IDENTIFICATION AND APPORTIONMENT OF POLLUTION SOURCES TO GROUNDWATER QUALITY USING RECEPTOR MODELS

Thesis

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

By

MOHAMMAD SHAHID GULGUNDI

DEPARTMENT OF APPLIED MECHANICS AND HYDRAULICS

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA

SURATHKAL, MANGALORE- 575 025

MARCH 2018

IDENTIFICATION AND APPORTIONMENT OF POLLUTION SOURCES TO GROUNDWATER QUALITY USING RECEPTOR MODELS

Thesis

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

MOHAMMAD SHAHID GULGUNDI

(AM11P02)

Under the guidance of

Dr. Amba Shetty

Associate Professor

Department of Applied Mechanics & Hydraulics

NITK, Surathkal

DEPARTMENT OF APPLIED MECHANICS AND HYDRAULICS

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA

SURATHKAL, MANGALORE- 575 025

MARCH 2018

D E C L A R A T I O N

By the Ph.D. Research Scholar

I hereby *declare* that the Research Thesis entitled **"Identification and Apportionment of Pollution Sources to Groundwater Quality using Receptor Models",** which is being submitted to the National Institute of Technology Karnataka, Surathkal in partial fulfilment of the requirements for the award of the Degree of Doctor of Philosophy in the Department of Applied Mechanics and Hydraulics is a *bonafide report of the research work carried out by me*. The material contained in this Research Thesis has not been submitted to any University or Institution for the award of any degree.

MOHAMMAD SHAHID GULGUNDI
(Register Number: 110630 AM11P02)
Department of Applied Mechanics and Hydraulics
National Institute of Technology Karnataka, India

Place: NITK Surathkal
Date:

C E R T I F I C A T E

This is to *certify* that the Research Thesis entitled **"Identification and Apportionment of Pollution Sources to Groundwater Quality using Receptor Models",** submitted by **Mohammad Shahid Gulgundi** (Register Number: 110630 AM11P02) as the record of the research work carried out by him, is *accepted as the Research Thesis submission* in partial fulfilment of the requirements for the award of degree of Doctor of Philosophy.

Dr. Amba Shetty
(Research Guide)
Department of Applied Mechanics and Hydraulics
National Institute of Technology Karnataka, India

Prof. Amai Mahesha
(Chairman – DRPC)
Department of Applied Mechanics and Hydraulics
National Institute of Technology Karnataka, India

ACKNOWLEDGEMENTS

With deep sense of gratitude, I express my heartfelt thanks to my research supervisor Dr. Amba Shetty, Associate Professor, Department of Applied Mechanics and Hydraulics, NITK, Surathkal. Her logical and tactical suggestions have been very valuable and encouraging during the development of this research.

I am grateful to Research Progress Assessment Committee members, Prof. G Srinikethan and Prof. A. Mahesha, for their critical evaluation and useful suggestions during the progress of the work.

I am greatly indebted to Prof. M. K. Nagaraj, Prof. Subba Rao and Prof. Dwarakish G.S, the former Head of the Department and Prof. A. Mahesha., the present Head of the Department, Department of Applied Mechanics and Hydraulics, NITK, Surathkal, for granting me the permission for using the departmental facilities in carrying out the present research work.

I thank the former Director of the NITK, Surathkal, Prof. Sawapan Bhattacharya, and present Director Prof. Karanam Uma Maheshwar Rao for granting me the permission to use the institutional infrastructure facilities, without which this research work would have been impossible.

I am gratefully indebted to Dr. Nagappa and Dr A. Mahesh of Karnataka State Pollution Control Board for helping me with the groundwater quality data and their valuable suggestions during the course of my research. Also i would like to extend my deep sense of gratitude to Dr. B. P. Lakshmikantha and M V Shashirekha of Karnataka State Remote Sensing Application centre for their help.

I also thank Prof. Shivakumar K, former Principal of Atria Institute of Technology and Dr. K. V. Narayanswamy, the present Principal for their support and encouragement. Also I thank Dr. Manjunath. G. R, Head of Department of Civil Engineering of Atria Institute of Technology, Bangalore for his support throughout my research.

I sincerely acknowledge the help and support rendered by all the Faculties, staffs and Research scholars of Department of Applied Mechanics & Hydraulics.

I gratefully acknowledge the support and all the help rendered by Ms Binny Gopal, Ms. Varadha Vijayaraghavan, Ms. Anusree K during my research.

I express heartfelt gratitude to authors of all those research publications, which have been referred in this thesis. I also express my gratitude to reviewers of journals for their excellent input for this work.

Finally, I wish to express gratitude, love and affection to my Mother, Wife, Brothers and all my family members for their continuous support and sacrifice they had to make.

Mohammad Shahid.

ABSTRACT

Characterizing groundwater quality and apportionment of pollution sources to groundwater pollution is important for managing water resources effectively. Owing to rapid industrialization and population growth in Bengaluru city, the groundwater quality is getting deteriorated. Application of source apportionment techniques to water quality problems, especially with respect to groundwater are limited in the Indian context. Therefore a scope exists for source apportionment of pollution sources to groundwater quality using receptor models.

Multivariate statistical techniques (Cluster analysis, Discriminant analysis and Principal component analysis) and Receptor oriented source apportionment models were used to evaluate groundwater quality. To have first-hand information on the quality of groundwater, samples were collected and analyzed for 14 physico-chemical parameters from 67 sites distributed across the western half of the city region during the year 2013. It was revealed that overall groundwater quality in the study area is found to be less than desirable. To find out the possibility of heavy metals contamination also, groundwater quality data obtained from Karnataka State Pollution Control Board (KSPCB) on 20 parameters (physical, chemical and heavy metals) from 41 sampling stations (monthly data) was collected for the year 2015 and was used in the final analysis of this study for peenya industrial region.

From the basic statistical analysis, it was observed that the average concentration of five groundwater quality parameters (turbidity, total hardness, iron, manganese chromium) considered for the study were exceeding permissible limit ,especially chromium which is known to be human carcinogen.

Multivariate statistical techniques such as Cluster analysis (CA) was useful in classifying the 41 sampling sites into 3 main clusters as high pollution and low pollution areas. Discriminant analysis (DA) revealed that T-Hard, NO_3, Ca, Mg, HCO_3 and TDS were the most significant parameters causing the temporal variations in groundwater quality and accounted for 94% assignation of seasonal cases. Fe, Cr, Cl, Mn, Cu and Cd were the most important parameters discriminating between the 3 clusters and accounted for 92% spatial assignation of cases thereby, delineating a few indicator parameters responsible for large variations in the groundwater quality. Principal component analysis (PCA) through varimax rotation achieved a simpler and more meaningful representation of the underlying factors by identifying 7 factors/sources with eigen value greater than one explaining 73.42% of the total variance.

Receptor oriented source apportionment modeling using Absolute Principal Component Score Multi-Linear Regression (APCS-MLR), Unmix and Positive Matrix factorization provided apportionment of various sources responsible for the groundwater quality characteristics in the study area. The percentage contribution of the identified sources was determined. Results indicated that most variables were primarily affected by rock water interactions, seepage of sewage, geology of the area and industrial discharges especially different types of electroplating industries. It was also noticed that few parameters gained significant contribution from unidentified sources.

Model performance was evaluated based on the ratio of estimated mean to measured mean (E/M). Results revealed that all three models provided good results regarding their ability to reproduce measured concentrations in most of the cases, with the APCS-MLR model showing better performance. This study concludes that these apportionment results provide useful help for policy and decision makers to enhance their ability to put in place effective policy and regulatory measures to reduce groundwater pollution.

Keywords: Groundwater quality, Multivariate statistical techniques, Cluster analysis; Discriminant analysis, Principal component analysis, Source apportionment, APCS-MLR, Unmix, PMF.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
APHA	American Public Health
APCS	Absolute Principal Component Scores
BIS	Bureau of Indian Standards
BWSSB	Bangalore Water Supply And Sewerage Board
CA	Cluster Analysis
Ca	Calcium
Cadmium	Cadmium
Cl	Chloride
CM	Classification Matrix
Cr	Chromium
Cu	Copper
DA	Discriminant Analysis
DF	Discriminant Function
EC	Electrical Conductivity
F	Fluoride
Factor Analysis	Factor Analysis
FCA	Fuzzy Comprehensive assessment
Fe	Iron
GIS	Geographical Information System
HCA	Hierarchical Cluster Analysis
HCO_3	Alkalinity
Potassium	Potassium

KIADB	Karnataka Industrial Area Development Board
KMO	Kaiser-Meyer-Olkin
KSPCB	Karnataka State Pollution Control Board
KSSIDC	Karnataka Small Industries Development Corporation
MDL	Minimum Detection Limit
MLD	Million Litres Per Day
Mg	Magnesium
Mn	Manganese
Na	Sodium
NO_3	Nitrate
Pb	Lead
PCA	Principal Component Analysis
PFA	Principal Factor Analysis
PMF	Positive Matrix Factorization
SO_4	Sulphate
SS	Sum of Squares
TDS	Total Dissolved Solids
T-Hard	Total Hardness
UIS	Unidentified Source
USEPA	United States Environmental Protection Agency
VF	Varifactor
WQI	Water Quality Index
Zn	Zinc

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Groundwater is one of most significant sources of consumable water which plays a very important part in shaping the economic and social health of majority of urban regions of India. It is accepted to be protected, free from pathogenic microorganisms and suspended matter. Worldwide extraction of groundwater is increasing to take care of increasing demands; the significance of the quality of groundwater likewise builds with respect to its monetary worth and helpfulness (Tziriti et al., 2015). Characterizing groundwater quality and recognizing potential contamination sources can substantially enhance our insight regarding natural and human effects on the groundwater quality (Helena et al., 2000). Composition of groundwater in an area relies upon natural and anthropogenic procedures which can adjust these systems by polluting them or altering the hydrological cycle. Contamination of groundwater due to increasing population, extension of commercial activities and other developmental activities are clear in numerous urban ranges.

The matter which can pollute groundwater can be essentially named natural and anthropogenic. Groundwater pollution can start above or beneath the surface of the earth. Infiltration of polluted water causes contamination beneath the surface of the earth. Numerous urban contamination sources like infiltration from sewers and storm water, solid waste disposal sites and fuel stockpiling tanks are prone to release underneath the ground surface, by passing any defensive cover presented by the soil layer. On comparison with water in streams and rivers, the movement of groundwater is slow and subsequently once the contamination occurs to the groundwater; there is little degree of dilution and dispersion (Kumar et al., 2017).

Global increase in population is causing higher utilization of water for household, industrial and irrigation purposes. In this regard, it has become important to avoid and control water contamination and to have dependable data on water quality for its systematic supervision. As a result of spatial and temporal changes in the hydrochemistry of subsurface water, periodic monitoring programs are essential (Singh et al., 2005). Generally monitoring is carried out through water sample collected from network of open/bore wells spread throughout the region of interest. Usually measurements are made seasonally which generates large datasets composed of many parameters that are difficult to understand and interpret. The contemporary data on the standing of groundwater resources in India is showing many disturbing trends. In India, for the past few years there has been an unregulated growth of urban areas, implying absence of infrastructural facilities like collection, transportation, treatment and disposal of domestic and industrial waste. These unscientific developments have caused groundwater problems to a stage beyond repair (Machiwal and Jha 2014).

1.2 NEED FOR THE STUDY

In India there is a growing dependence on groundwater resources to meet the increasing demands of growing population, urbanization, rapid industrialization, and agriculture. Added to this existing surface water are getting polluted because of variety of reasons. These progressive activities additionally have the unfavorable impact of polluting the groundwater resources making it unsuitable for various vital functions. Hence there is a need to develop and manage the available groundwater resources in a scientific and efficient manner and also to identify and mitigate the impacts of pollution in order to maintain good groundwater quality (Khanam and Singh 2014).

Contaminants are continuously being added to the groundwater by means of manmade projects and natural processes. The deposition of solid waste from industrial units which are close to the industrial facilities reacts with infiltrated rainwater and joins the groundwater level. This percolating water collects a lot of dissolved constituents and joins the aquifer framework and pollutes the groundwater. Anthropogenic activities such as increasing exploitation of groundwater resources and leaching of chemicals from agriculture, industry and domestic sources towards the aquifer cause contamination of groundwater affecting its quality (Ravikumar et al., 2011).

Chemical constituents like fluoride, arsenic and selenium represent a serious health risk in the nation. It is assessed that around 70 million individuals in 20 States are at danger because of heavy fluoride and around 10 million individuals are at risk because of heavy arsenic in ground water. Apart from this, increment in the concentration of Chloride, TDS, Nitrate, Iron in groundwater is of incredible concern towards a sustainable drinking water scheme (Bhutiani et al., 2016).

The issue of groundwater contamination in many parts of the nation has turned out to be so intense that unless dire steps are taken to prevent it, groundwater assets might be contaminated permanently. Groundwater quality estimation and management requires regular and extensive monitoring of various water quality parameters at different locations (Jaiswal et al., 20013). This generates a huge and complicated data matrix involving many parameters/variables that are not simple enough for meaningful explanation. To draw worthwhile interpretations from these huge datasets multivariate statistical techniques are used.

1.3 MULTIVARIATE STATISTICAL APPROACH

Multivariate statistics have been broadly connected in earth and ecological sciences. Cases incorporate orders of hydrogeology and geochemistry **(**Vega et al., 1998; Shrestha & Kazama 2007), hydrology (Ali et al., 2012) and science (for species recognizable proof and scientific classification). Multivariate techniques include the concurrent investigation of numerous factors instead of an examination of every factor separately.

These techniques are especially suited for the recognition of common features and additionally contrasts between extensive informational data, for example, water chemistry in the particular regions. There are different statistical analysis methods, each with its own type of analysis according to the problem being selected. Important ones being Principal component Analysis (PCA), Discriminate Analysis (DA) and Cluster Analysis (CA) (Alberto et al., 2001).

Principal component Analysis (PCA) is a mathematical tool that uses orthogonal transformation to convert a set of observations of possibly correlated variables called principal components. Factor Analysis (FA); factor analysis is closely related to PCA. It is a method used to describe variability among observed, correlated variables in terms of the lower number of variables called factors (Bengraine and Marhaba 2003).

Discriminate Analysis (DA) or Canonical Variate Analysis, is a statistical analysis to predict a categorical dependent variable or grouping variable by one or more binary variables, independent of continuous variables called predictor variables. It attempts to establish whether a set of variables can be used to distinguish between two or more group classes (Bu et al., 2010).

Cluster Analysis (CA) or Clustering is the grouping of a set of variables in such a way that objects in the same group called clusters are more similar to each other than those of other groups, also called dissimilar clusters (Cloutier et al., 2008).

One such use of multivariate statistical methods to environmental pollution research is known as source apportionment, receptor modeling technique, which has the ability to quantify the identified pollution sources.

1.4 RECEPTOR ORIENTED SOURCE APPORTIONMENT MODELLING

In the past few years, there has been an increasing enthusiasm for the utilization of multivariate statistical methods to various environmental research areas (water, soil, air pollution etc). Receptor modeling is an assemblage of techniques for recognizing significant pollution sources and evaluating the contribution of every source. It takes into account the surrounding measurements of air/water pollutants extracted from a given monitoring site. Primarily, multivariate receptor models are utilized to find out the observed air pollutant mixtures into contributions from different sources of multiple air pollutants, such as Volatile Organic Compounds (VOCs) or distinct metal constituents of fine particulate matter (PM2.5), at a receptor site (Hopke 2016). Different types of factor analysis or principal component analysis methods have been used in multivariate receptor modeling. Among those methods, Positive Matrix Factorization (PMF), (Paatero and Tapper 1994) and Unmix (Henry and Kim 1990, 2003) have gained most prevalence among environmental engineers and researchers and have been broadly utilized as part of practice.

1.5 PROBLEM IDENTIFICATION

India is rapidly moving towards a crisis of ground water overuse and tainting. The overuse of ground water is portrayed as a situation in which, over some stretch of time, normal extraction rate from aquifers is more prominent than the normal recharge rate.

From the past decade groundwater quality has been deteriorating in numerous parts of Bangalore city as an after effect of anthropogenic activities, hence it is important to characterize groundwater quality and find out the potential contamination sources for its ideal usage and sustainable environment. A study conducted by the Mines and geology division, Karnataka, India in 2011-12 revealed that the bore well water near Peenya industrial region have heavy metals like zinc, copper, lead, manganese, chromium and aluminum, well past the permissible limits. High Total dissolved solids and fluoride content in groundwater are the signs of deteriorating groundwater. Excessive quantity of nitrates, low and high pH values and the presence of e-coli and total coliform microorganisms are present in the groundwater due to the release of untreated sewage waste into the normal seepage system.

As per the Karnataka State Pollution Control Board, Groundwater and Soil contamination with Cr^{+6} was observed from an engineering industry located at Northern Bengaluru. The source of Cr^{+6} is due to plating activity. Fifteen bore wells were also found to be contaminated with Cr^{+6} ranging from 0.03 mg/L to 70.5 mg/L due to improper management of liquid waste generated from plating operation.
In this regard proper knowledge regarding the spatial distribution, temporal variation, factors influencing groundwater quality and apportioning the sources of groundwater pollution is necessary for the organized supervision of groundwater resources.

1.6 OBJECTIVES OF THE PRESENT STUDY

From the preceding sections it may be noted that various groundwater quality assessment study programs have been conducted in the region, however the application of source apportionment receptor model techniques to water quality problems, especially with respect to groundwater has not been carried out. Also, the ability of multivariate techniques to handle multiple parameters and large volume of data together to give a wholesome idea about the existing conditions makes it highly suitable for the assessment of groundwater quality data.

The main objective therefore of this research is to identify and determine the impact of different sources on the quality of subsurface water of the study area through statistical approaches. The specific objectives are;

1. Application of different Multivariate statistical techniques like

- Cluster analysis to extract information about the similarities or dissimilarities between sampling sites by detecting spatial patterns in groundwater quality and thereby classifying the sampling sites accordingly.
- Discriminant analysis to identify the most significant variables responsible for spatial and temporal variations in groundwater quality.
- Principal component analysis/factor analysis to identify the influence of possible sources (natural and anthropogenic) on the groundwater quality parameters of the study area.

2. Application of Receptor oriented source apportionment models like APCS-MLR, Positive Matrix Factorization (PMF) and Unmix for

- Apportioning responsibility for contamination observed at the receptor site among the sources and evaluating their performances.

In this way receptor models can be utilized to evaluate contributions from various sources in light of continuous monitoring at inspecting sites. Further, for effective groundwater pollution control and water resource management, it will be helpful to identify the pollution sources and their quantitative contributions. Thus, this study presents usefulness of multivariate statistical techniques in groundwater quality assessment, identification and apportionment of pollution sources/factors with a view to get better information about the groundwater quality and design of monitoring network/strategy for effective management of water resources.

1.7 OUTLINE OF THESIS

Chapter 1 – Introduction: Describes the background of the study, short review of groundwater quality problems in India. Along with a brief introduction on Multivariate statistical approach and Receptor oriented source apportionment modeling, Objectives of the present research and outline of the thesis are presented in this chapter.

Chapter 2 - An overview of the groundwater use and quality: Brings out the global and Indian groundwater use and its availability. Also outlines the various sources to groundwater quality contamination in urban areas and their effects with the summary.

Chapter 3 – Literature Review: Provides the detailed review of the published literature with reference to (i) urbanization and its impacts on groundwater quality, (ii) application of multivariate statistical techniques to explain water quality data. This chapter also gives a brief overview of various source apportionment techniques used worldwide and finally identifies the literature gap.

Chapter 4 - Materials and Methodology: Briefly outlines the methodology adopted, description of the study area, climate, hydrogeology, data collection, parameters used, detailed discussion of various multivariate statistical techniques like cluster analysis (CA), discriminant analysis (DA), principal component analysis (PCA)/factor analysis (FA), detailed explanation of different source apportionment models like APCS-MLR, PMF, Unmix software's utilized for the analysis.

Chapter 5 - Multivariate Statistical Techniques: The results of basic statistical analysis, detailed discussion of results from different multivariate statistical techniques like cluster analysis (CA), discriminant analysis (DA), principal component analysis (PCA)/factor analysis (FA).

Chapter 6 - Receptor oriented Source Apportionment modeling: This chapter deals with the detailed discussion of results from different source apportionment models like APCS-MLR, PMF, Unmix and compares the performance evaluation of the three source apportionment models used for the study.

Chapter 7 – Summary and Conclusions: This chapter summarizes the work carried out and highlights important conclusions, along with limitations and future direction for research.

The following chapter discusses the overall groundwater use and its quantity and quality status globally and specially in the Indian context setting the need for identification of pollution sources.

CHAPTER 2

AN OVERVIEW OF GROUNDWATER QUALITY AND USE

2.1 INTRODUCTION

The greater part of the earth's freshwater is not located in lakes or waterways, but rather is in subsurface aquifers (Amadi et al., 2010). The contribution from groundwater is very important; more than two billion people rely straight away on aquifers for drinking water, and around 40% of the world's food supplies delivered are through agriculture which are groundwater dependent (Chabukdhara et al., 2017). In the coming years, aquifer evolution will keep on being crucial to economic development and dependable water supplies will be required for domestic, industrial and irrigation purposes.

