at Stations 2 and 4 during the given months. The water drawn during these seasons should be properly disinfected before consumption.
- In Functional Data Analysis, the indices from Relative Pollution Degree overestimate while those of entropy weights underestimate the water quality. Both the methods show similar trends of water quality. The indices developed for the stations using entropy weights almost converge with those obtained using conventional method showing minimum error. Water Quality Indexing using Principal Component Analysis also shows similar results. The principal components obtained have strong correlations with the three significant parameters of each station.
- Finally an error analysis is done on the methods used to find the best suitable water quality indexing technique. It is found that Functional Data Analysis using Entropy Weights has least error with for example, 9.3% Root Mean Square Error, 1.2% Mean Absolute Error and 1.4% Mean Absolute Percentage Error for Station 1.

References

- Abbasi, T., Abbasi, S.N. (2012). "Water Quality Indices" Published at Elsevier, Amsterdam, New York.
- Ahmad, S., Khan, I., Parida, P. (2001) "Performance of stochastic approaches for forecasting river water quality", *Wat. Res.* Vol. 35, No. 18, pp. 4261–4266.
- Bjørnstad, O.N., Robinet, C. and Liebhold, A.M. (2010). "Geographic variation in North American gypsy moth cycles: sub harmonics, generalist predators and spatial coupling". *Ecology* 91:106-118.
- Brown, R.M., McClelland, N.I., Deiningner, R.A., Tozer, R.G., (1970). "A water quality index e do we dare?" *Water Sewage Works* 117, 339e343.
- Champely, S., Doledec, S. (1997) "How to separate long-term trends from periodic variation in water quality monitoring", *Water Res.* 31, 2849–2857.
- Deshkar, S., Lakhmapurkar, J., Gavali, D. (2014) "Temporal variation in nutrient fluxes in the Narmada Estuary", *Indian Journal of Geo-Marine Sciences*, Vol. 43(2), pp 284-288.
- Feng, Y., Ling, L. (2014) "Water quality assessment of the Li Canal using a functional fuzzy synthetic evaluation model", *Environ. Sci.: Process. Impacts* 16, 1764–1771.
- Feng, Y., Ling, L., Yanfeng, L., You, Z., Musong, C., Xigang, X. (2015) "A dynamic water quality index model based on functional data analysis", *Ecological Indicators* 57, 249–258.
- Pitombo, C.Z., Gomes, M.M., (2014). "Study of Work-Travel Related Behavior Using Principal Component Analysis", *Open Journal of Statistics*, 2014, 4, 889-901.
- Haggarty, R., Miller, C., Scott, E., Wyllie, F., Smith, M. (2012) "Functional clustering of water quality data in Scotland", *Environmetrics* 23, 685–695.
- Henderson, B. (2006) "Exploring between site differences in water quality trends: a functional data analysis approach" *Environmetrics* 17, 65–80.
- Horton, R.K., (1965). "An index number system for rating water quality". *Journal of Water Pollution Control Federation* 37 (3), 300e306.
- Iglesias, C., Sancho, J., Peneiro, J.I., Martinez, J., Pastor, J.J., Taboada, J. (2015) "Shewhart-type control charts and functional data analysis for water quality analysis based on a global indicator" *Desalination and Water Treatment*, Taylor and Francis.
- Jha, R., Singh, V. P. (2008) "Evaluation of River water Quality by Entropy", *KSCE Journal of Civil Engineering*, Vol. 12(1), 61-69.
- Chatfield, C. and Collins, A. J. (2013). *Introduction to Multivariate Analysis*, Springer US, 1st edition, Page 8.