2.2 GLOBAL GROUNDWATER USE

Globally, there is tremendous use of groundwater, yet it is by and large perceived that the degree of its utilization has a tendency to be underestimated. The very ease and universality of groundwater improvement implies that much imperative little scale use is not included in official statistics. Groundwater is frequently assumed to be always available by governments and society. In the year 2000, twenty three urban cities in the globe were having a population exceeding 1 crore, and are in this way referred to as megacities. Over half of the megacities depend upon, or make critical utilization of, local available groundwater (Chéné, 1996). China solely encompasses 500 or more urban cities, and 2/3rd of the water demand for them is abstracted from aquifers (Fig. 2.1).

Urban dependence on groundwater does not depend upon climate and latitude. Hence, right around 1/3rd of the biggest urban communities of Russia fulfill their water demands predominantly by subsurface water, as do a large portion of the central and West African countries. It is assessed that numerous several urban communities worldwide are groundwater dependent. The utilization of groundwater for local supply exceeds across the board in little towns and villages (Llamas and Custodio 2003). This is very much outlined in eastern China, where the Huang-Huai-Hai aquifer framework supplies almost 160 million individuals. It is evaluated that just about 1/3rd of Asia's drinking water supply originates from groundwater. In USA, more than 95% of the country rural population relies upon aquifers for their drinking water.

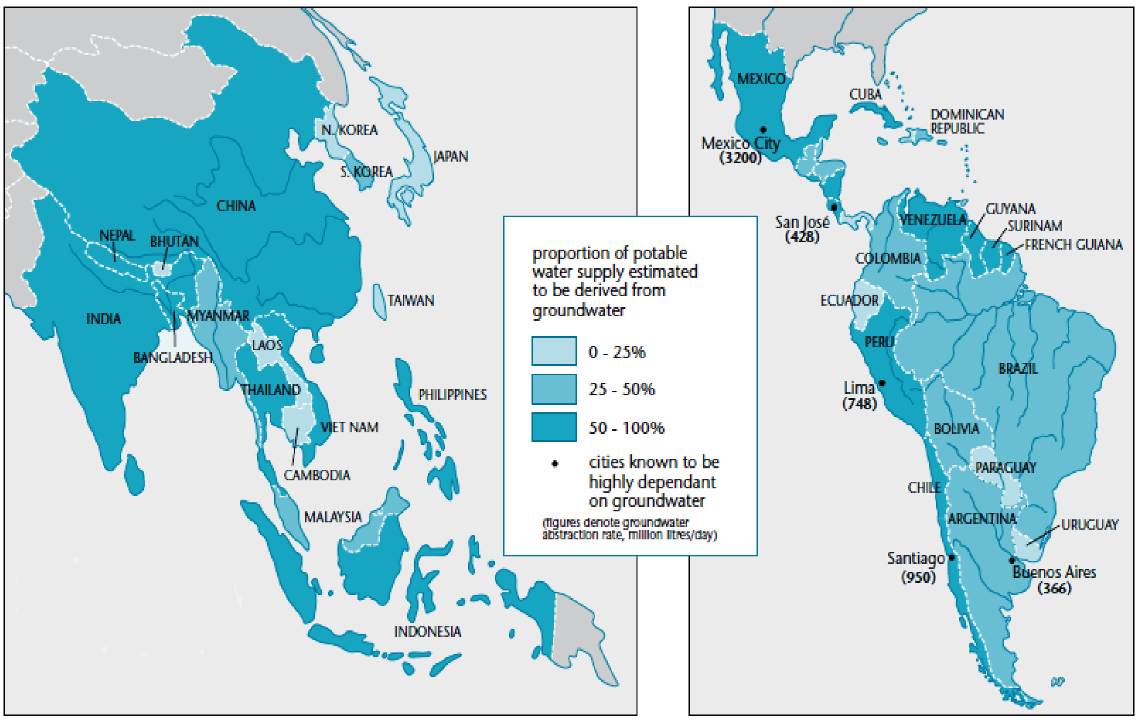


Fig 2.1: Estimated range of groundwater extracted for drinking water in Asia and Latin America (Chéné, 1996)

In the Fig. 2.1, it can be inferred there is high urban reliance on groundwater which is reflected in Asia and in Central and South America. It can be observed that cities like Mexico City; San José; Lima; Santiago; Buenos Aires from Latin America are highly dependent on groundwater, where as in Asian countries like India, Nepal, Bhutan, China and Thailand groundwater dependency for drinking varies from 50% in certain places to 100% in other places.

2.3 TEMPORAL PATTERNS OF GROUNDWATER WITHDRAWAL AROUND THE WORLD

The data available on subsurface water use is inadequate. Likewise the data on the effect of agricultural groundwater use, on food security, rural livelihoods, and ecological systems are even more so. However there is minimal uncertainty that groundwater use has continued to increase in many parts of the world. Along with that, groundwater usage throughout the globe in the past century has evolved and moved ahead in waves.

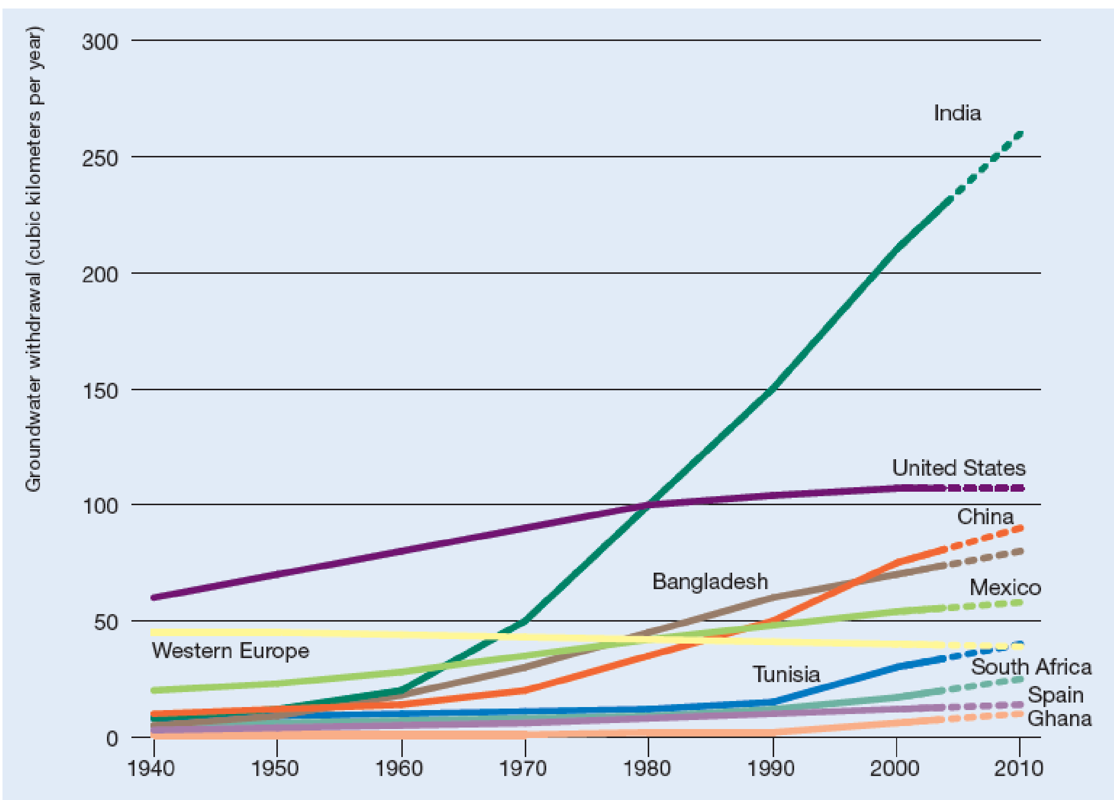


Fig 2.2: Groundwater withdrawal in selected countries (Shah 2005)

From Fig. 2.2 , it can be observed that the first wave groundwater usage was in Europe, Mexico, Spain, and the United States, where huge quantity of groundwater utilization started in the early parts of the 1900s and appears to have mountain topped or if nothing else to have quit increasing. The second wave started in South Asia (India), parts of the North China, and parts of the Middle East and North Africa amid the 1970s is as yet proceeding. The major substantial change in the groundwater scenario in Indian agricultural sector was the usage of tube wells in irrigated areas, which rose from a mere 2 % in 1960-61 to 51 % in 2006-07, which has caused a steep increase in the groundwater use of the country.

Dependable and unequivocal worldwide figures are hard to acquire either in light of the fact that the part of private local supply is unquantified or since numerous towns and urban communities meet their demands from a combined mix of surface and sub-surface water. The extent of usage changes with respect to the season of year or with usage trends. The approximate quantification of the utilization of sub-surface water for potable supply around the globe is presented in Table 2.1.

Table 2.1: Assessed level of drinking water supply acquired from groundwater in percentage (Sampat 2000)

Region	Per cent	Population served (millions)
Asia–Pacific	32	1000 – 2000
Europe	75	200 – 500
Central and South America	29	150
USA	51	135
Australia	15	3
Africa	NA	NA
World	-	1500 –2750

Table 2.1, shows drinking water supply obtained by groundwater in various continents of the world on an approximate basis. In Africa and Asia, a large portion of the biggest urban areas utilize surface water, yet a huge number of individuals in the provincial zones are reliant on groundwater. In USA, more than 95% of the countryside people rely upon groundwater for their drinking purpose. A large number of urban communities in Europe are also depending on groundwater. In countries like Austria, Denmark, Portugal, Iceland and Switzerland which have sufficient groundwater reservoirs, more than 75% of the water demand of the people is pumped from groundwater. Also 50-75% in Belgium, Finland, France, Germany, and Luxembourg, and less than 50% in Norway, Spain, Sweden, and the UK meet their water demands by relying on groundwater.

2.4 GROUND WATER AVAILABILITY IN INDIA

As of April 2015, the water resource potential or yearly water availability of the Nation with respect to natural runoff in rivers is around 1,869 Billion Cubic Meter (BCM)/year. But, the usable water resources of the nation have been assessed as 1,123 BCM/year. This is because of imperatives of geology and uneven dispersion of the resource in different river catchments, which makes it hard to extricate the whole accessible 1,869 BCM/year. Out of the 1,123 BCM/year, the share of surface water and groundwater is 690 BCM/year and 433 BCM/year respectively as shown in Table 2.2.

Table 2.2: Statistics regarding water resources in India

Parameter	Unit (Billion Cubic Meter/Year)
Annual water availability	1,869
Usable water	1,123
Surface water	690
Ground water	433

Sources: Water and Related Statistics, April2015, Central Water Commission; PRS.

Table 2.2, shows annual water availability with respect to surface and groundwater. It can be inferred from Table 2.1, that groundwater constitutes to about 40% of the country's usable water indicating that it a major source of potable water in the country for various purposes.

Experts maintain that, India is quickly moving towards an emergency of ground water overuse and contamination. Ground water overuse or overexploitation is characterized as a circumstance in which, over a period of time, average extraction rate from aquifers is more prominent than the average recharge rate. In India, the availability of surface water is more noteworthy than ground water (Londhe et al., 2004).

However, owing to decentralized availability of groundwater, it is effortlessly available and shapes the biggest share of India's agriculture and drinking water supply. 89% of ground water extricated is utilized as a part of their irrigation sector, making it the most astounding category user in the country (Shah 2005). This is trailed by ground water for household utilize which is 9% of the extracted groundwater. Industrial utilization of ground water is 2%. Half of urban water prerequisites and 85% of rural water necessities are in like manner fulfilled by ground water.

2.5 GROUNDWATER DEPENDENCE VERSUS SIZE OF CITY IN INDIA

The extent of a region is a solid marker for the amount of surface water it can import or the amount it needs to depend on the locally available sources of water. In India, 56% of metropolitan, class-I and class-II cities are dependent on groundwater either completely or to a certain extent (NIUA, 2005). Towns which are smaller than these for the most part don't have imported surface water. Subsequently, general reliance on sub-surface water for urban water supply in India is huge.

Bigger urban spots (million or more urban areas) in India are developing quickly. Be that as it may, numerous little spots (Class-I and Class-II urban communities) are also developing in the country, at a rate substantially quicker than the million or more urban areas (Mahmood & Kundu, 2004). With time, the reliance of urban cities on groundwater inside as far as possible and from encompassing territories has been increasing. (Phansalkar et al., 2005). (1 Million Plus cities, Class-I = 100,000 and above; Class-II = 50,000 – 99,999; Class-III).

In Fig. 2.3, growth in population and need for urban water supply with respect time has been plotted for a typical city. It can be observed that, at first, there is greater reliance on available local water assets i.e. water bodies, tapping shallow aquifers and so forth. As population grows in number, neighborhood water assets may never again have the capacity to satisfy the requirements.

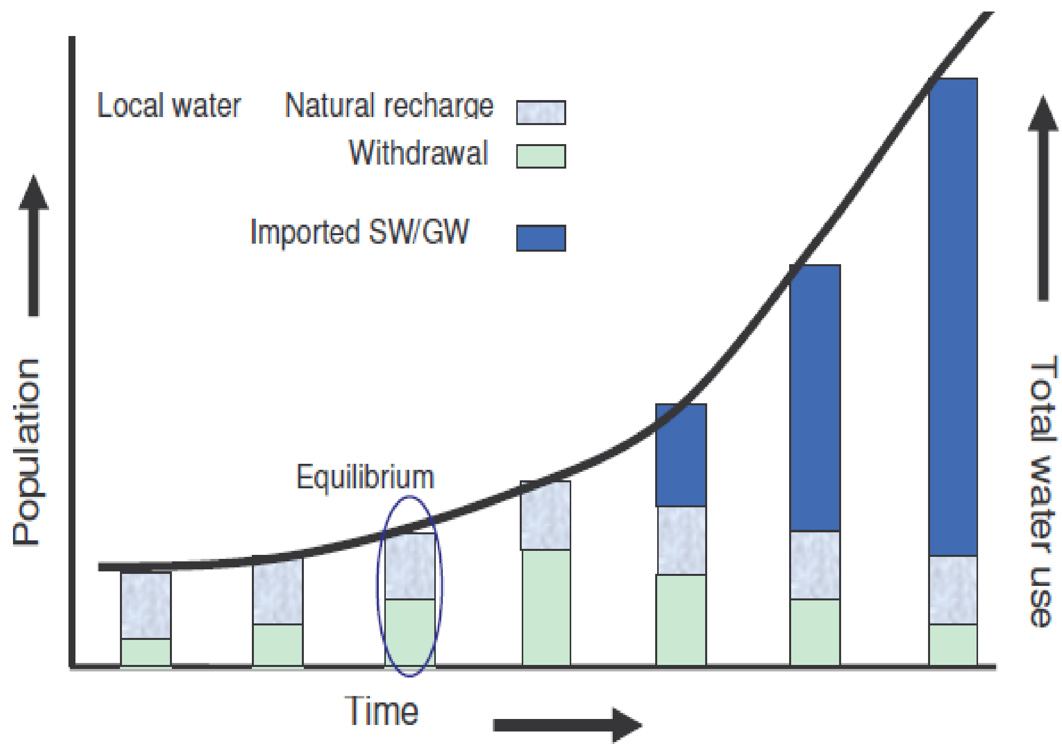


Fig 2.3: Population growth with time versus water supply (Phansalkar et al. 2005)

Subsequently falling of sub-surface water levels occurs because of utilization of bore-wells and tube-wells which causes the city to cross the equilibrium (column three from left to right in Fig. 2.3).

Huge growth in the number of private tube wells inside the city and also tankers providing drinking water to urban territories has been witnessed. The volume and extent of water brought in to urban zones increments with its development and the water supplies given by shallow unconfined aquifers may not that were initially acquired. They may also never be adequate; either on the grounds that the accessibility of the resources is limited or in light of the fact that contamination has brought about its quality to decline (Phansalkar et al., 2005). The additional water resources required are abstracted from deeper aquifers or from aquifers or surface water bodies/reservoirs in the cities distant places. In this manner, importing of water gets to be inevitable for any city in the event that it keeps on growing.

The connection between urban development and groundwater contamination is a link that necessities more investigation and can be one of the critical determinants of the two receptors, together with the drivers of urban development in a water focused environment. The same is discussed next.

2.6 URBAN SOURCES OF GROUND WATER CONTAMINATION

Providing potable water, sewage system and drainage are the essential components to manage the fast changing urban environment, and the subsurface plays a very important role in each of these components of urban facilities (Fig. 2.4).

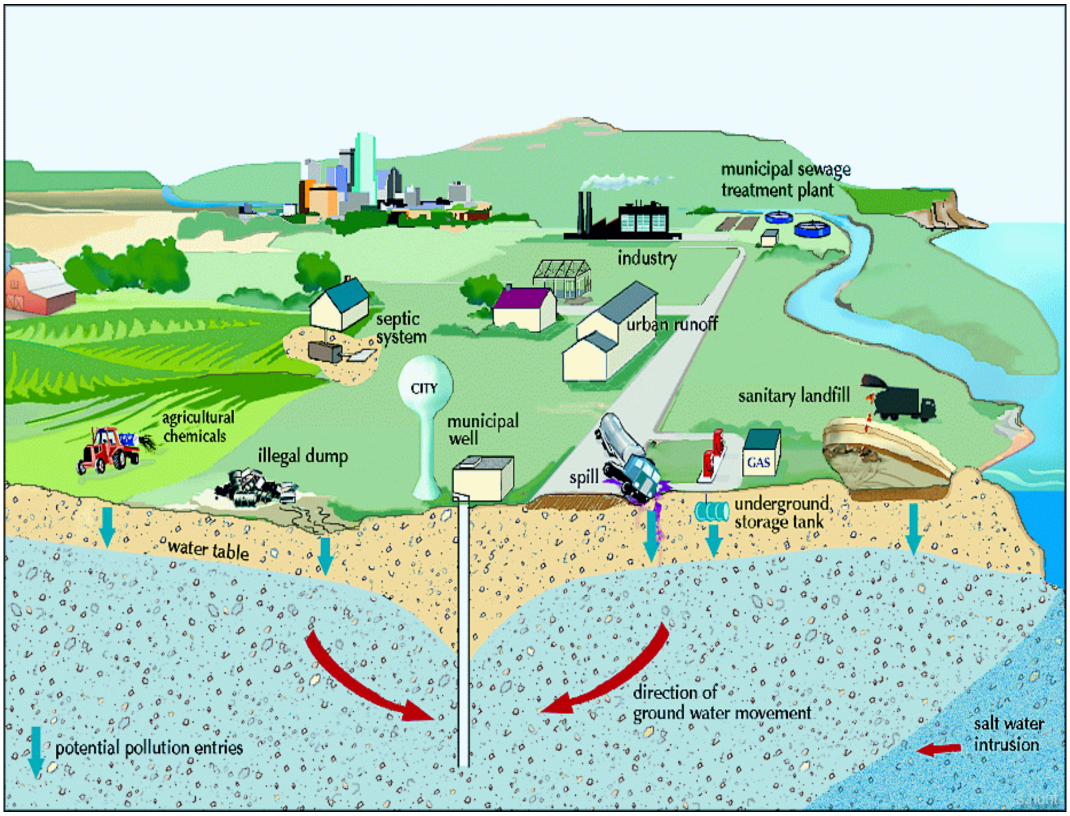


Fig 2.4: Sources of groundwater contamination (Zaporozec and Miller, 2000)

Groundwater under the cities gets affected by these components immediately. The use of subsurface for providing infrastructure (water supply and sewage pipes, tunnels, roads, metro system and deep foundations), applying fertilizers and pesticides for agricultural purpose also affects the shallow groundwater under the cities indirectly. The advantages from these activities seem worthy at the beginning, only to find out later that the environmental impacts and their related costs are not more damaging and unappreciable (Foster 1990).

From Fig. 2.4, it can be observed that groundwater quality can get significantly affected by the changing ways, amount of recharge and abstraction. The overall effect of the modified recharge on basic groundwater quality is generally unfavorable as the majority of the origins of recharge are of low quality (Table 2.3). Among these, unsewered sanitation is especially an essential source where septic tanks, soak ways, cesspits and pit latrines are utilized by thick urban populations living on shallow, unprotected aquifers.

Table 2.3, presents the different urban recharge sources and their pollution indicators which can cause the contamination of groundwater. While some are considered excellent and good in terms of the quality of recharge, few others are poor quality recharge sources which can cause contamination of groundwater. The contamination caused by them also comes with many harmful effects which are discussed next.

Table 2.3: Impact of urban recharge sources on groundwater quality and their pollutants (Morris et al, 2003)

Source of recharge	Relevance	Water quality	Pollutants/Pollution indicators
Leaking water mains	Major	Excellent	Generally no obvious indicators
On-site sanitation systems	Major	Poor	N, Cl, FC, DOC
On-site disposal or leakage of industrial wastewater	Minor-to-major	Poor	HC, industrial chemicals, N, Cl, FC, DOC
Leaking sewers	Minor	Poor	N, B, Cl, FC, SO4, industrial chemicals
Pluvial drainage from surfaces by soakways	Minor-to-major	Good-to-poor	N, Cl, FC, HC, DOC, industrial chemicals
Seepage from canals and rivers	Minor-to-major	Moderate-to-poor	N, Cl, FC, SO4, DOC, industrial chemicals
Amenity watering of parks, playing fields, private gardens	Minor-to-major	Good-to- moderate	No obvious indicators if from potable supplies, N, Cl, FC, DOC if with untreated or partially treated wastewater

B: Boron, Cl: Chloride, DOC: Dissolved organic carbon, FC: Faecal coliforms, HC: Hydrocarbons, N: Nitrogen compounds, SO_4: Sulphate

2.7 SOURCES OF CONTAMINANTS IN GROUNDWATER AND THEIR EFFECTS

Rapidly developing urban communities with a lacking wastewater framework have possibly significant impact on expanding groundwater recharge than urban communities with sewerage framework. Groundwater contamination can come from a number of natural and human-made sources as shown in Table 2.4.

Table 2.4: Contaminants Found In Groundwater and their potential effects

Contaminant	Sources to groundwater	Potential health and other effects
Arsenic	Enters environment from natural processes, industrial activities.	Causes acute and chronic toxicity, liver and kidney damage; decreases blood hemoglobin. Possible carcinogen.
Barium	Occurs naturally in some limestones, sandstones, and soils.	Can cause a variety of cardiac, gastrointestinal, and neuromuscular effects.
Cadmium	Found in low concentrations in rocks, coal, and petroleum and enters the ground and surface water when dissolved by acidic waters.	Replaces zinc biochemically in the body and causes high blood pressure, liver and kidney damage, and anemia. Destroys testicular tissue and red blood cells.
Chromium	Enters environment from old mining operations runoff and leaching into groundwater, fossil-fuel combustion, cement-plant emissions, mineral leaching, and waste incineration.	Chromium VI is much more toxic than Chromium III and causes liver and kidney damage, internal hemorrhaging, respiratory damage, dermatitis, and ulcers on the skin at high concentrations.

Copper	Enters environment from metal plating, industrial and domestic waste, mining, and mineral leaching.	Can cause stomach and intestinal distress, liver and kidney damage, anemia in high doses.
Cyanide	Often used in electroplating, steel processing, plastics, synthetic fabrics, and fertilizer production; also from improper waste disposal.	Poisoning is the result of damage to spleen, brain, and liver.
Fluoride	Occurs naturally or as an additive to municipal water supplies; widely used in industry.	Decreases incidence of tooth decay but high levels can stain or mottle teeth. Causes crippling bone disorder at very high levels.
Lead	Enters environment from industry, mining, plumbing, gasoline, coal, and as a water additive.	Affects red blood cell chemistry; delays normal physical and mental development in babies and young children.
Mercury	Occurs as an inorganic salt and as organic mercury compounds. Enters the environment from industrial waste, mining, pesticides, coal, electrical equipment (batteries, lamps, switches), smelting, and fossil-fuel combustion.	Causes acute and chronic toxicity. Targets the kidneys and can cause nervous system disorders.
Nickel	Occurs naturally in soils, groundwater, and surface water. Often used in electroplating,	Damages the heart and liver of laboratory animals exposed to large amounts over their lifetime.

	stainless steel and alloy products, mining, and refining.	
Thallium	Enters environment from soils; used in electronics, pharmaceuticals manufacturing, glass, and alloys.	Damages kidneys, liver, brain, and intestines in laboratory animals when given in high doses over their lifetime.
Pesticides	Enter environment as herbicides, insecticides, fungicides, rodenticides, and algicides.	Cause poisoning, headaches, dizziness, gastrointestinal disturbance, numbness, weakness, and cancer. Destroys nervous system, thyroid, reproductive system, liver, and kidneys.
Coliform bacteria	Occur naturally in the environment from soils and plants and in the intestines of humans and other warm-blooded animals.	Bacteria, viruses, and parasites can cause polio, cholera, typhoid fever, dysentery, and infectious hepatitis.

2.8 SUMMARY

Groundwater is a universally vital, significant and renewable resource. Its significance comes from its capacity to provide vast supply of freshwater that gives "buffer storage" amid times of dry spell. Majority of sub-surface water is of good quality as a result of natural purification process. Its modest treatment prerequisites makes it even more of a significant source of potable water which can be recharged, developed economically and effectively in a timely manner.

Groundwater is facing the risk of deterioration because of pollution and by improper utilization. In spite of its significance, groundwater is regularly misused, ineffectively comprehended and once in a while well managed. The primary dangers to groundwater maintenance emerge from the constant increment for water requirement (from increasing population and per capita use, expanding requirement for irrigation and so on) and from the expanding use and disposal of chemicals to the land surface.

An important guide for effective groundwater administration is an effectively thought out and systematically supported monitoring framework. 'Out of the picture, therefore irrelevant' is a poor reasoning for sustainable development. Blatant disregard of groundwater resources as far as national planning and surveillance can be controlled, once successful monitoring is viewed as an investment as opposed to just a waste on the resource. Consequently monitoring systems ought to be occasionally reassessed to ensure that they stay equipped for educating administration make choices, in order to caution early deterioration and give significant time to devise a successful methodology for sustainable management.

Groundwater quality estimation and management requires regular and extensive monitoring of various water quality parameters which generates a huge and complicated data matrix involving many parameters/variables which are difficult to interpret meaningfully. The ability of multivariate techniques to handle multiple parameters and large volume of data together with quantitative assessment by source apportionment models makes them highly suitable for the assessment of water quality.

Next chapter reviews the use of multivariate statistical techniques and different source apportionment models which are used to identify and apportion the pollution sources to surface/groundwater quality and derive meaningful information from the large data sets.

CHAPTER 3

LITERATURE REVIEW

3.1 INTRODUCTION

This chapter reviews the literature pertaining to the objective of the study, i.e., receptor oriented source apportionment of groundwater quality. Significant amount of work has been reported on the quality of groundwater, application of a number of multivariate statistical techniques to interpret the data matrix, identify and apportion the sources of pollution. A brief review of the different multivariate statistical techniques and source apportionment models has been included. With the increasing administrative consideration being given to groundwater pollution, it is turning out to be vital to recognize particular sources or contributors to a specific pollution issue. Receptor oriented source apportionment modeling is extensively used statistical method for source apportionment of environmental pollutants in air pollution studies (Guo and Wang 2004). Its use for water pollution source apportionment is limited (Simenov et al., 2003).

Multivariate statistical methods and explorative data examination are the suitable techniques for the significant data reduction and explanation of multi-constituent physical and chemical parameters (Juahir et al., 2008). To draw meaningful interpretations multivariate statistical techniques such as cluster analysis (CA), factor analysis (FA), principal component analysis (PCA) and discriminant analysis (DA) have been broadly utilized as a part of investigation of water-quality information (Helena et al., 2000; Wunderlin et al., 2001). As such several studies using different techniques have been carried out around the world on issues relating to groundwater quality. However the choice of selection of a particular method depends upon the objective of study i,e cluster analysis is used to address the heterogeneity in each set of data. It also provides the characteristics of each data object to the clusters to which they belong. Discriminant analysis is used when there is a need to predict a variable from a set of independent variables. When there is huge amount of complex data matrix, factor analysis is used mostly for data reduction purposes (Nosrati 2011).

Multivariate techniques allow researchers to look at relationships between variables in an overarching way and to quantify the relationship between variables. They can control association between variables by using cross tabulation, partial correlation and multiple regressions, and introduce other variables to determine the links between the independent and dependent variables or to specify the conditions under which the association takes place. This gives a much richer and realistic picture than looking at a single variable and provides a powerful test of significance compared to univariate techniques (Adams 1998).

However Multivariate techniques are complex and involve high level mathematics that requires a statistical program to analyze the data. The results of multivariate analysis are not always easy to interpret and tend to be based on assumptions that may be difficult to assess. For multivariate techniques to give meaningful results they need a large sample of data; otherwise, the results are meaningless due to high standard errors. Standard errors determine how confident you can be in the results, and you can be more confident in the results from a large sample than a small one. Running statistical programs is fairly straightforward but does require a statistician to make sense of the output (Adams 1998).

The literature reviewed in this area has been presented in this chapter and classified in to 3 categories.

- Urbanization and groundwater quality
- Multivariate statistical methods in water quality analysis
- Source apportionment

3.2 URBANIZATION AND GROUNDWATER QUALITY

Skidmore et al., (1997) stated that groundwater quality plays a critical part in groundwater conservation and quality preservation. Subsequently, it is critical to evaluate the groundwater quality for its present use as well as from the perspective of a potential source of water for future utilization. Groundwater has turned into a vital asset in the course of recent decades because of the expansion in its use for human consumption for different purposes (drinking, agriculture etc.). The quality of groundwater is similarly imperative like its quantity. Remote sensing and GIS are powerful means for water quality mapping and land spread mapping key for mapping, modeling and ecological change detection.

Ministry of water resources (MOWR 2000) report says water contamination is a grave issue in India as around 70% of its surface water resources and increasing number of its groundwater reserves are now polluted by biological, organic and inorganic toxins. Much of the time, these sources have been rendered unsafe for human utilization and in addition for different exercises such as irrigation and industrial needs. This indicates that polluted water quality can in actuality add to water shortage as it restricts its accessibility for both human use and environment.

Worldwide ecological variations due to natural variability and anthropogenic activities have caused impact on both water quantity as well as quality at regional and neighborhood scales, and additionally worldwide also (Kim et al., 2005). These changes are caused by urban expansion, growing industries, over irrigation, and extreme use of fertilizers in agricultural domain (Hassen et al., 2016). A contaminated environment detrimentally affects the well-being of individuals, animal life, and ecology **(**Sujatha & Reddy, 2003). Therefore, for the successful use of water resources maintaining the quality of water at acceptable levels is a vital necessity.

Lawrence et al., (2005) studied hazards induced by groundwater recharge under rapid urbanization in Lima;Peru, Santa Cruz, Merida, La Molina Alta Bolivia; Mexico, Hat Yai ;Thailand found out elevated nitrogen in urban groundwater. Increased concentrations of chloride originating from on-site sanitation systems, sulphate from detergents and road runoff) and bicarbonates derived from degradation of organic wastes, were frequently observed. It implies that groundwater is getting polluted from variety of sources in Thailand.

Amadi (2010) determined outcomes of urbanization on sub-surface water quality of Port-Harcourt metropolis, Nigeria. Groundwater sampling was done from different parts of the city and analyzed for relevant physico-chemical and bacteriological parameters. The effect on the groundwater quality was shown through the observed low pH (average 4.9), high nitrate, iron, manganese and copper in some locations in the area. The bacteriological investigation also indicated the high concentration of total coliform, which can be attributed the upsurge in human population in the area, coupled with poor sanitary system. In underdeveloped regions like Nigeria, the problem of groundwater contamination is very severe.

Xia, et al., (2012) used correlation analysis with GIS tools to analyze relationships between landscape pattern and water quality in Baiyangdian, in central Hebei Province of China. DEM data, land use data, and water quality monitoring data were used for the same. The results showed that the rising percentages of farmland and construction land were the major causes for water quality deterioration.

Rao et al., (2016) evaluated the groundwater quality, its variations with respect to urbanization and hydro chemical characteristics in rapidly emerging Vijayawada urban district of Andhra Pradesh. Groundwater sampling was carried out with an objective to produce base line information. The groundwater was found to be brackish, hard to very hard, most of the times supplemented with nitrate, phosphate, and fecal coliform. The study indicated there was the influence of anthropogenic sources on groundwater because of urbanization.

Groundwater quality assessment generates a huge and complicated data matrix involving many parameters/variables which are difficult to interpret meaningfully. Multivariate statistical methods are the appropriate techniques for the significant data reduction and explanation of multi-constituent physical and chemical parameters. The same is discussed in the section 3.3.

3.3 MULTIVARIATE STATISTICAL METHODS IN WATER QUALITY ANALYSIS

Groundwater quality estimation and administration requires regular and broad checking of a number water quality parameters at many locations. This produces an enormous measure of complex data which includes a large number of parameters/variables that is hard to understand. Multivariate statistical techniques have been used broadly for better comprehension of water quality in order to understand the complex data generated. Techniques like cluster analysis (CA), discriminant analysis (DA), principal component analysis (PCA)/factor analysis (FA), multiple linear regression on absolute principal component scores (APCS-MLR),etc. continually used extensively to appraise the water quality. The results from this multivariate treatment of data are being broadly used to describe and evaluate the surface and groundwater quality. The techniques are also helpful in confirming the temporal and spatial variations brought about by natural and manmade components connected to seasonality (Reisenhofer et al., 1998; Helena et al., 2000).

Vega et al., (1998) did exploratory examination of dataset of Pisuerga stream, (Center-North of Spain) using principal component analysis and cluster analysis in order to segregate sources causing dissimilarity in water quality of the region. PCA facilitated the distinguishing proof of a diminished number of ``latent" components with a hydro chemical meaning. Spatial (contamination from manmade sources) and temporal (occasional and climatic) sources influencing quality and hydrochemistry of stream water were separated and allocated to the polluting sources. The utilization of PCA and CA accomplished a significant grouping of stream water samples in light of seasonal and spatial criteria.

Alberto et al., (2001) applied different multivariate statistical techniques to assess the changes both spatially and temporally for the water quality of Suquı´a River. The methods rendered great outcomes as a first exploratory strategy to assess both spatial and temporal contrasts. In any case it neglected to demonstrate subtle elements of these distinctions. FA/PCA grouped the chosen parameters as indicated by common highlights and in addition to assess the occurrence of each gathering on the general change in water quality. DA served to incredibly diminish the dimensionality of the data set by bringing up a couple of parameters that show the greatest changes in water quality and in addition various sequences related with seasonal variation, exhibiting a novel approach for water quality assessment.

Guler et al., (2002) carried out comparative studies on the performance of graphical and statistical approaches used to differentiate water tests which included collins bar chart, schoeller plot, Piper outline and HCA. All the techniques were examined and contrasted as with their capacity to cluster, usability, and simplicity of understanding. Graphical procedures were found to have impediments when compared to multivariate strategies for extensive information sets. Principal components analysis was observed to be valuable for information reduction and to survey the progression/cover of groups or bunching/likenesses in the information. The most productive grouping was accomplished using statistical clustering techniques. It was inferred that the blend of graphical and statistical techniques gives a reliable way to characterize vast quantities of information present in the data, while holding the simplicity of great graphical presentations.

Bengraine and Marhaba (2003) studied chemical, biological and physical data monitored at 12 locations along the Passaic River, New Jersey. PCA was utilized to obtain the primary sources (solute content, temperature, supplements and organics) which were connected with the hydrochemistry variability. The spatial and temporal variations in water quality were also acquired. This study demonstrated the significance of environmental supervision connected with simple yet effective statistics which helped to better understand a complicated water system.

Singh et al., (2004) performed various multivariate statistical techniques for a huge complex water-quality information set of Gomti River in India. Cluster analysis indicated three distinctive groups of comparability between the sampling stations showing the diverse water-quality parameters. FA/PCA recognized the causative factors of the information structure clarifying 71% of the total variance of the data set. Discriminant analysis demonstrated data reduction and pattern acknowledgment amid both temporal and spatial examination. This study exhibited the need and handiness of multivariate statistical techniques, with a perspective to show signs of improvement in the information about the water quality and outline of checking system for powerful administration of water assets.

Panda et al., (2006) applied FA and CA for 3 sets of data in Mahanadi river systems. The study was aimed to identify the natural processes and manmade factors which were causing enrichment of hydrological features. R-mode FA uncovered that anthropogenic contribution of nutrients was the main reason to bring down dissolved oxygen and pH level of water. However its intensity was distinctive in fresh and saline systems and in various seasons. The connections amid the stations were depicted by cluster analysis, which was able to sort the diverse levels of pollution. The study confirmed the noteworthiness of water quality characteristics, which varies with a particular system.

Kowalkowski et al., (2006) applied chemometric techniques to the dataset which monitors the pollution of Brda River (Poland). The results permitted in deciding common clusters and groups of observing locations with alike contamination character and recognizing essential discriminant in the dataset. With the help of chemometric techniques, additional information captured was that, a few areas were under the high impact of metropolitan contamination and some others affected by agriculture (released from fields) inside the observed time period.

Ouyang et al., (2006) applied multivariate techniques to assess the seasonal correlations of water quality parameters. The principal factor analysis method was utilized to extricate the parameters that are most essential in surveying seasonal variations of stream water quality in LSJR bowl situated in upper east Florida, USA. Examination demonstrated that a variable that is most essential in adding to water quality variety for a particular season may not be critical for another season.

The spatial and temporal variations in surface water quality of the Fuji river basin were evaluated by Shrestha and Kazama (2007), with the help of different multivariate techniques. Taking into account the obtained data, it was conceivable to outline a future, ideal testing strategy, which could decrease the quantity of sampling stations and related expenses. The FA/PCA investigation did not bring about a huge data reduction. But it separated and recognized the factors/sources which were causing the change in stream water quality at the sampling sites. DA permitted a diminishment in the dimensionality of the huge data set, portraying a couple pointer parameters causing substantial variation in water quality. Therefore the value of statistical techniques for investigation and translation of complex information sets and in water quality appraisal was portrayed.

Zhou et al., (2007) used cluster analysis and discriminant analysis in order to find out the variations (temporal and spatial) in the water quality for the north-western new territories in Hong Kong. HCA grouped the 12 months into two periods and characterized the 23 monitoring sites into three groups based on likeness in water quality characteristics. DA identified 6 parameters (pH, temperature, 5 day BOD, fecal coliforms, Fe, and Ni) with higher discriminatory capacity for temporal and spatial analysis respectively. In this way, DA facilitated in reduction of the dimensionality of the huge data matrix and pointed out a few critical parameters that were causing most of the changes in water quality. Consequently, this study showed the usefulness of multivariate statistical methods for understanding complicated data matrix.

Zhou et al., (2007) used different chemometric techniques (CA, DA and PCA/FA) to identify spatial and temporal variations in marine water quality of Southern Hong Kong. Four years data comprising 19 parameters measured at 16 various sites was used for the study. Cluster analysis grouped the data in to two groups based on similarities. Discriminant analysis rendered data reduction by using only 8 parameters for causing temporal variations. Principal component analysis identified four sources (organic/eutrophication pollution, natural pollution, mineral pollution, and nutrient/fecal pollution) which were influencing the water quality.

Andrade et al., (2008) analyzed the similarities or dissimilarities among the sampling sites using cluster analysis. Factor analysis/principal component analysis (FA/PCA) was used to identify the sources of contamination in Trussu Valley, Brazil. The CA technique segregated two similar groups, which were upland and down land regions. This outcome was helpful in minimizing the sample collection and analyzation, with regard to space and time and minimal loss of data. FA/PCA helped to identify the sources which were accountable for the groundwater quality in the two different regions, pointing towards the fact that, the variables affecting the water quality composition were primarily related to soluble salts variables.

Cloutier et al., (2008) studied the groundwater hydro geochemistry of the sedimentary rock aquifer system in Que´bec by applying HCA and principal components analysis PCA to a dataset consisting of 144 samples and 14 parameters. HCA resulted in 7 geochemically distinct clusters. The combination of HCA and PCA, with conventional classification of groundwater types, as well as with the hydrogeological and geological contexts, facilitated the segregation of the area into four important geochemical areas. This gave an enhanced local picture of the aquifer framework and hydro geochemical evolution of groundwater.

Omo-Irabor et al., (2008) studied the natural and manmade pathways which were affecting the chemistry of surface and groundwater in western Niger Delta region, with the help of multivariate statistical techniques. The chemical data set generated was subjected to PCA/ FA and HCA. The different sources and their haphazard distribution pointed out by this study indicated that appropriate land use planning and strict application of current environmental regulations were important in the oil producing area keeping in mind the end goal to have effectual surface and groundwater resource management.

Shrestha et al., (2008) assessed the spatial and temporal variations of Mekong waterway utilizing multivariate methods. Information framework including 18 parameters created from 6 years was utilized. Varieties got from PCA/FA were clarified by disintegrated mineral salts along the whole stream extend. Information optimization and pattern recognition was accomplished through discriminant analysis. Hence translation of the extensive and complex datasets was accomplished through the use of multivariate techniques.

Venugopal et al., (2008) used multivariate statistical techniques to identify the factors that were behind the chemical composition of the groundwater, which was close to the gravely polluted Adyar River. Two main clusters were recognized, reflecting the groups of polluted and unpolluted stations. The results of the R-mode factor analysis showed that the groundwater chemistry of the study area was due to the influence of human activities, rock-water interactions, saline water intrusion into the river water, and subsequent infiltration into the groundwater.