- Esbensen, K H., Guyot, D., Westad, F. and Houmoller, L. P. (2002). *Multivariate Data Analysis - in Practice: An Introduction to Multivariate*, CAMO Sosftware; 5th Edition, Page 2-3.
- Everitt, B. and Hothorn, T. (2011). *An Introduction to Applied Multivariate Analysis with R*, Springer New York; 2011 edition.
- Fabrice, R., Delannay, N., Conan-Guez, B., Verleysen, M. (2007). “Representation of Functional Data in Neural Networks”, Elsevier Science, Published in *Neurocomputing* (Volume 64, pages 183–210).
- Höppner, F. (1999). *Fuzzy Cluster Analysis: Methods of Classification, Data Analysis and Image Recognition*, Wiley; 1st edition, Page 8.
- Iscen, C F., Emiroglu, Ö. Ilhan, S., Arslan, N., Yilmaz, V., and Ahiska, S. (2008). Application of multivariate statistical techniques in the assessment of surface water quality in Uluabat Lake, Turkey. *Environmental Monitoring and Assessment* September 2008, Volume 144, Issue 1-3, Pages 269-276.
- Kim, J. and Mueller C. W. (1978). *Factor Analysis Statistical Methods and Practical Issues*. Volume 14; SAGE Publications Inc. Page 8
- Kunwar, P., Saxena, K K., Singh, N O., Nayak, A K., Tyagi, B C., Ali, S., Pandey, N N. and Mahanta P C. (2011). Application of multivariate statistical techniques for water quality characterization of Sarda Sagar Reservoir, India, *Indian J. Fish.*, 58(4): 21-26, 2011.
- Liping, Z., Jie, P., Wang, Y., Muqui, Y., Yuanyuan, S. and Liu, Y. (2010). SPSS for Water Quality Assessment of Beijing Typical River Based on Principal Component Analysis. *Proceedings of the 2010 International Conference on Digital Manufacturing & Automation - Volume 02*, Pages 395-398.
- Malviya, P., Dwivedi, A.K., (2015) “Physico- chemical parameters of Narmada River Water: A review”, *International Journal of Chemical Studies* 2015; 3(2): 01-04.
- Malviya, P., Dwivedi, A.K., Choubey, O.N., (2010) “Chemical assessment of narmada river water at Hoshangabad city and Nemawar as navel of river in Central India”, *Oriental Journal of Chemistry*, Vol. 26(1), 319-323 (2010).
- Mukati, P., Naqvi, S.M.A., Aske, D.K., Sainkhediya, J., Thakur A., (2014) “Studies of Phytoplankton Ecology in Narmada River of West Nimar, MP, India”, *Research Journal of Animal, Veterinary and Fishery Sciences* Vol. 2(4), 13-16, April (2014).
- Díaz Muñiza, C., García Nietob P.J., Alonso Fernández J.R., Martínez Torrec J., Taboadad J. (2012). “Detection of outliers in water quality monitoring samples using functional data analysis in San Esteban estuary (Northern Spain)”. *Science of The Total Environment* Volume 439, 15 November 2012, Pages 54–61.
- Najar, A.N., Khan, A.B. (2011) “Assessment of water quality and identification of pollution sources of three lakes in Kashmir, India, using multivariate analysis”. *Environmental Earth Sciences*, August 2012, Volume 66, Issue 8, pp 2367-2378
- Noori, R., Sabahi, M.S., Karbassi, A.R., Baghvand, A., and Zadeh, H. T. (2010). Multivariate analysis of surface water quality based on correlations and variations

in the dataset. *Desalination* Volume 260, Issues 1–3, 30 September 2010, Pages 129–136.

- Ozguler, H. (2005). Use of Principal Component Analysis in determining the relationships between flow and water quality parameters. *Sustainable Water Resources Solution for Large Cities IHAS Publ* 293, 2005. P-289.
- Qian, Y., Migliaccio, K.W., Wan, Y., Li, Y., (2007). “Surface water quality evaluation using multivariate methods and a new water quality index in the Indian River Lagoon, Florida”. *Water Resources Research* 43 (8)
- Ramsay, J.O., Silverman, B.W., 2002. “Applied Functional Data Analysis: Methods and Case Studies”. Springer, New York
- Sahoo, M. M. (2014) Analysis and Modelling Of Surface Water Quality in River Basins, Page 82, Masters in NIT Rourkela, Unpublished.
- Sayadi, MH, Rezaei, A., Rezaei, M. R. and Nourozi K. (2014). Multivariate statistical analysis of surface water chemistry: A case study of Gharasoo River, Iran. *Proceedings of the International Academy of Ecology and Environmental Sciences*, 2014, 4(3): 114-122.
- Selle, B., Schwientek, M., Lischeid G., (2013). “Understanding processes governing water quality in catchments using principal component scores”, *Journal of Hydrology*, Volume 486, 12 April 2013, Pages 31–38.
- Soni, V., Khwaja S., Visavadia M. (2013) “Preimpoundment studies on Water Quality of Narmada River of India”, *International Research Journal of Environment Sciences* ISSN 2319–1414 Vol. 2(6), 31-38, June (2013)
- Sharma, S., Dixit, S., Jain, P., Shah, K. W., Vishwakarma, R., (2008). “Statistical evaluation of hydrobiological parameters of Narmada River water at Hoshangabad City, India”, *Environ Monit Assess* (2008) 143:195–202.
- Sharma, S., Dixit, S., Jain, P., Shah, K. W., Vishwakarma, R., (2011). “Evaluation of Water Quality of Narmada River with reference to Physicochemical Parameters at Hoshangabad city, MP, India”, *Research Journal of Chemical Sciences* Vol. 1(3) June (2011).
- Shrestha S., Kazama F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling & Software* Volume 22, Issue 4, April 2007, Pages 464–475. Special section: Environmental Risk and Emergency Management.
- Srichandan, S., Kim, J.K., Bhadury, P., Barik, S.K., Pradipta R., Samal, M.R.N., Pattnaik, A.K., Rastog G. (2015). “Spatiotemporal distribution and composition of phytoplankton assemblages in a coastal tropical lagoon: Chilika, India”, *Environmental Monitoring and Assessment* February 2015, 187:47.
- Telang, S., Saxena, Y., Chaturvedi, A. (2009). “Effect of mass bathing on the water quality of Narmada river at district Hoshangabad, (M.P.) India”, *Current World Environment* Vol. 4(1), 211-213 (2009).
- Wang, C., Feng, Y., Gao, P., Ren, N., Li, B.L. (2012) “Simulation and prediction of phenolic compounds fate in Songhua River, China”, *Sci Total Environ.* 2012 Aug 1; 431:366-74.