Kazi et al., (2009) concluded that the utilization of various multivariate statistical techniques like CA, FA/PCA and DA helps to interpret complicated data sets for improved understanding of water quality and ecological status of a particular area. They also permit to identify the likely sources that affect water systems and offer a significant tool for dependable administration of water resources and additionally a fast answer for the contamination issues.

Suvedha et al., (2009) evaluated the usefulness of two multivariate statistical methods HCA and FA in order to classify the groundwater samples and find out the geochemical processes affecting geochemistry of the sedimentary rock aquifer system in Veeranam catchment area in India. Q- and R- mode factor and cluster analysis was applied to the hydro chemical data for 52 groundwater samples. R-mode analysis showed the inter-relations among the variables while the Q-mode analysis revealed the inter-relations among the samples. Both Q-mode factor and Q-mode cluster analyses indicated an interaction among the river water and the groundwater in the region.

Zhang et al., (2009) analyzed the dataset containing 13 water quality parameters various sites of the Daliao River Basin using multivariate statistical methods. Cluster analysis, discriminant analysis and principal component analysis were used in order to find out the temporal and spatial variations and to locate the pollution sources.

Hierarchical CA classified the 12 months into three periods and the sampling sites into three different groups. Six and five important variables were identified by DA for distinguishing temporal and spatial groups. PCA was helpful in finding out five latent pollution sources responsible.

Bu et al., (2010) investigated water quality using multivariate statistical techniques at 12 sampling sites in the Jinshui river China. The objective was to find out the temporal and spatial variations of water quality and recognize the prime reasons/sources causing pollution. Factor analysis showed that the total variance in the data was due by five components namely salinity, trophicity, organic pollution, oxide related process, and erosion. The results indicated that pollution in water was the result of domestic wastewater and agricultural runoff, thus aiding in water resource conservation.

Yidana (2010) determined the prominent sources causing variation in the hydrochemistry and the fitness of groundwater, in the aquifers of the Birimian system in Ghana. R-mode factor and Q-mode HCA, together with conventional graphical techniques were used. It was revealed that, hydrochemistry of groundwater in the area was affected by 3 main factors: silicate mineral weathering, cation exchange, carbonate mineral weathering and chemical fertilizers from farms in the area. Groundwater clusters formed by Q-mode HCA were found to have low sodium content and did not pose the sodium hazard when used for irrigation.

Nasir et al., (2011) applied multiple linear regressions (MLR) and principle component analysis (PCA) on the collective water quality dataset of over five years of Klang River, Malaysia. The objective was to quantify the contributions of various sources affecting the water quality of the river. MLR was used as a tool for surface water modeling and forecasting. PCA was used to simplify and understand the complex relationship among water quality parameters. Nine principle components were found which were responsible from the data matrix out of which the urban domestic pollution was identified as the main pollution contributor. Results proved that the using the inputs from PCA, enhanced the MLR model prediction by bringing down their complicatedness and removing data collinearity.

Zhang et al., (2011) applied PCA technique to examine the groundwater quality data gathered from Shizuishan city, China, in order to investigate the significant impact factors on water quality in the area. The results demonstrated that the fundamental pollution sources were industrial and municipal wastewaters, and the arid climate also played an imperative part in the strong evaporation which imparted great influence on the groundwater chemical components.

Kovács et al., (2012) used CA, DA and Wilks' lambda distribution to identify the different sub-areas in Lake Balaton, in order to determine the number of representative sampling locations required for water quality monitoring. CA results revealed that a total of 5 sub-areas were present, changing in number and alignment. This was confirmed utilizing DA and also the parameters that impacted the sub-zones the most were found out using Wilks' lambda dispersion. It was concluded that a minimum of 5 sampling locations were required, one in each of the sub-area in order to monitor the changes in the sub-areas and to get a comprehensive picture of the lake.

Mustapha et al., (2013) analyzed dataset consisting physicochemical variables from twenty seven sampling locations in upper Jakara River Basin using environmetric techniques. The results seemed to give proof on the cause behind the water quality disparity in the selected region. When PCA coupled with FA was applied on the data from sampling locations, the results showed that the water quality fluctuations were primarily due to anthropogenic and natural processes. The environmental tools used supplied with objective interpretation of surface water quality parameters which in turn helped in finding out water pollution source apportionment with a view to manage a sustainable river basin.

Wang et al., (2013) used cluster analysis and principal component analysis/factor analysis in order to assess the temporal/spatial variations in water quality and identify the sources of water pollution, for Songhua River Harbin region. The data matrix comprised of 15 parameters for 6 different monitoring stations in the area over a 5-year observation time span (2005–2009). Hierarchical CA assembled the 6 monitoring stations into 3 clusters in light of their similarity and on the basis of low, moderate and high pollution. PCA/FA of the 3 unique groups brought about five dormant factors in the water quality datasets of LP, MP and HP, separately.

Mavukkandy et al., (2014) applied PCA and PFA to determine the effectiveness and appropriateness of existing water quality monitoring network, in the Kabbini River basin of Kerala, India in order to understand the complicated data set of the river basin. The principal goal was to distinguish critical sampling stations which must basically be incorporated in evaluating yearly and variations with respect to season. Additionally, the importance of seasonal overhaul of the sampling network of stations was likewise researched to obtain important data on water quality from the network. It was seen that, few of the monitoring stations were redundant in explaining the annual variance of the dataset. In this manner, the study delineates that different multivariate statistical techniques can be adequately utilized in economical administration of water assets.

Khan et al., (2015) characterized the groundwater quality from the quantitative analytical data of the alluvial aquifers of Ganga-Sot Sub-Basin (GSSB) using multivariate statistical analysis. The data matrix was composed of 10 variables from 34 groundwater samples which were collected from equally spaced location points. The HCA resulted in 6 clusters and each of the 6 clusters group were subjected to Principal component analysis (PCA) individually. The reason for dissimilarity among the clusters was found to be because of the anthropogenic activities on the groundwater regime.

Qian et al., (2016) utilized multivariate methods keeping in mind the end goal to comprehend aquifer network. Major ion chemistry of groundwater from the primary aquifers and groundwater geochemical development close to the Dingi Coal Mine (Anhui, China). A total of 57 groundwater samples were analyzed. With expanding depth of the aquifers, groundwater turned out to be more mineralized indicating more noteworthy hardness and saltiness. PCA gave two primary segments, the principal segment exhibiting hardness variety and the second segment controlled basically by saltiness and sulfate diminishment process. CA demonstrated that the coal strata aquifer was to a great extent confined from the likely invasion at concealed areas.

Hassen et al., (2016) investigated the hydro chemical processes leading to mineralization in order to assess the suitability of the water quality for agriculture and drinking purposes in Oum Ali-Thelepte, Tunisia using CA and PCA. Water analysis was conducted on 16 groundwater samples during the study period. The investigation of this data information uncovered that the major and trace concentrations were inside the allowable level for human utilization. The supportability of groundwater for drinking and water system was surveyed in light of the water quality (WQI) and on Wilcox and Richards' graphs. This aquifer was designated as to have "excellent water" serving great water system in the zone.

Kumar et al., (2017) investigated on spatio-chemical, contamination sources by making use of multivariate statistics. Health risk assessment was also carried out by analyzing bore well samples for 28 parameters, which was suspected to be due to the groundwater, which was contaminated with trace and toxic elements in the industrial area of Uttar Pradesh, India. Multivariate measurements (PCA and CA) demonstrated that the natural and anthropogenic exercises like agrochemical disposal and waste from industries were in charge of water nature of the investigation region.

Chabukdhara et al., (2017) surveyed the nature of groundwater and potential wellbeing hazard because of ingestion of heavy metals in the semi-urban and urban-areas predominated with industries of Ghaziabad, Uttar Pradesh, India. The investigation was expected to assess heavy metals sources and their contamination level utilizing multivariate techniques and fuzzy comprehensive assessment (FCA), individually. This investigation demonstrated that diverse methodologies are required for the coordinated evaluation of the groundwater contamination subsequently giving a logical premise to the vital future arranging and far reaching administration.

Kazakis. et al., (2017) used multivariable procedures like CA and FA to hydro chemical data set from 3 areas of Northern Greece. The utilization of cluster analysis in the groundwater tests of every considered zone came about gainfully in the most hydro chemically complex range. Nonetheless it was inferred that, the utilization of this approach ought to be tried in different districts, despite the fact that in this examination the delimitation of the influenced zones of each procedure was as per the traditional hydrogeological translation and field experience.

The summary of surface and groundwater quality assessment using multivariate statistical techniques with respect to the type of parameters used, the technique applied, the pollution source identified in the study region are presented in Table 3.1(a) to Table 3.1(c) for surface water quality assessment and Table 3.2(a) to 3.2(b) for groundwater quality assessment.

Table 3.1(a): Summary of surface water quality assessment, using multivariate statistical techniques

Sl. No	No of Variables	Variable type	Technique	Pollution Sources identified	Country	Authors
1	22	Physico-chemical	PCA and CA	Mineral contents, man-made pollution and water temperature	Spain	Vega et al. (1998)
2	22	Physico-chemical and biological	CA,DA,PCA/ FA	Seasonal variations, urban run-off, and anthropogenic pollution sources	Argentina	Alberto et al. (2001)
3	19	Chemical, biological and physical	PCA	Solute content, temperature, supplements and organics	USA	Bengraine and Marhaba (2003)
4	24	Physico-chemical and biological	CA,DA,PCA/ FA	Natural and anthropogenic	India	Singh et al. (2004)
5	21	Physico-chemical and biological	CA and FA	Anthropogenic contribution of nutrients	India	Panda et al. (2006)
6	12	Physico-chemical and biological	CA,DA,PCA/ FA	Metropolitan contamination and Agricultural release	Poland	Kowalkowski et al. (2006)
7	14	Physico-chemical and biological	PCA	Urban run-off, heavy metal pollution, fecal pollution	USA	Ouyang et al. (2006)

Table 3.1(b): Summary of surface water quality assessment, using multivariate statistical techniques

Sl. No	No of Variables	Variable type	Technique	Pollution Sources identified	Country	Authors
8	12	Physico-chemical and biological	CA,DA,PCA/FA	Organic pollution and nutrients	Japan	Shrestha and Kazama (2007)
9	48	Chemical, biological, physical and heavy metals	CA and DA	-	HongKong	Zhou et al. (2007)
10	13	Physico-chemical	CA and PCA/FA	Soluble salts variables, nutrients and anthropogenic activities	Brazil	Andrade et al. (2008)
11	18		DA and PCA/FA	Dissolved mineral salts, organic components and nutrients	Thailand, Laos, Cambodia, Vietnam	Shrestha et al.(2008)
12	22	Physico-chemical and heavy metals	CA and FA	Human activities, rock-water interactions, saline water intrusion	India	Venugopal et al. (2008)

Table 3.1(c): Summary of surface water quality assessment, using multivariate statistical techniques

Sl. No	No of Variables	Variable type	Technique	Pollution Sources identified	Country	Authors
13	36	Physical, chemical, biological and heavy metals	CA,DA,PCA/FA	Effluent from industrial, domestic, agricultural and saline seeps	Pakistan	Kazi et al. (2009)
14	13	Physico-chemical and biological	CA,DA,PCA	Organic pollution, toxic organic pollution, heavy metal pollution, fecal pollution and oil pollution	China	Zhang et al. (2009)
15	25	Physical, chemical, biological and heavy metals	FA	Domestic wastewater and agricultural runoff	China	Bu et al. (2010)
16	14	Physico-chemical	CA and PCA/FA	Rock water interaction	India	Khan (2011)
17	30	Physico-chemical and biological	PCA	Urban domestic pollution	Malaysia	Nasir et al. (2011)

Sl. No	No of Variables	Variable type	Technique	Pollution Sources identified	Country	Authors
18	23	Inorganic	CA and DA	-	Hungary	Kovács et al. (2012)
19	15	Physico-chemical and heavy metals	PCA/FA	Irrigation agricultural, construction activities, clearing of land, and domestic waste disposal	Indonesia	Mustapha et al. (2013)
20	15	Physical, chemical, biological and heavy metals	CA,DA,PCA	Animal husbandry and agricultural activities, temperature (natural), heavy metal and toxic pollution	China	Wang et al. (2013)
21	22	Physico-chemical and biological	PCA/FA	Organic pollution, industrial pollution, diffuse pollution and fecal contamination	India	Mavukkandy et al. (2014)

CA : Cluster Analysis, DA: Discriminant Analysis, PCA: Principal component analysis, FA: Factor analysis

Table 3.2(a): Summary of groundwater quality assessment, using multivariate statistical techniques

Sl. No	No of Variables	Variable type	Technique	Pollution Sources identified	Country	Authors
1	39	Hydro chemical	CA and PCA	-	USA	Guler et al. (2002)
2	14	Physico-chemical	CA and PCA	Sea water and solute diffusion	Canada	Cloutier et al. (2008)
3	15	Physico-chemical	CA and PCA	Natural and anthropogenic	Nigeria	Omo-Irabor et al. (2008)
4	14	Physico-chemical	CA and FA	Rock water interaction	India	Suvedha et al. (2009)
5	26	Physico-chemical and trace metals	CA and FA	Silicate mineral weathering, cation exchange, carbonate mineral weathering and chemical fertilizers	Ghana	Yidana (2010)
6	14	Physico-chemical	PCA	Farmland irrigation return water and wastewater from ditches	China	Zhang et al. (2011)

Table 3.2(b): Summary of groundwater quality assessment, using multivariate statistical techniques

Sl. No	Number of Variables	Variable type	Technique	Pollution Sources identified	Country	Author
7	10	Physico-chemical	CA and PCA	Influence of top soil	India	Khan et al. (2015)
8	6	Physico-chemical	CA and PCA	-	China	Qian et al. (2016)
9	26	Physico-chemical and trace metals	CA and PCA	Water–rock interaction	Tunisia	Imen Hassen et al. (2016)
10	28	Trace and toxic metals	PCA	Agrochemical waste and industrial effluent	India	Manoj Kumar et al. (2017)
11	16	Physico-chemical and trace metals	PCA and FCA	Industrialization	India	Chabukdhara et al. (2017)
12	10	Physico-chemical	CA and PCA	Anthropogenic (nitrate pollution) and natural	Greece	Kazakis et al. (2017)

CA : Cluster Analysis, DA: Discriminant Analysis, PCA: Principal component analysis, FA: Factor analysis

- From Table 3.1(a) to 3.1(c) and 3.2(a) to 3.2(b) it is evident that several statistical methods and models have been used for the assessment of surface and groundwater quality around the globe.
- Among the several techniques applied, the important ones are cluster analysis, discriminant analysis and principal component analysis.
- It implies that these techniques are used successfully and can be employed to evaluate environmental problems, which help to identify of possible factors/sources that influence water systems and offers a robust tool for reliable water resources management as well as quick solution to pollution problems.
- Pollution sources were different in different countries, in some cases it could be return flow from agriculture, anthropogenic activities, rock-water interactions etc. Anthropogenic and natural source were more frequent sources of pollution in India.
- In the Indian context these techniques have been used to find out the similarities between sampling sites, different sources of pollution like natural and anthropogenic which leads to the conclusion that it can be applied to find the different sources of pollution.

Multivariate statistical methods like PCA can only help to identify the sources. To quantify the source contribution to environmental pollution research source apportionment, receptor modeling technique, is been used worldwide which has the ability to quantify the identified pollution sources. The use of different source apportionment models is discussed in section 3.4.

3.4 SOURCE APPORTIONMENT OF WATER QUALITY

Source apportionment investigations of water contamination can significantly enhance the learning of the human intervention on nature. Receptor oriented source apportionment modeling is broadly utilized measurable procedure for source apportionment of environmental contaminants in air pollution studies and has been as of late connected to water contamination source apportionment also (Simenov et al., 2003). Receptor modeling is the utilization of multivariate statistical methods intended to identify and apportion air pollutants to their sources (Hopke et al., 2006). Amid the most recent years, these different models have already been acknowledged for creating viable and proficient air and water quality management plans.

Distinctive models including principal component analysis/absolute principal component scores (PCA/APCS) (Song et al., 2006; García et al., 2006), edge analysis Unmix (Olson and Norris, 2008), chemical mass balance (CMB) (Chow and Watson 2002) and positive matrix factorization (PMF) (Paatero, 1997; Paatero, 1999; Gildemeister et al., 2007) have been connected to recognize and to build up the sources commitment to observed ambient concentrations.

Thurston and Spengler (1985) developed a new procedure to apportioning mass of air pollutants within different PCA source components, by calculating Absolute Principal Component Scores, and the subsequent regression of these mass and elemental concentrations on these scores, i.e. APCS-MLR. This method was connected to distinguish and measure the significant particle pollution source classes which were influencing an observation site in metropolitan Boston, MA. PCA of particulate elemental data allowed the estimation of mass contributions for five fine mass particle source classes and six coarse particle source classes and also indicated the elemental characteristics of the sources. The contribution of these sources to the total recorded elemental concentrations was then estimated using APCS-MLR technique.

Sharma et al., (1994) developed up a source-receptor model for PAH source-apportionment for different PAH species. Air concentration was related to deposition level by deposition capture and long haul maintenance properties of snow packs.

Anttila et al., (1995) applied a new variant of factor analysis (positive matrix factorization, PMF) to a finish data set of monthly bulk wet deposition concentrations of strong acids. It was noted that the elements showed by the PMF were quantitative. This in like manner demonstrates the uniqueness of the PMF outcomes i,e no rotations were required. The normal quantitative assertion was obtained with the first unrotated result.
Polissar et al., (1998) applied positive matrix factorization(PMF) method to analyze the aerosol composition data from 7 service locations in Alaska for the years 1986 to 1995. It was reasoned that positive matrix factorization was a powerful strategy for recognizing conceivable sources of aerosol in remote areas. PMF distinguished more factors and gives a quantitative distribution of aerosol mass, and estimated the related vulnerabilities in the resolved values.

Anderson et al., (2002) evaluated 4 receptor models by implementing them to reenacted individual presentation information for select unpredictable natural volatile organic compounds (VOCs). The source was monte carlo sampling from known source donors and profiles. The receptor models investigated were CMB, APCS-MLR, PMF and Unmix models. Each of the models used were able to distinguish only the utmost contributors to total exposure concentrations PMF created factor profiles were the most firmly represented significant sources that were utilized to deliver the simulated information. None of the models could perceive sources with relative chemical profiles.

Lewis et al., (2003) used Unmix receptor model to analyze $PM_{2.5}$ ambient aerosol data set collected in Phoenix. The analysis results yielded source profiles and total average percentage source contribution estimates (SCEs) for 5 sources. The Unmix SCEs were by and large predictable with a prior multivariate receptor investigation of basically the same information utilizing the PMF model. This study gave the primary exhibit to an urban region of the capacity of the Unmix receptor model.

Pekey et al., (2004) used multivariate receptor modeling approach based on factor analysis (FA) and factor analysis-multiple regression (FA-MLR) to the dataset of the polluted Dil Deresi stream, in order to apportion quantitatively the trace metal sources contributing to the surface water pollution in the stream. FA and FA-MR were able to identify and quantify the major polluting sources. It was also noticed that predicted concentrations were calculated with uncertainties lower than 15%.This study showed the importance of multivariate statistical analyses in aquatic environment studies.

Singh et al., (2005) used receptor modeling technique of APCS-MLR for apportionment of several sources causing the river pollution of the Gomti River India. The modeling pointed out the main sources causing river quality degradation. This study displayed the value of multivariate statistical techniques in water quality evaluation together with the identification and apportionment of pollution sources. Also signs of improvement in data understanding regarding the water quality and configuration of monitoring network for viable administration of water assets.

Song et al., (2006) carried out source apportionment of fine particulate matter in Beijing, China, utilizing PCA/APCS and Unmix. The information utilized as a part of this study were from the chemical analysis of 24-h tests, which were gathered at 6- day for the study region. Both models recognized five sources of PM2.5: secondary sulfate and secondary nitrate, a blended source of coal combustion and biomass burning, industrial emission, motor vehicles exhaust, and road dust. The sources identified were practically identical to past evaluation utilizing positive matrix factorization (PMF) and chemical mass balance (CMB) receptor models.

Zhou et al., (2007) applied APCS-MLR (a receptor based source apportionment method) to calculate source contributions. 5 possible sources were distinguished for the 2 clusters by rotated principal component analysis. Soil weathering, organic pollution, nutrient pollution, mineral pollution, and physicochemical and biochemical pollution were found to be the potential pollution sources for the first cluster. Soil weathering, agricultural runoff, physicochemical and mineral pollution, and natural sources were the latent pollution sources for cluster 2. Receptor-based source apportionment through APCS-MLR uncovered the fact that majority of the variables were fundamentally impacted by soil weathering, organic, nutrient pollution and mineral pollution.

Chemical mass balance (CMB) and positive matrix factorization (PMF) modeling was carried out to find out the source of discharge and contaminants in combined sewer overflows by Soonthornnonda and Christensen (2008). Based on overflow events, it was found that between 27% and 56% of the total overflow was from sanitary sewage and most of the remaining from storm water with possible minor contribution (8%) from groundwater. Majority of total suspended solids and metals were from storm water, while sanitary sewage conveyed substantial commitments (28%) of BOD5, NH_3, and total phosphorus.

Shukla and Sharma (2008) utilized source apportionment method to distinguish and allot the sources of PM10 in Kanpur. The vital sources responsible for PM10 in the study region were recognized. The study inferred that NH_3 assumed an essential part in the development of secondary particles and was clear through air quality sample collection, receptor demonstrating, and furthermore through the atmospheric chemistry.

Boamponsem et al., (2010) quantified the atmospheric heavy metal deposition in the mining area of Ghana by making use of PMF, PCA and CA techniques. The PCA and CA classified the examined elements into anthropogenic and natural sources, and PMF resolved three primary sources. Add up to substantial heavy metal concentrations acquired by instrumental neutron activation analysis (INAA) were handled by positive matrix factorization, principal component and cluster techniques. Positive matrix factorization (PMF) model was effective in recognizing three physically important factors, which represented over 90% of the variation of the 10 investigated elements.

Huang et al., (2010) used unmix and fuzzy comprehensive analysis to quantify the contributions from located pollution sources to the thirteen water quality variables, collected at the monitoring sites along Qiantang river(china). Fuzzy comprehensive analysis masterminded the informational index into three noteworthy contamination zones in view of national quality measures for surface waters. Factor investigation distinguished two and three potential contamination sources in low and high contamination zones. Unmix was utilized to calculate contributions from distinguished contamination sources to each water quality variable and each observing site. Results demonstrated that most water quality factors were principally impacted by contamination because of industrial wastewater, farming exercises and urban runoff. These outcomes provided with data to create predominant contamination control procedures for the Qiantang river.

Li and Zhang (2011) used receptor modeling technique comprising factor analysis-multiple linear regression (FA-MLR) to recognize the sources and apportion them to the pollution caused by the heavy metals in the river water of the Han in China. Results uncovered that land use was an essential element in water metals in the snow melt stream period and area use in the riparian zone was not a superior indicator of metals than area utilize far from the river. FA-MLR examination distinguished 5 sources which were found to be the major sources causing pollution in the surface waters. The outcomes exhibited extraordinary effects of human exercises on metal concentrations in the subtropical stream of China.

Su et al., (2011) evaluated the spatial and temporal variations in water quality together with apportionment of the sources to pollution in Qiantang River, using receptor-based source apportionment technique APCS–MLR. The study was carried out using 4-year data set which uncovered that, most of the parameters were fundamentally affected by local and rural sewage contamination together with mineral weathering. Be that as it may, many sources that were not identified sources in all groups, added to contamination in

Qiantang River for a large portion of the water quality parameters, which pointed to another major latent source.

Ielpo et al., (2012) applied multivariate techniques like Cluster Analysis (CA), Principal Component Analysis (PCA) and Absolute Principal Component Scores (APCS) to the groundwater quality data from Apulian agricultural sites with a view to quantify, observe and control the local groundwater quality. CA and PCA uncovered that few sampling locations showed dissimilarities, because of the orientation of the site, the land utilize and administration methods and groundwater abuse. By APCS technique three sources of pollution were recognized as farming contamination because of manure applications and microelements for horticulture and groundwater exploitation and a third source distinguished as soil run off.

Selvaraju et al., (2013) utilized a thorough receptor modelling way to distinguish and quantify the contributions of different air pollutants, utilizing factor examination (FA), positive matrix factorization (PMF), and chemical mass balance (CMB) for manufactured and field information. Data on emission inventory was utilized and a strategy was proposed for taking out a portion of the subjectivity and weaknesses of individual methodologies. The mix of the models evacuates the vulnerability in the information. The distinctions in the outcomes between the models accentuate the sensitivity of the techniques to the exactness of the receptor concentration and source profile utilized from the emission inventory.

Chen et al., (2013) connected PCA and APCS– MLR to explore attributes of water quality, recognize potential sources, and allocate their relative contributions to water contamination in Jinjiang River, China. Horticultural exercises, industrial wastewater release, household sewage, and mineral contamination were distinguished as real contamination sources. Rotated PCA and receptor modelling through (APCS– MLR) uncovered potential contamination sources and their comparing commitments. Results demonstrated that exhaustive utilization of different multivariate techniques were compelling for water quality evaluation and administration.

Yang et al., (2013) gathered bimonthly water samples in Wen-Rui-Tang (WRT) river watershed, near the East China Sea at 12 monitoring sites. Multivariate statistical techniques such as principal component analysis (PCA), and absolute principal component score—multiple linear regression (APCS-MLR) were used to discover the spatial dispersion of water quality and to allot the contamination sources. The results demonstrated that most water quality parameters had no huge contrast between the urban and rural zones.

Bhuiyan et al., (2015) examined concentrations of heavy metals in water and sediment of Buriganga River in the capital city Dhaka, Bangladesh. The purpose was to comprehend the level of heavy metals and their source distribution. By using positive matrix factorization (PMF) and examining correlations, the end goal to clarify the substance, conduct, and source apportionment of metals. PMF brought about an effective dividing of fluctuations into sources identified with back-ground geochemistry and contaminant impacts. However, the PMF approach effectively delineated the significant sources of metals from different activities in the area.

Bhutiani et al., (2016) performed source apportionment of groundwater of an industrial zone in north India. PCA was utilized to distinguish the fundamental sources of contamination while, HCA was used to for clustering. Along these lines the distinctive sources of contamination like anthropogenic and geogenic were recognized**.**

Hopke (2016) assessed various developed mathematical data analysis techniques that can be connected to data sets to evaluate source apportionment. This techniques can be effortlessly utilized, there has been broad application to various sorts of information. Accordingly these enhanced techniques and their application to particular air quality contamination will give extra data in to source-receptor affiliation that will define productive air quality administration plans.

Gholizadeh et al., (2016) utilized receptor modeling techniques like APCS-MLR and PMF to evaluate the water quality and distinguish and measure the potential contamination sources influencing the water nature of 3 rivers in South Florida. The information network has included 16 sampling stations for 12 water quality parameters. Five and four potential contamination sources in wet and dry seasons were distinguished by PCA/FA. Also PMF and APCS-MLR allocated their contributions to each water quality variable. Effluents from farming waste, local and industrial wastewater were recognized as the real sources causing pollution in the river. The APCS-MLR receptor displaying approach was observed to be all the more physically compelling in distinguishing the real sources causing stream water pollution.

Zhang et al., (2017) examined the groundwater quality parameters in the Hutuo river which is located in fan district of northern China. Outcome of principal component investigation (PCA) uncovered three main sources. Utilizing PCA and APCS-MLR, it was demonstrated that household wastewater and horticultural overflow are the fundamental sources of groundwater contamination in the stream. In this way, the most fitting techniques to anticipate groundwater quality contamination were recognized which were to enhance capacities, with regards to wastewater treatment and to revise fertilization methods in the area.

Guo et al., (2017) distinguished groundwater contamination sources utilizing FA and PMF techniques in the Jinji groundwater source northwestern China. The source allotment with the PMF technique distinguished three predominant groundwater contamination sources and FA display recognized four sources. It was presumed that the most noteworthy characteristic of the PMF is its logical translation and physical clarification of the outcomes, which depends upon nonnegative confinement of the contamination source profiles and its contributions.

The summary of source apportionment using receptor models with respect to the type of variable used, the technique applied, study region and the type of pollution problem addressed are presented in Table 3.3(a) to 3.3(c).

Table 3.3(a): Summary of source apportionment models used for environmental monitoring

Sl. No	Pollution Type	No of Variables	Variable type	Technique	Result R^2	Country	Authors
1	Particulate Matter	16	Fine and coarse inhalable particle	APCS-MLR	> 0.8	USA	Thurston and Spengler (1985)
2	Bulk wet deposition	12	Strong acids	PMF	-	Finland	Anttila et al. (1995)
3	Volatile organic compounds (VOCs)	13	Volatile organic compounds	APCS-MLR, Unmix, PMF	0.93,0.73, 0.90	USA	Anderson et al. (2002)
4	Stream pollution	11	Trace Metals	FA-MLR	0.52 – 0.96	Turkey	Pekey et al. (2004)
5	River pollution	33	Physical, chemical, biological, organic and heavy metals	APCS-MLR	0.42 – 0.94	India	Singh et al. (2005)
6	Coastal water pollution	14	Physical, chemical and organic	APCS-MLR	-	Hong - Kong	Zhou et al. (2007)
7	Combined sewer overflows	30	Physical, chemical, biological, organic and heavy metals	CMB and PMF	-	USA	Soonthornnonda and Christensen (2008)

Table 3.3(b): Summary of source apportionment models used for environmental monitoring

Sl. No	Pollution Type	Number of Variables	Variable type	Technique	Result R^2	Country	Authors
8	Atmospheric heavy metal deposition	10	Lichens	PMF	0.996	Ghana	Boamponsem et al. (2010)
9	River pollution	13	Physico-chemical and heavy metals	Unmix	0.85	China	Huang et al. (2010)
10	River pollution	15	Trace metals	FA - MLR	0.65 – 0.94	China	Li et al. (2011)
11	River pollution	13	Physico-chemical and biological	APCS-MLR	-	China	Su et al. (2011)
12	Ground water pollution	14	Physico-chemical and biological	APCS-MLR	0.6	Italy	Ielpo et al. (2012)
13	River pollution	18	Physico-chemical and heavy metals	APCS-MLR	-	China	Chen et al. (2013)

Table 3.3(c): Summary of source apportionment models used for environmental monitoring

Sl.No	Pollution Type	Number of Variables	Variable type	Technique	Result R^2	Country	Authors
14	River pollution	11	Physical, chemical, biological and heavy metals	APCS-MLR	>0.7	China	Yang et al. (2013)
15	River pollution	18	Physico-chemical and heavy metals	PMF	-	Bangladesh	Bhuiyan et al. (2015)
16	River pollution	12	Physical, chemical, biological and organic	APCS-MLR, PMF	>0.7	USA	Haji Gholizadeh M et al. (2016)
17	Groundwater pollution	5	Heavy Metals	PCA, HCA	-	India	Bhutiani et al. (2016)
18	Groundwater pollution	16	Physico-chemical and heavy metals	APCS-MLR	0.34 – 0.98	China	Zhang et al. (2017)
19	Groundwater pollution	15	Physical, chemical, biological	FA - PMF	0.42 – 0.97	China	Guo et al. (2017)

APCS-MLR: Absolute principal component scores – Multilinear regression, PMF: Positive matrix factorization, CMB: Chemical mass balance, FA-MLR : Factor analysis – Multilinear regression

From Table 3.3(a) to Table 3.3(c), it is evident that different receptor models have been used for to quantify the source contribution to environmental pollution research around the globe.

- Among the several techniques applied, the important ones are APCS-MLR, Unmix and Positive Matrix factorization.
- It implies that these techniques are used successfully and can be employed to can be employed to assess contributions from different sources quantitatively based on observations at sampling sites which will help researchers establish priorities for sustainable water management.
- Large number of applications were to surface water across the globe
- In the Indian context the application of these techniques to water quality problems, especially with respect to groundwater is limited in number. Source apportionment of pollution sources to groundwater quality using receptor models can be explored further.
- Comparisons of different models can be studied to understand the choice of source apportionment techniques.

3.5 LITERATURE SUMMARY

- Various studies have been reported in the domain of groundwater quality appraisal. The importance of having a proper understanding of the existing conditions and the factors affecting water quality for the efficient management of water resources cannot be disputed. For this purpose, various methods have been adopted by various researchers, of which multivariate techniques have proved to be quite successful.
- The ability of multivariate techniques to handle multiple parameters and large volume of data together to give a wholesome idea about the existing conditions makes it highly suitable for the assessment of water quality data.

- Amid the various multivariate techniques, Cluster Analysis (CA), Principal Component Analysis (PCA)/ Factor Analysis (FA), Discriminant Analysis (DA) and APCS-MLR emerged as the most suitable methods for water quality assessment. Each of these methods has a different purpose and provides different type of information about the water quality.
- CA is found to be useful in detecting spatial and temporal patterns in water quality and thereby classifying the data into different types based on their similarities/dissimilarities. PCA and FA are useful in identifying the underlying factors affecting the water quality and also help in source identification, DA helps in determining the most significant parameters affecting the spatial and temporal variation in water quality, while receptor oriented source apportionment modeling can be used for determining the source contributions to the various parameters quantitatively.
- Most groundwater pollution investigation and research depend on general physical and chemical portrayal of subsurface waters and gives data just with respect to regardless of whether a particular sample from a specific zone is being polluted. It is troublesome, if certainly feasible, to recognize the sources of the pollution or allocate different pollution sources influencing a similar water body. In this way Receptor models can be utilized to evaluate contributions from various sources in light of continuous monitoring at inspecting sites.
- Unmix takes care of the general blend issue where the information is thought to be a linear mix of an unknown number of sources of unknown configuration, which contribute an unknown add up to every sample while PMF weights information focuses by their analytical vulnerabilities, constrains factor loadings and factor scores to non-negative values and in this way minimizing the uncertainty created by rotating the factors.

- Having the capacity to recognize distinctive contamination sources precisely is a key component in a compelling water quality administration framework. Subsequently characterization of spatial variation and source apportionment of water quality variables can give an enhanced comprehension of the ecological conditions and help scientists/researchers build up needs for sustainable water administration. Subsequently portrayal of spatial differences and source apportionment of water quality factors can give an enhanced comprehension of the natural conditions and enable scientists to set up needs for reasonable water administration.

Chapter to follow discussed the materials and methodology adopted in order to achieve the objectives of research.

CHAPTER 4

MATERIALS AND METHODOLOGY

4.1 INTRODUCTION

This chapter illustrates the proposed methodology followed to achieve the objectives of the study. Groundwater frameworks are dynamic in nature and modify ceaselessly to short and long term changes in atmosphere, groundwater withdrawal and land utilize. Water quality estimations from wells regularly called 'groundwater quality observing', gives imperative and truly necessary data about the progressions the aquifer experiences and how these progressions influence the groundwater quality in a specific district. Long haul methodical groundwater observing gives basic information expected to assess the changes in the asset additional time, to create groundwater models and anticipate patterns.

4.2 OVERALL METHODOLOGY

The present study aims to understand the sources of groundwater contamination, spatial and temporal variations in water quality and apportioning the sources of groundwater pollution, in the Peenya industrial region of Bangalore Urban District of Karnataka, India. Multivariate analysis (CA, PCA) of the data was carried out using SPSS® 20.0 software. DA was carried out using Statistica® 10.0 software. Receptor modeling was carried out by applying multi-linear regression on the absolute principal component scores (APCS-MLR) Unmix and Positive Matrix Factorization (PMF) models. Finally the performance evaluation of the models used was carried out. The overall methodology adopted in the research is shown in Fig. 4.1.

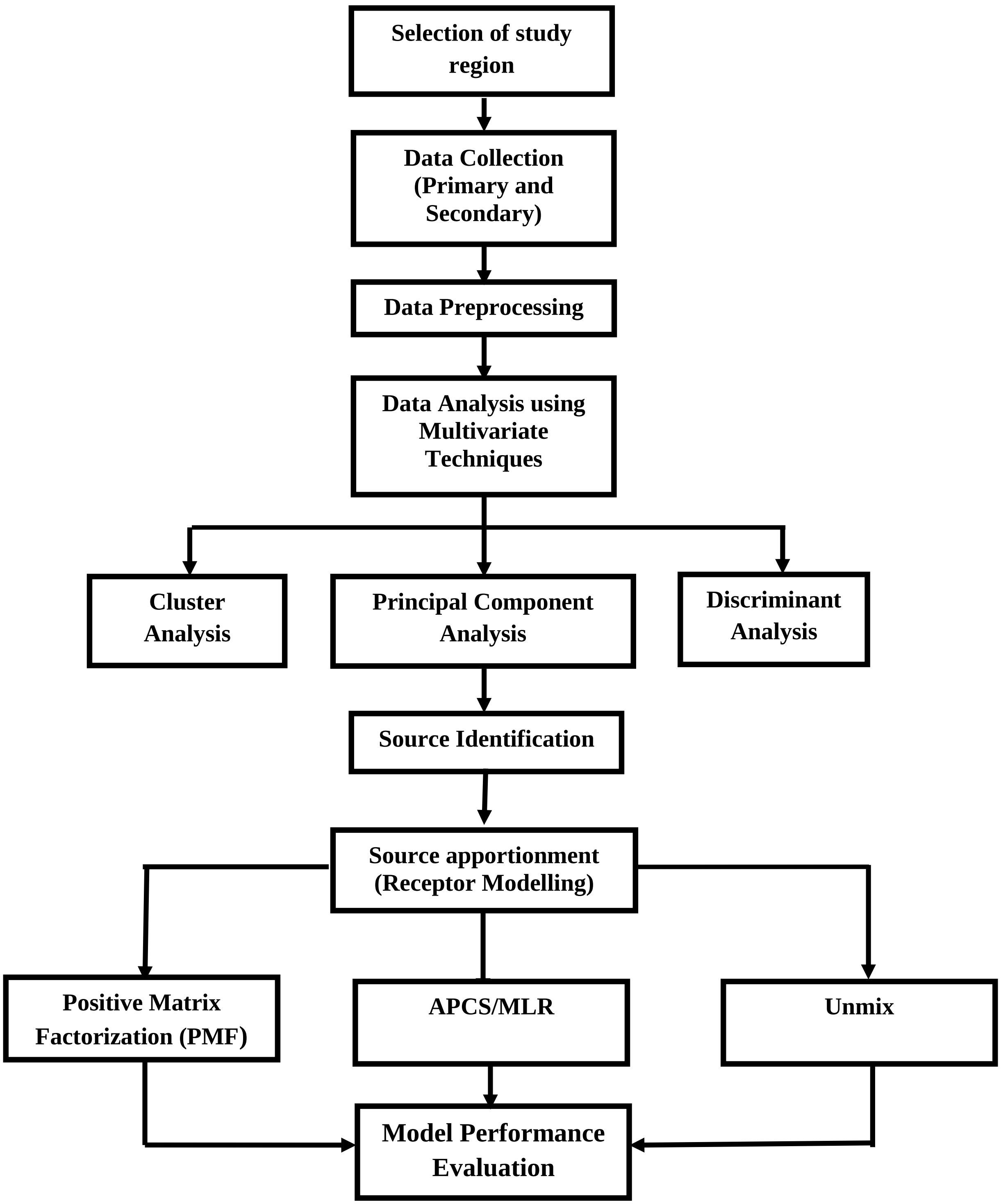


Fig 4.1: Overall Methodology

As part of the methodology some of the activities involved are

- Selection of study area,
- Primary and Secondary Data Collection
- Data Preprocessing
- Application of Multivariate techniques such as Cluster Analysis , Principal Component Analysis and Discriminant Analysis
- Source Identification
- Source Apportionment using PMF, APCS/MLR and Unmix

4.3 LOCATION OF STUDY AREA

The present study has been carried out in the Bangalore Urban district located in the southeastern part of Karnataka (Fig.4.2). The district has a geographical area of 2174 sq.km. According to the census carried out in 2011, the total population of the district is 95,88,910 along with a density of 4378 persons per sq.km. The study area is Peenya industrial area which is situated in the north-western part of Bangalore between $13^{o}1'42''$N and $77^{o}30'45''$E. Peenya is known a one of the most established and biggest industrial zones in the whole south-east Asia. The industrial estate of Peenya was built up in the late 1970s by the Karnataka Small Industries Development Corporation (KSSIDC) in three phases. Later the Karnataka Industrial Area Development Board (KIADB) developed Peenya Industrial territory in four Phases and comprises of more than 2600 industries. The location map of the study area is shown in Fig. 4.2.

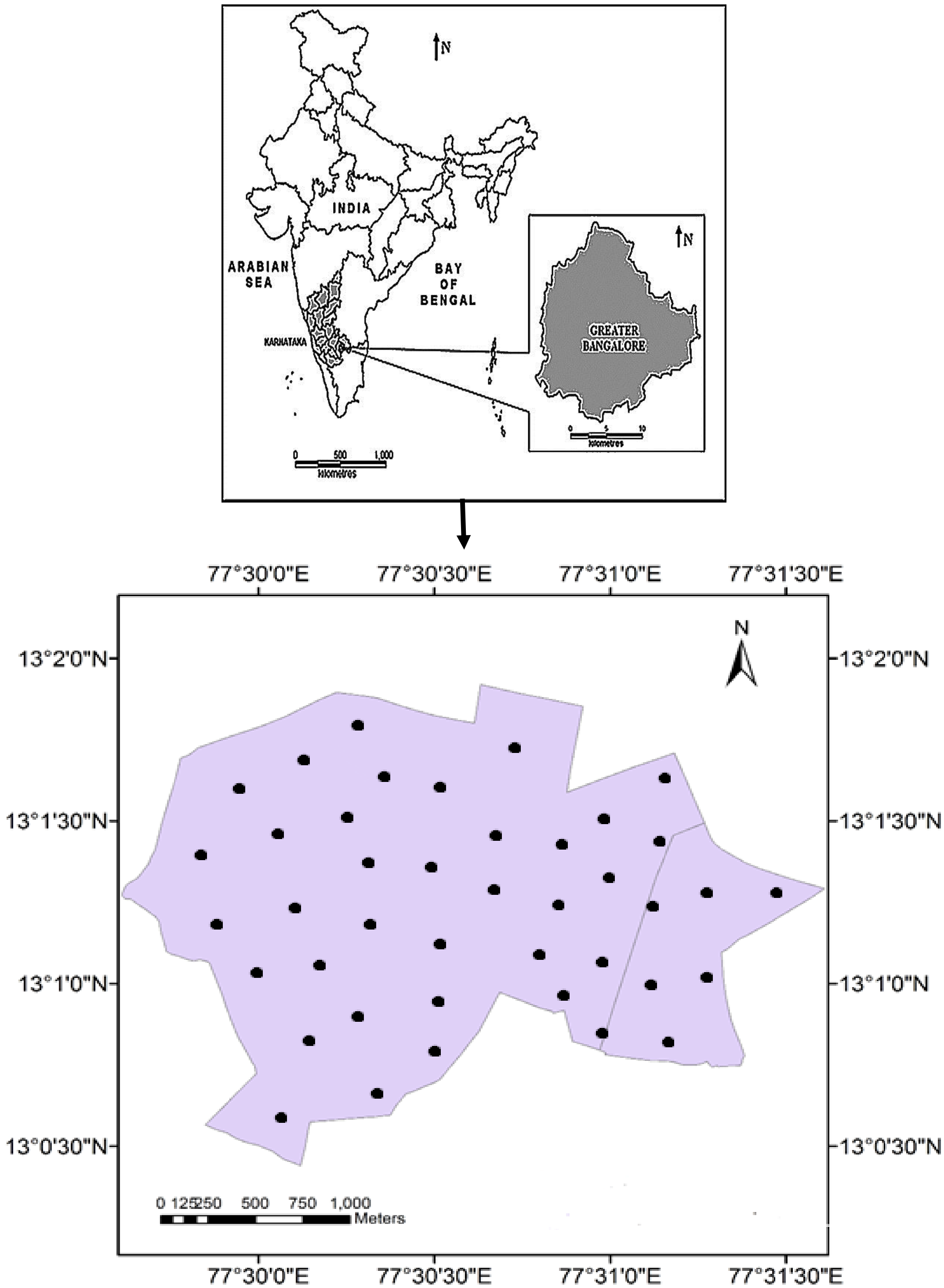


Fig 4.2: Study Area

4.3.1 Justification for Selection of Study area

The study area is enclosed by residential and private diverse industrial activity. There is no buffer zone existing between designated Peenya Industrial Area/estate and surrounding area. The industrial area houses many type of industries. As such these industries are likely to generate large quantities of wastes both liquid and solid which are causing groundwater pollution.

This site is considered as plausible site of contamination to recognize the contaminated bore wells, whose groundwater is polluted well beyond permissible limits for drinking water.

Central Groundwater Board (1999) performed various investigations on pollution caused by industries in Bangalore city, which covered the prominent industrial sectors for 80 km^2 area. It was outlined that the groundwater quality was marginally alkaline in nature. The domination by calcium and magnesium as cations and chlorides and nitrates as anions was also observed. Non-potability in the range of 12.5 to 50% was observed in groundwater of these regions with higher concentrations of nitrate also.

An examination by the Department of Mines and Geology, Government of Karnataka which studied about the condition of groundwater quality in Bangalore (2003) uncovered that 50.34% of the groundwater was found not found potable. Nitrate pollution was as high as 747 mg/L in a few areas with the permissible limit being 45 mg/L. Bacteriological pollution was tested independently at 100 areas, and contamination was found in 74 of the 100 specimens.

The fact that as far back as 1999 and 2003, groundwater quality was this terrible underlines the earnestness of the groundwater quality issue.

4.3.2 Major activities causing groundwater pollution

The industries/commercial enterprises which are notable from water contamination perspective are engineering with surface treatment, painting, pickling, electroplating, drugs, pesticides, clothing and textiles. Apart from these, effluents from industries in the unorganized sectors situated around the industrial area and domestic sewage are likewise major sources of pollution of this area.

4.3.3 Climate

The precipitation of the district comprises of the Pre monsoon (Mar-May), South-West monsoon (Jun-Sept) and North-East monsoon (Oct-Dec). The yearly precipitation of the district is 1049 mm with most of the precipitation being contributed by the South-West monsoon. Generally, humid to semi-arid climatic conditions prevail in the district. When the rainfall is deficient, there is a risk of higher concentration of surface pollutants getting infiltrated into the groundwater. Above average rainfall increases the dilution effect of rainfall recharge.

4.3.4 Hydrogeology and Drainage

Geo-morphologically, the Bangalore Urban district can be divided into three physiographic units as rocky uplands, plateau & flat topped hills at a general elevation of about 950m above MSL. Granites and Gneisses of peninsular gneissic group form the primary aquifers in the study district (Fig. 4.3). Laterites of Tertiary age occur as isolated patches capping crystalline rocks in the Bangalore north taluk. Groundwater occurs in phreatic conditions or unconfined conditions in the weathered zone and under semi-confined to confined conditions in fractured and jointed rock formations. The versatility, presence and aquifer refill of groundwater event is dominated by the measure of weathering, fracture pattern, geomorphological setup and rainfall.

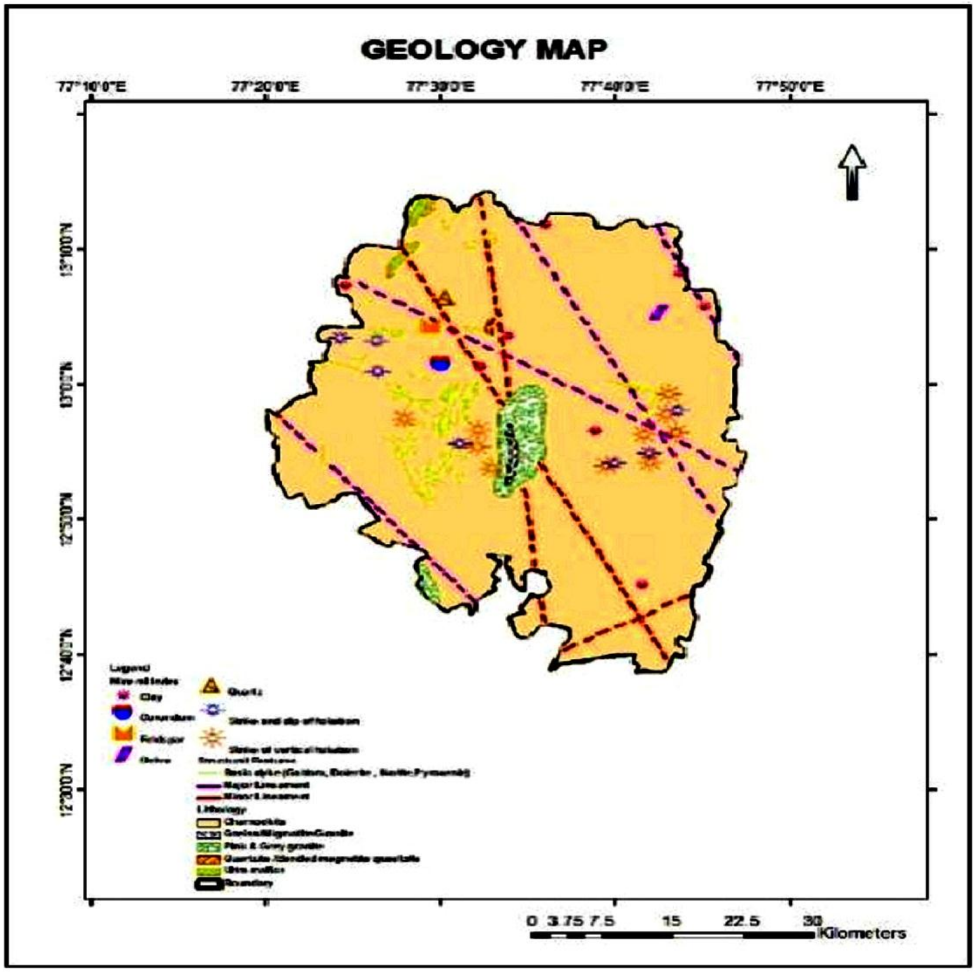


Fig 4.3: Geology Map of Bangalore urban District (CGWB 2012)

The aquifers in Bangalore urban district can be categorized into 3 zones i,e shallow zone, moderately deep zone and deep zone as shown in Table 4.1. Data collected from the Bore wells collectively coming from all the three aquifers.

Table 4.1: Aquifer types in Bangalore Urban District (CGWB 2012)

Aquifer type	Depth (bgl)	Yield
Shallow	upto 25m	1-2 lps
Moderately deep	upto 60m	2-6 lps
Deep	beyond 60m	2-8 lps

The study area is on a water divide with the area sloping towards west. Streams of various watersheds begin from this locale. Significant piece of the investigation zone is involved by streams streaming towards west from this zone (Fig. 4.4). A couple of surface water bodies or tanks are available in the territory yet are outside the limit of modern zone.

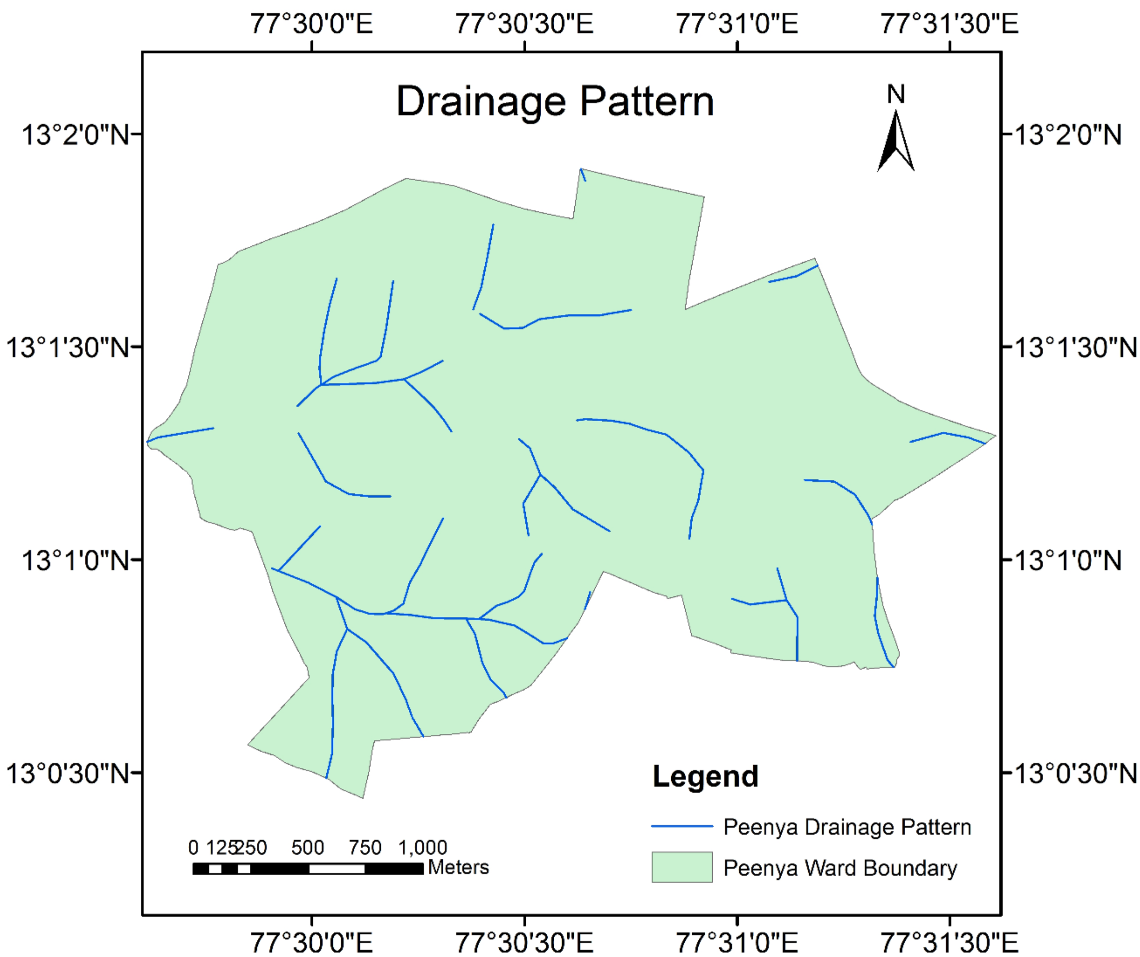


Fig 4.4: Drainage Map of Peenya, Bangalore.

4.3.5 Water supply demand and groundwater dependence

In the previous 2 decades the Bangalore city has seen fast development in the field of Industrial, Commercial and Institutional divisions. This unplanned development has caused major urban exercises encompassing Bangalore. Fast development in population is putting huge weight on framework, particularly on water supply and underground seepage framework. The anticipated water supply figures for the city of Bangalore and population projections are shown in the Table 4.2.

Table 4.2: Population and Water supply requirements (BWSSB 2016)

Year	Population (Million)	Water Demand (MLD)	Water Demand (TMC)	Present Supply		Shortfall in Demand	
				MLD	TMC	MLD	TMC
2011	8.499	1400	18.05	950	12.25	450	5.80
2021	10.581	2100	27.1	1450	26.7	650	0.4
2031	14.296	2900	37.39	2070	26.7	1450	10.69
2041	17.085	3400	43.84	2070	26.7	1950	17.14
2051	20.561	4100	52.86	2070	26.7	2650	26.16

From Table 4.2, it can be inferred that, the demand supply gap is met by groundwater exploitation, which leaves the present deficit of around 450 MLD to be pumped from groundwater sources. So there is a tremendous pressure on this natural resource to meet domestic, commercial and industrial water supply demands.

4.4 DATA COLLECTION

To have firsthand information on the quality of groundwater, sampling was done and analysis was carried out by the author for 14 physico-chemical parameters namely pH , Electrical conductivity(EC), Total Dissolved solids(TDS), Calcium(Ca^{2+}), Magnesium(Mg^{2+}), Sodium(Na^{+}), Potassium(K^{+}), Iron(Fe), Alkalinity (HCO^{-3}), Chloride(Cl^{-}), Nitrate(NO_3^{-}), Sulphate(SO_4^{2-}), Total hardness(TH) and Fluoride(F^{-}) from 67 sites distributed across the western half of the city region for pre-monsoon and post monsoon during the year 2013. Box plots were constructed for all the 14 parameters and are shown in Fig. 4.4.On analyzing the data it was revealed that, most of the parameters were found to exceed the specified desirable limits while few parameters were found to exceed the permissible limits as well.

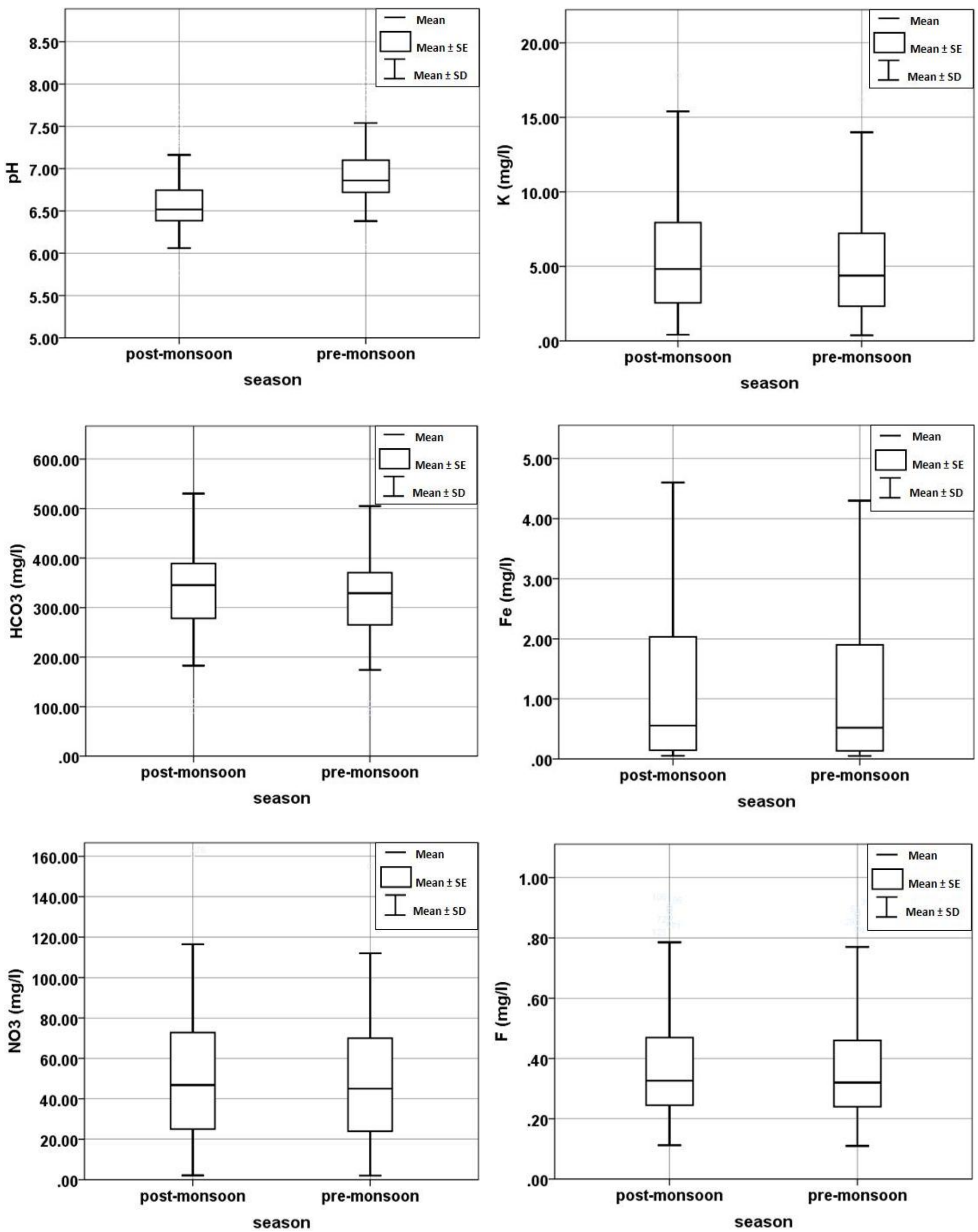


Fig. 4.4(a): Box plots showing temporal variation of pH, K HCO_3, Fe, NO_3 and F

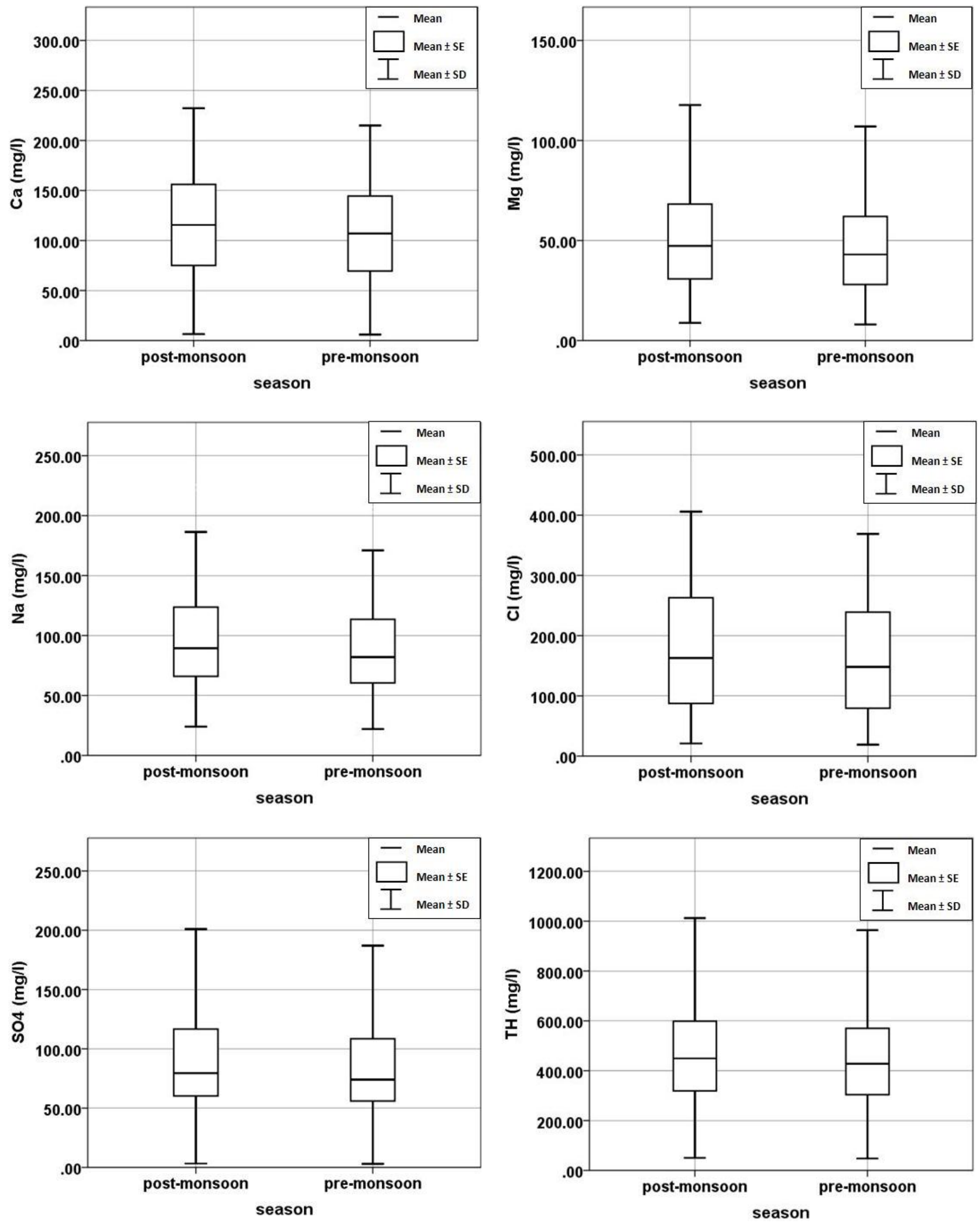


Fig. 4.4(b): Box plots showing temporal variation of Ca, Mg, Na, Cl, SO_4 and TH

It can be observed that the concentration of each parameter is found to be slightly higher in the post-monsoon season as compared to the pre-monsoon season. Rainfall data published by Indian Meteorology Department (IMD) revealed that Bangalore region received comparatively higher pre-monsoon rainfall and normal monsoon rainfall while the rainfall in the post-monsoon season was about 31% deficient in the year 2013, when the data was collected. Thus the dilution effect of rainfall recharge is observed to be higher in the pre-monsoon season.

4.4.1 Chemical Analysis

The groundwater samples were collected from bore wells after 10 min of pumping and transferred into pre-cleaned polyethylene bottles and stored at 10 °C. Electrical conductivity and pH were measured in the field immediately after sampling and the remaining parameters were determined in laboratory within 24 h. The analyses for various chemical parameters to assess the groundwater quality were carried out using standard procedures (APHA 2005). Calcium and Magnesium were determined by EDTA titrations method, Sodium and Potassium by flame emission photometry, Iron by Phenanthroline Spectrophotometry, Bicarbonate and Carbonate by Titrimetry, Chloride by Argnetometric Titration, Nitrate by UV Spectrometry, Sulphate by Nephelometry, Total Dissolved solids by Gravimetry, Total Hardness by Potentiometry, Fluoride by Ion Selection Electrode method.

4.4.2 Groundwater Quality Index

To decide the reasonableness of the groundwater for drinking purposes, water quality index was computed for the groundwater quality dataset using equation developed by Tiwari & Mishra(1985) ground water quality index of the region was calculated. From the groundwater quality index computation it is observed that the number of samples rated as poor, very poor and unfit constitute to about 50% of the total samples.

The computed WQI values range from 19 to 145 and 24 to 164 for pre-monsoon and post-monsoon, respectively. Out of 67 groundwater quality data points 24 stations (35%) falls in the "excellent" category, 16 stations (23%) in "good" category, 18 stations (26%) in "poor" category and 7 stations (10%) in "very poor" category and 3 stations (4%) in unfit category for pre-monsoon season. During post-monsoon, 20 stations (30%) falls in the "excellent" category, 16 stations (23%) in "good" category, 16 stations (23%) in "poor" category and 9 stations (13%) in "very poor" category and 5 stations (7%) in unfit category for post-monsoon dataset. Consequently the general groundwater quality in the region was observed to be less than desirable.

It was also evident that the physico-chemical groundwater quality of the region was of poor water quality. Further to find out the possibility of heavy metals contamination, groundwater quality data was obtained from Karnataka State pollution Control Board (KSPCB) on 20 parameters from 41 sampling stations (borewells) and subsequently was used in the study. The data collected was for the year 2015, and contained monthly sampled information from the month of January to December for the 20 physico-chemical and heavy metal parameters.

4.4.3 Parameters and Analysis

Parameters collected were, pH , Turbidity, Total Dissolved solids(TDS), Calcium(Ca^{2+}), Magnesium(Mg^{2+}), Iron(Fe), Alkalinity(HCO^{-3}), Chloride(Cl^{-}), Nitrate(NO_3^{-}), Sulphates(SO_4^{2-}), Total hardness(TH) ,Fluoride(F^{-}), Sulphide(S^{-}), Lead(Pb), Copper(Cu), Hexavalent Chromium(Cr), Zinc(Zn), Manganese(Mn), Cadmium(Cd). The groundwater samples were collected from bore wells after 10 min of pumping and transferred into pre-cleaned polyethylene bottles and stored at 10 °C. Electrical conductivity and pH were measured in the field immediately after sampling and the remaining parameters were determined in laboratory within 24 h. The detection levels for each parameter and method of analysis carried out for the data by the Karnataka State Pollution Control Board is indicated in the Table 4.3.

Table 4.3: Physico-Chemical parameters, analytical methods and detection limits

Sl. No.	Parameters	Methods	Detection Limits
1.	pH	Electrometric	0.01 pH unit
2.	Total Dissolved Solids	Gravimetric	2.0 mg/L
3.	Sulphate	Turbidity	0.1 mg/L
4.	Chloride	Argenrtometric	0.1 mg/L
5.	Iron	Instrument (AAS)	0.02 mg/L
6.	Nitrate	Specific Ion meter	0.05 mg/L
7.	Total Hardness	EDTA Titration	1.0 mg/L
8.	Calcium	EDTA Titration	1.0 mg/L
9.	Magnesium	Calculation	1.0 mg/L
10.	Fluoride	Spands	0.01 mg/L
11.	Turbidity	Turbidity	0.1 NTU
12.	Alkalinity	Titration	1.0 mg/L

Table 4.4: Heavy Metals, analytical methods and detection limits

Sl. No.	Parameters	Methods	Detection Limit
1	Lead	Instrument (AAS)	0.05 mg/L
2	Copper	Instrument (AAS)	0.01 mg/L
3	Hexavalent Chromium	Colorimetric	0.001 mg/L
4	Zinc	Instrument (AAS)	0.005 mg/L
5	Cadmium	Instrument (AAS)	0.002 mg/L
6	Manganese	Instrument (AAS)	0.01 mg/L

The data obtained was multivariate in nature. So in order to process the large amount of data and to report overall trends from it, multivariate statistical techniques were applied for the organization, analysis, interpretation and presentation of sample data. The same is discussed next.

4.5 MULTIVARIATE STATISTICAL METHODS

Statistical analysis is the study of collection methods which includes planning, designing, collecting data, analyzing, drawing meaningful interpretation and reporting of the research findings. Multivariate statistical analysis involves examination and interpretation of more than one variable at a specific time. Multivariate analysis worries about various points and background of each of the distinctive types of multivariate investigation and how they relate with each other. Multivariate Statistics include univariate and multivariate investigation keeping in mind the end goal to comprehend the connections amongst variable and their pertinence to the real issue being concentrated on.

Among the several techniques applied, the important ones are APCS-MLR, Unmix and Positive Matrix factorization. These techniques are used successfully and can be employed to can be employed to assess contributions from different sources quantitatively based on observations at sampling sites which will help researchers establish priorities for sustainable water management. In the Indian context the application of these techniques to water quality problems, especially with respect to groundwater is limited in number. Hence source apportionment of pollution sources to groundwater quality using receptor models can be explored further. Also comparisons of different models can be studied to understand the choice of source apportionment techniques.

4.5.1 Data Pretreatment

Checks for missing data and outliers revealed that there were no missing values in the data or significant outliers. Skewness and kurtosis was analyzed in order to check the normality of the data. Data standardization is very important in multivariate analysis as it ensures that all the parameters are close with respect to their variances. Standardization minimizes the effect of variance difference on variables and removes the effect of different units of measurement, thus producing a dimensionless number (z-scores) (Güler et.al., 2002, Cloutier et.al., 2008, Yidana 2008, Yidana 2011).

In order to attain normal distribution and uniformity, the data was standardized corresponding to their z-scores as in Equation 4.1.

$$z = \frac{x - \bar{x}}{s} \qquad 4.1$$

Where x represents the value, $\bar{x}$ represents the mean and s represents the standard deviation of the parameter, at a given sampling site.

4.5.2 Cluster Analysis

Cluster analysis is a gathering of multivariate statistical techniques whose basic role is to assemble objects in view of the attributes they have. CA is one of the multivariate techniques, which; groups the objects based on their characteristics. It arranges the objects, such that every object is same as the others in the cluster according to a predefined selection criterion. The clusters of objects obtained should then display high internal (within-cluster) resemblance and high external (between clusters) diversity. The results of CA help in interpreting the data and indicate patterns.

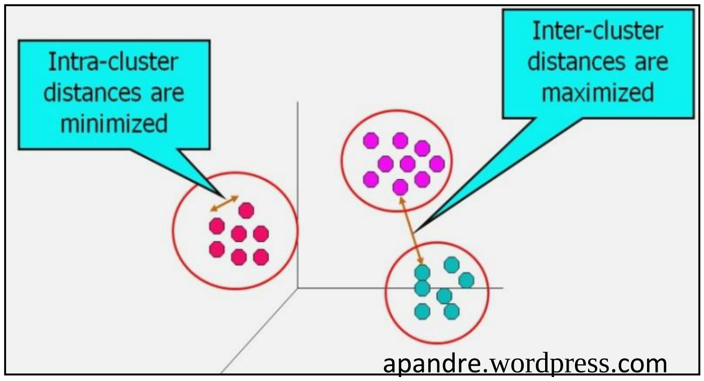


Fig 4.5: Cluster Analysis

Distinctive sorts of cluster analysis techniques have been utilized to evaluate water quality information for both surface and groundwater. Hierarchical agglomerative clustering is the most commonly used approach which; supplies with instinctive similarity relationships between any one sample and the entire data set. It is represented by a dendrogram (tree diagram).

Hierarchical Cluster Analysis (HCA) is an efficient means to recognize groups of samples that have similar chemical and physical characteristics. Clusters are formed in a sequential manner with the most comparative combine of objects and framing higher clusters in a step by step manner. The two key strides inside cluster analysis are the estimation of distances amongst objects and to aggregate the objects in light of the outcomes in the distances (linkages). The distances accommodate a measure of likeness amongst objects and might be measured using different methods, for example, Euclidean and Manhattan metric distance. Linkages depend upon the type of measurement happens within the group. Different types of linkages are:

1. Simple linkage or Nearest neighbour distance - It measures the distance to the nearest object in a group.
2. Complete linkage or Furthest neighbour distance - It measures the distance between furthest objects.

3. Average between group linkage - It is based upon the distance from all objects in a group.
4. Centroid linkage – It has a new value, representing the group Centroid, which is compared to the ungrouped point to weigh inclusion.
5. Ward's method – It is a variance based method with the groups variance assessed to enable clustering. The group which sees the smallest increase in variance with the iterative inclusion of a case will receive the case.

Ward's method is a popular default linkage which produces compact groups of well distributed size. Standardization of variables is undertaken to enable the comparison of variables to minimize the bias in weighting which may result from differing measurement scales and ranges. The ward's method makes use of an analysis of variance approach for evaluating the distances between clusters, in order to minimize the sum of squares (SS) of any two clusters that can be formed at each step

HCA with squared Euclidean distance as a similarity measure and Ward's method as a linkage measure has been determined as the best combination for revealing the most unique set of spatial sample associations (Güler et.al., 2002, Cloutier et.al., 2008, Yidana 2011).

4.5.3 Discriminant Analysis

Discriminant analysis determines the variables that discriminates between two or more expected occurring groups. The purpose of DA is to increase the similarity between-group relative to the within-group variance It constructs a discriminant function (DF) for each group as in Equation (4.2):

$$\mathrm{f}(\mathrm{G_i}) = \mathrm{k_i} + \sum_{j=1}^{n} \mathrm{w_{ij}} \mathrm{p_{ij}} \tag{4.2}$$

Where i is the number of groups (G), k_i the constant inherent to each group, n the number of parameters used to classify a set of data into a given group, w_j the weight coefficient assigned by DA to a given selected parameter (p_j) (Yidana 2008).

The main use of discriminant analysis is to predict group membership from a set of predictors. Accordingly, an attempt is made to delineate based upon maximizing between group variance while minimizing within group variance. DA may be performed in three different modes:

1. Standard mode – In this mode, Discriminant functions are constructed using all parameters.
2. Forward stepwise mode - In this mode, variables are added step by step beginning with the most significant until no significant changes are obtained.
3. Backward stepwise mode - Here the variables are eliminated in a well ordered start with the less significant until the point that no noteworthy changes are acquired.

4.5.4 Principal Component Analysis/Factor Analysis

Principal component analysis supplies the details of most significant parameters, which describes the whole data set, thereby reducing the data with minimal loss of original information (Wunderlin et al. 2001). PCA is a technique, which; converts the original variables into new uncorrelated variables (axes), known as principal components, which; are linear combinations of the original variables

The principal component (PC) is expressed as in Equation (4.3):

$$z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}x_{3j} + \cdots + a_{im}x_{mj} \tag{4.3}$$

where a is the component loading, z the component score, x the measured value of a variable, i the component number, j the sample number, and m the total number of variables (Singh et al., 2005).

Normality provides for an enhanced solution, but some inference may still be derived from non-normal data. Rotation tries to put the principal components (PCs) in an easier position with respect to the original variables, which helps in the translation of factors. Varimax, quartimax, and equimax are all orthogonal rotations, while oblique rotations are non-orthogonal. The most commonly used type of rotation is the varimax rotation which maximizes the variance of the loading.

In order to carry out PCA, there should be a certain redundancy between the variables that can be summarized with a few factors. Hence before performing PCA, the data should be checked using the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) and the Bartlett's test of sphericity.

The KMO index compares the values of correlations between variables and those of the partial correlations. If the KMO index is high (0.6-0.9), PCA can act efficiently; if KMO is low (< 0.5), PCA is not relevant.

The number of factors/PCs to be extracted from PCA is determined based on the Kaiser criterion and the Scree plot. As per the Kaiser criterion, the factors/PCs, whose Eigen values are greater than 1 are retained. Factors with Eigen values less than one are ignored as it accounts for less variance than an original variable. The Scree plot is a graph of magnitude of Eigen values (Y axis) versus the factors/PCs (X axis). Based on this, the factors/PCs which are above the inflection point of the slope are extracted.

PCA is continued with factor analysis. The main purpose of FA is to reduce the contribution of less significant variables to simplify even more of the data structure coming from PCA. This purpose can be achieved by rotating the axis defined by PCA, according to well established rules, and constructing new variables, also called varifactors (VF). In FA, the basic concept is expressed as in Equation 4.4.

$$z_{ji} = a_{f1} f_{1i} + a_{f2} f_{2i} + a_{f3} f_{3i} + \cdot \cdot \cdot + a_{fm} f_{mi} + e_{fi} \tag{4.4}$$

where z is the measured value of a variable, a the factor loading, f the factor score, e the residual term accounting for errors or other sources of variation, i the sample number, j the variable number, and m the total number of factors.

Further to find out the source contributions from the possible sources by principal component analysis, source apportionment techniques were applied. The same is discussed next.

4.6 SOURCE APPORTIONMENT

Source Apportionment (SA) is the demonstration of surmising information about contamination sources and the sum they add to surrounding contamination levels. The customary methodology is dispersion modeling, in which a pollutant emission rate and meteorological information serve as an input to a mathematical model that scatters the emitted pollutant, producing an expectation of the subsequent resulting pollutant concentration at a point in space and time. The other method is receptor modeling, which is a predefined mathematical procedure for distinguishing and evaluating the sources of ambient air/water contaminants at a receptor fundamentally on the premise of concentration measurements at that receptor. These are the two basic approaches to determine the sources of pollution:

(1) Top-down or Receptor based source apportionment methods, and

(2) Bottom-up or Source-based methods.

The top-down approach begins by taking samples in a given area and comparing the chemical and physical properties of the sample to the properties of emission sources. Top down strategies offer the guarantee of giving data on the sorts of sources of emmisions and their relative commitments to measured contamination, which thusly distinguishes and evaluate the sources that would be best to control. The top-down methodology starts by taking samples in a given zone and looking at the physical and chemical properties of the sample to the properties of emission sources.

Bottom up strategies start by recognizing contamination sources and evaluating the factors causing emission by utilizing the dispersion models. Using this data and point by point meteorological information an environmental dispersion model assesses the surrounding contamination levels.

The receptor oriented source apportionment modelling techniques used in the study are discussed next in detail.

4.6.1 Receptor Oriented Modeling

A general receptor-oriented model is based on the assumption that the total concentration of each contaminant is made up of the linear sum of elemental contributions from each of the *j* pollution source components collected at the receptor site and can be expressed mathematically as in Equation 4.5:

$$\mathrm{Z_{jk}} = \sum_{j=1}^{p} \mathrm{w_{ij} p_{jk}} \tag{4.5}$$

where z_{jk} is the normalized concentration of contaminant (variable), *j* the number of pollution sources, w_{ij} the factor loadings, the coefficient matrix of the components relating the pollution sources to their elemental concentrations; and p_{jk} the factor scores, the value of the j^{th} source's components on observation *k.* Both w_{ij} and p_{jk} are dimensionless.

The most common methods utilized for receptor modeling are

- Absolute Principle Component Scores-multiple Linear Regression (APCS-MLR)
- Unmix
- Positive Matrix Factorization (PMF).

A simplified description of receptor modeling is depicted in Fig. 4.4.

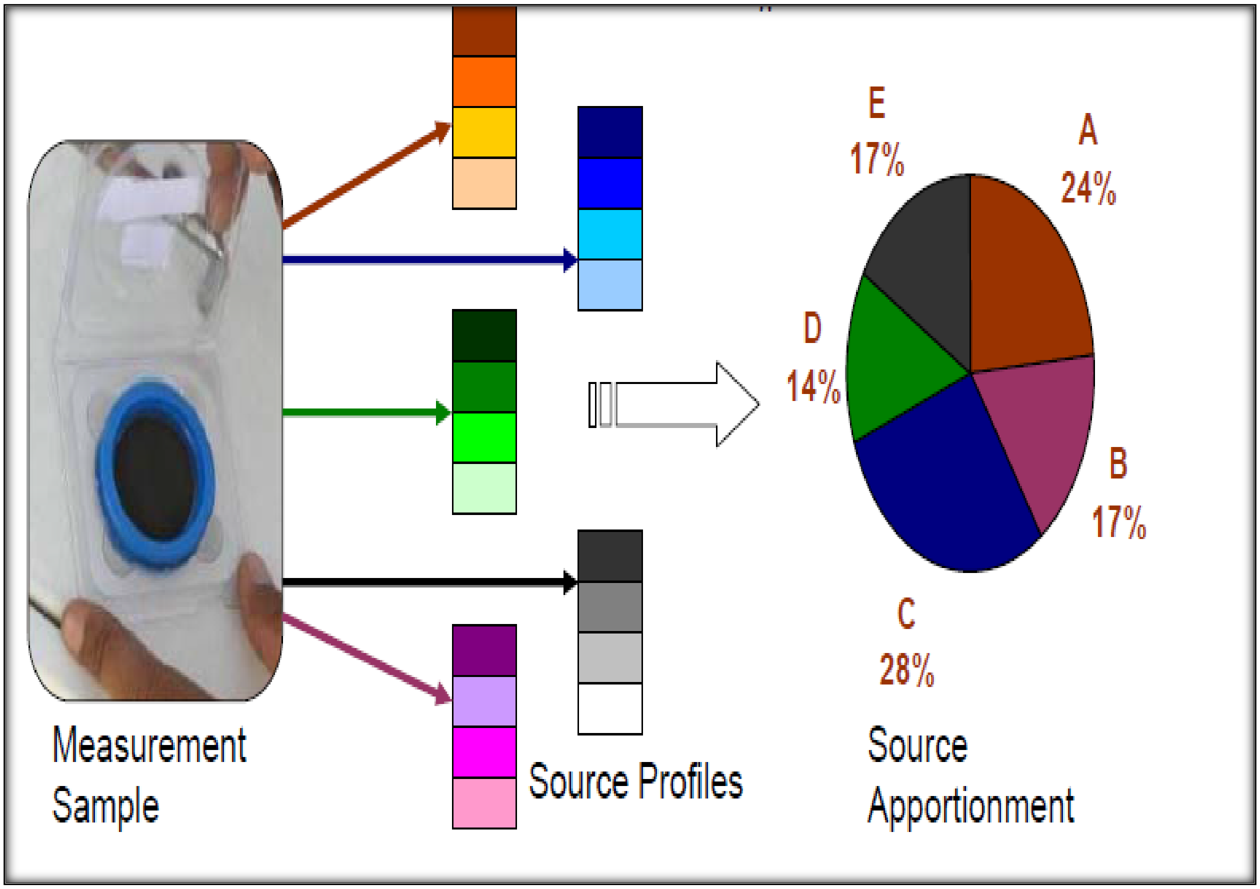


Fig 4.6: Depiction of receptor modeling

4.6.2 Absolute Principle Component Scores-multiple Linear Regression (APCS MLR)

Source apportionment of environmental contaminants can be carried out using receptor modeling approach which is based on multi-linear regression of the absolute principal component score (APCS-MLR). It can be expressed mathematically as shown in Equation 4.6.

$$Z_{jk} = k_i + \sum_{j=1}^{p} w_{ij} p_{jk} \tag{4.6}$$

Where Z_{jk} is the normalized concentration of contaminant (variable), j the number of pollution sources, w_{ij} the factor loadings, the coefficient matrix of the components relating the pollution sources to their elemental concentrations; and p_{jk} the factor scores, the value of the jth source's components on observation k. Both w_{ij} and p_{jk} are dimensionless.

Since the normalized value of the variables cannot be directly used for computation of quantitative source contributions, the normalized principal component scores (PCS) were converted to un-normalized absolute principal component scores (APCS). This is done by subtracting the principal component score of a fictitious sample (true zero), with all concentrations as zero, from the principal component score of each sample (Thurston and Spengler, 1985) as in Equation (4.7).

$$(\mathrm{APCS})_{jk} = \mathrm{Z_{jk}} - Z_0 \tag{4.7}$$

The commitment from each factor would then be able to be assessed by multiple linear regression (MLR), utilizing the APCS values as independent variables and the measured concentration of the specific parameter as the reliant variable, as shown in Equation (4.8).

$$\mathrm{M_{jk}} = \mathrm{a_{i0}} + \sum_{\mathrm{j=1}}^{\mathrm{p}} \mathrm{A_{ij}}(\mathrm{APCS})_{\mathrm{jk}} \tag{4.8}$$

Where M_{jk} is the contaminant's concentration; a_{i0} the average contribution of the jth contaminant from sources not determined by PCA, A_{ij} the linear regression coefficient for the ith contaminant and the jth factor, and $(APCS)_{jk}$ the absolute factor score for the jth factor with the kth measurement. The values for M_{jk}, a_{i0} and A_{ij} have the dimensions of the original concentration measurements (Singh et al., 2005).

4.6.3 Unmix

Unmix is one of the receptor models that the United States Environmental Protection Agency´s Office of Research and Development (ORD) has developed. The hidden rationality of Unmix is to give the information a chance to justify itself with real evidence. The data in a general mixture problem is assumed to be a linear combination of an unknown number of sources of unknown composition, which contribute an unknown amount to each sample. Unmix also assumes that the compositions and contributions of the sources are all positive. Also, that for each source there are some samples that contain little or no contribution from that source. Using concentration data for a given selection of species, Unmix estimates the number of sources, source compositions, and source contributions to each sample. UNMIX uses the singular value decomposition (SVD) method to estimate the source number by reducing the dimensionality of data space m top (Henry, 2003). The UNMIX model can be expressed as in Equation (4.9).

$$C_{ij} = \sum_{I=1}^{p}\left(\sum_{k=1}^{p} U_{ik} D_{kI}\right) V_{Ij} + \epsilon_{ij} \tag{4.9}$$

where U, D, and V are n×p, p×p diagonal, and p×m matrices, respectively; and ε_{ij} is the error term consisting of all the variability in C_{ij} not accounted for by the first p principal components.

The methodology adopted in Unmix study is as shown in Fig. 4.4

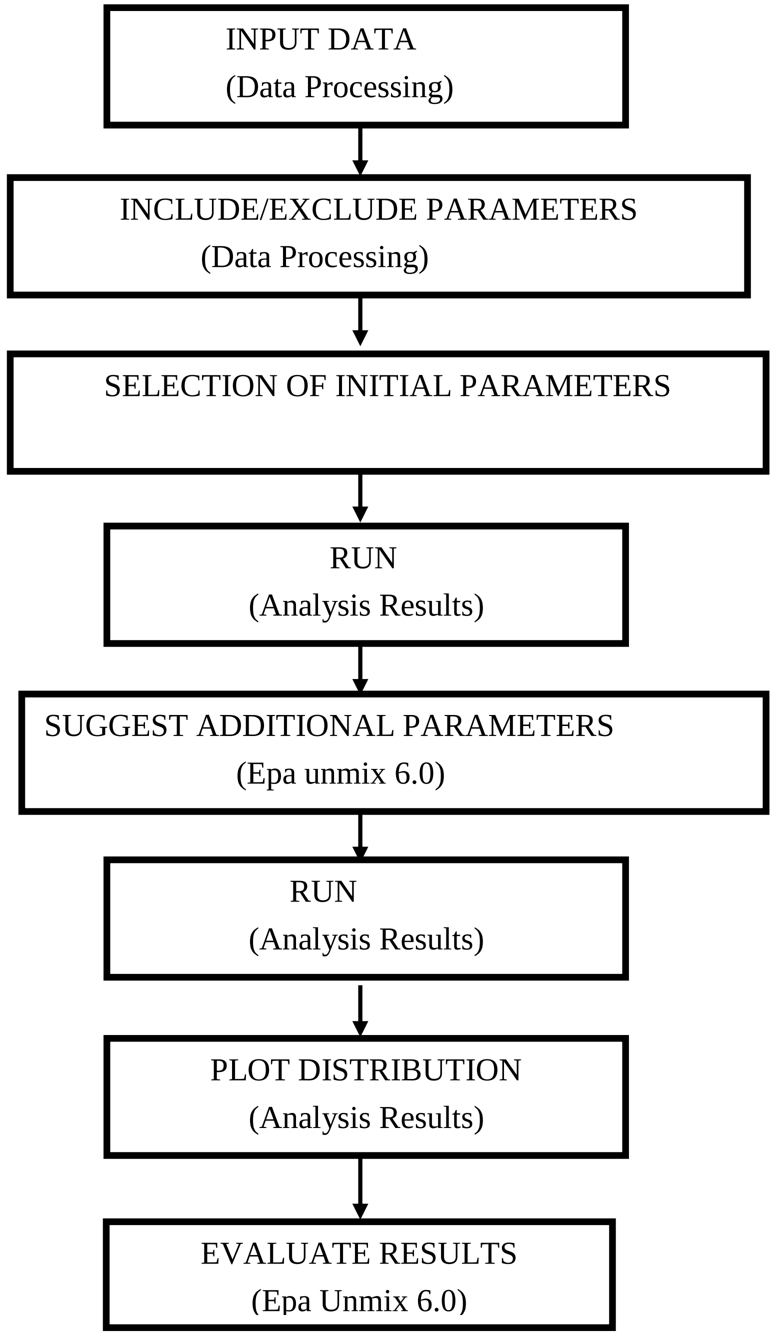


Fig 4.7: Unmix Results Evaluation process

4.6.4 Positive Matrix Factorization (PMF)

This model was developed by Paatero (Paatero and Tapper, 1993) and implemented by the EPA. PMF is a multivariate factor investigation tool that breaks down a framework of speciated test data into two lattices namely factor commitments (G) and factor profiles (F). These factor profiles should be translated to distinguish the source type that might be adding to the sample utilizing measured source profile data, and emission or release inventories.

The results are extracted utilizing the requirement that no example can have fundamentally negative source commitments. The PMF general receptor model assumes that there are p independent sources (factors) that contribute to a receptor and that linear combinations of these p factors give rise to the observed composition as shown in Equation (4.10).

$$x_{ij} = \sum_{k=1}^{p} g_{ik} f_{kj} + e_{ij} \tag{4.10}$$

where x_{ij} is the concentration at a receptor for the j^{th} species on the ith day, g_{ik} is the contribution of the k^{th} factor to the receptor on the i^{th} day, f_{kj} is the fraction of the k^{th} factor that is species j, and e_{ij} is the residual concentration for the j_{th} species on the i^{th} day.

The contributions of the elements are compelled to be nonnegative with the goal that the physical implications of the factor loadings and scores are all the more effectively translated. An uncertainity is expressly relegated for every observation. The PMF model then looks to limit the aggregate of squares of an object function, Q, in view of the uncertainities for every perception as shown in Equation (4.11)

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\frac{x_{ij} - \sum_{k=1}^{p} g_{ik} f_{kj}}{u_{ij}} \right]^2 \tag{4.11}$$

where s_{ij} is the uncertainty in the j^{th} species for day i.
Evaluation with the quantity of variables and rotation parameters is carried out until the point when the most sensible solution is acquired, i.e., the arrangement creates the most physically significant outcomes (Paatero 1997).

The most vital quality of PMF is that the user can allocate singular uncertainties to indicate the quality and certainty of every data point. Subsequently, it is probably going to lessen the heaviness of missing and below-detection-limit data in the least squares solution and furthermore to "downweight" any information having lower certainty. PMF utilizes both sample concentration and client furnished uncertainty related with the specimen information to weight each points.

Uncertainty associated with test species ought to envelop mistakes, for example, examining and investigative errors. Observation-based and equation-based uncertainty are the two sorts of uncertainty documents EPA PMF 5.0 acknowledges. This component enables examiners to represent the trust in the estimation. For instance, information below detection can be held for use in the model, with the related uncertainty balanced so these data points have less effect on the arrangement than estimations over as far as possible.

If the concentration is less than or equal to the MDL provided, the uncertainty (*Unc*) is calculated using Equation 4.12, a fixed fraction of the MDL (Polissar et al., 1998).

$$\text{Unc} = \frac{5}{6} * \text{MDL} \qquad (4.12)$$

If the concentration is greater than the MDL provided, the calculation is based on a user provided fraction of the concentration and MDL

$$\text{Unc} = \sqrt{(\text{Error Fraction} * \text{Concentration})^2 + (0.5 * \text{MDL})^2}$$

Some portion of the information readiness is to choose if an species categories should be prohibited. Explanations behind the expulsion of species incorporate the twofold counting of information. The signal-to-noise (S/N) ratios of the species are helpful for assessing their value. On the off chance that the S/N value is too low the species uncertainty is expanded with the goal that it has less effect on the outcomes. The associated equation for signal-to-noise is as shown in Equation (4.13):

$$\left(\frac{S}{N}\right)_j = {}^{1}\!/_{n} \sum_{i=1}^{n} d_{ij} \tag{4.13}$$

The outcome with this new S/N estimation is that species with concentrations dependably underneath their uncertainty have a S/N of 0. Species with concentrations that are double the uncertainty value have a S/N of 1. S/N more noteworthy than 1 may regularly demonstrate a species with "great" signal. However this relies upon how uncertainties were resolved. In view of these insights and learning of explanatory and examining issues, the client can classify a species as “Strong,” “Weak,” or “Bad”.

4.7 SOFTWARES USED

In this study, the multivariate statistical analysis of the groundwater quality data was carried out with the help of different software such as SPSS 20.0, STATISTICA 10.0.

4.7.1 SPSS® 20.0

SPSS version 20 was used for carrying out the multivariate statistical analysis techniques like Cluster Analysis and Principal Component Analysis employed in this study and also for Basic statistical analysis. SPSS - initially, Statistical Package for the Social Sciences is among the most broadly utilized projects for statistical investigation as a part of sociology.

Statistics incorporated into the base software are:

- Descriptive statistics: Cross arrangement, Frequencies, Descriptives, Explore, Descriptive Ratio Statistics
- Prediction for numerical results: Linear regression
- Prediction for recognizing bunches: Cluster examination (2 step, K-means, hierarchical), Discriminant analysis, Factor analysis.

4.7.2 STATISTICA® 10.0

STATISTICA® is a statistics and analytics programming bundle created by StatSoft. The product incorporates a variety of data investigation, data administration, data perception and information mining methods; and in addition an assortment of predictive modeling, grouping, characterization, and exploratory strategies. Extra procedures are accessible through integration with the free, open source R programming environment. STATISTICA incorporates analytic and exploratory diagrams in addition to standard 2- and 3-dimensional charts. Brushing activities (intuitive naming, stamping, and information exclusion) take into account examination of outliers and exploratory data investigation. Various bundles of logical methods are accessible in six product offerings: (1) Desktop, (2) Data Mining, (3) Enterprise, (4) Web-Based, (5) Connectivity and Data Integration Solutions, and (6) Power Solutions.

4.7.3 Unmix 6.0

EPA's Unmix Model is a numerical receptor model created by EPA researchers that gives logical help to the improvement and audit of the air and water quality standards. Unmix can examine an extensive variety of environmental sample data: sediments, wet deposition, surface water, ambient air, and indoor air. EPA's Unmix display diminishes the vast number of factors in complex investigative data collections to mixture of species called source types and source contributions. The source types are distinguished by contrasting them with measured profiles. Source contributions are utilized to decide how much each source added to a specimen.

4.7.4 Positive Matrix Factorization (PMF) 5.0

Positive Matrix Factorization (PMF) Model developed by the EPA researchers, is a numerical receptor model that gives logical help to the improvement and survey of air and water quality measures, exposure research and environmental forensics. EPA's PMF display diminishes the vast number of factors in complex investigative data collections to mixture of species called source types and source contributions. The source types are distinguished by contrasting them with measured profiles. Source contributions are utilized to decide how much each source added to a specimen.

The assessment of groundwater quality has been carried out in this study with this detailed understanding of all these methods. The results obtained with respect to each of the objectives are discussed in the section to follow.

CHAPTER 5

GROUNDWATER QUALITY DATA ANALYSIS USING MULTIVARIATE STATISTICAL TECHNIQUES

5.1 INTRODUCTION

Statistical investigation offers more attractive options in environmental science, though the results may deviate more from real situations. This chapter deals with the statistical trend analysis of groundwater quality of the Peenya industrial area of Bengaluru city, viz., Mean, Maximum, Minimum and Standard Deviation of the parameters and Multivariate analysis of groundwater quality data. The techniques employed were;

- Cluster analysis for extraction of information with respect to the similarities or dissimilarities among the monitoring sites.
- Discriminant analysis for identification of groundwater quality parameters causing spatial and temporal changes in groundwater quality.
- Principal component analysis to identify the underlying factors describing the data and the effect of likely sources on the quality of groundwater in the study region.

The results of the analysis are presented under the following headings

- Basic Statistical Analysis.
- Multivariate Analysis of groundwater quality data
    1. Cluster analysis,
    2. Discriminant analysis
    3. Principal component analysis.

Table 5.1: Results of statistical analysis on groundwater quality

Sl No	Parameter	Min	Max	Mean	Standard Deviation	IS 10500:2012 Allowable Limit	IS 10500:2012 Permissible Limit
1	**pH**	5.9	7.5	6.75	0.23	6.5 to 8.5	6.5 to 8.5
2	**Turbidity (NTU)**	0	345	**9.79**	35	1	5
3	**TDS (mg/L)**	208	4734	1404	603	500	2000
4	**Sulphate (mg/L)**	21	1236	202	170	200	400
5	**Chloride (mg/L)**	25	1710	340	179	250	1000
6	**Nitrate (mg/L)**	0.62	213	40	30	45	45
7	**Total Hardness (mg/L)**	158	1878	**718**	232	200	600
8	**Calcium (mg/L)**	34	392	165	51	75	200
9	**Magnesium (mg/L)**	5.62	236	72	28	30	100
10	**Fluoride (mg/L)**	0.03	1.7	0.31	0.18	1	1.5
11	**Alkalinity(mg/L)**	25	1292	322	108	200	600
12	**Ammonia (mg/L)**	0.001	10	0.36	0.61	0.5	0.5
13	**Sulphide (mg/L)**	0	0.025	0.02	0.009	0.05	0.05
14	**Copper (mg/L)**	0.01	2	0.03	0.11	0.05	1.5
14	**Zinc (mg/L)**	0.007	4.7	0.25	0.44	5	15
16	**Iron (mg/L)**	0.001	7.3	**0.46**	0.73	0.3	0.3
17	**Manganese (mg/L)**	0.005	26	**0.31**	1.27	0.1	0.3
18	**Lead (mg/L)**	0.005	0.393	0.01	0.02	0.01	0.01
19	**Cadmium (mg/L)**	0.001	0.04	0.001	0.003	0.003	0.003
20	**Chromium (mg/L)**	0.001	78.2	**5.21**	10.95	0.05	0.05

5.2 BASIC STATISTICAL ANALYSIS

A statistical summary of the groundwater quality properties along with the Bureau of Indian standards (BIS) desirable limits is presented in Table 5.1. Results of the present investigation like, Minimum, Maximum, Mean, Median, and Standard Deviation (SD) are presented in the Table 5.1. The importance of groundwater quality parameters from human health point of view is summarized below.

- pH is an estimate of the power of corrosiveness or alkalinity and measures the concentration of hydrogen particles in water. It has no direct unfriendly impact on the wellbeing, be that as it may, a low pH underneath 4.0 will create sour taste and higher pH over 8.5 imparts alkaline taste to the water. pH value of the dataset varied from 5.9 to 7.6.
- Measurement of turbidity reflects the transparency in water. Turbidity doesn't have any serious effect on health impacts, however can meddle with disinfection and act as a medium to microbial development. In our study area, the turbidity of the groundwater dataset varied between 0 – 345.
- Sulphates of groundwater dataset lie in the range from 21 to 1236 mg/L. The existence of sulphates in drinking water can influence taste and direct physiological impact of sulphates on people is cartharsis (purgation of ailementary waterway).
- Chlorides are important in detecting the contamination of groundwater by wastewater. Chloride in abundance grants a salty taste to water and individuals who are not acclimated to high chlorides can be subjected to laxative effects. The chloride content of the groundwater dataset lies in the range 25 to 1710 mg/L.
- Nitrate in the groundwater dataset lies in the range from 0.62 to 213 mg/L. Nitrate signifies the contamination in ground water because of sewage penetration underneath the surface. Two risks associated to human health are identified with

utilization of water containing high nitrate content. These are methaemoglobinaemia in babies and generation of cancer-causing nitrosomines.

- Hardness in water is due to basically by cations like calcium and magnesium and presence of anions such as carbonate, bicarbonate, chloride and sulfate in water. Total hardness of groundwater dataset was found to be in the range from 158 to 1878 mg/L. Hard water does not cause any unfavorable impact on human health. Be that as it may, moderately milder water improves consumer adequacy.
- The concentration of Calcium and Magnesium of the groundwater dataset varied from 34 to 392 mg/L and 5.62 to 236 mg/L respectively. Although calcium and magnesium have been called as one of the important electrolytes in the body, its higher concentrations can cause diarrhea and it has a high solubility and is geologically abundant.
- Fluoride, the most usual transpiring type of fluorine, is one of the normal contaminant of water. Groundwater for the most part contains fluoride broken down by geographical development. Fluoride in groundwater data collected lies in the range from 0.03 to 1.7 mg/L. Introduction to higher measures of fluoride can cause dental fluorosis. In its mildest frame this outcomes in discoloration of teeth, while serious dental fluorosis causes setting and adjustment of tooth enamel.
- Alkalinity is a estimate of the ability of water to counterbalance acids. Basic mixes in the water, for example, bicarbonates, carbonates, and hydroxides expel H+ particles and lower the acridity of the water (which implies expanded pH). The alkalinity of the groundwater data lies in the range of 25 to 1292. Consuming natural alkaline water is by and large viewed as protected, since it contains minerals in in their natural form.
- Ammonia might be available in groundwater because of the debasement of normally happening organic matter or anthropogenic sources like industrial processes, sewage infiltration.

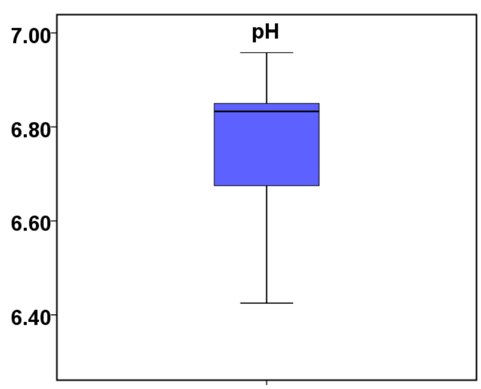
- Ammonia in the groundwater dataset collected lies in the range from 0.01 to 10 mg/L. Although ammonia is displeasing to the respiratory tract, the predetermined number of studies that have been directed demonstrate no long haul sick impacts.
- Sulphide in the groundwater data lies in the range from 0 to 0.025 mg/L. Existence of sulphide in drinking water results in disagreeable taste and odour.
- Copper finds its way in to the groundwater environment from metal plating, industrial activities, mining, and mineral leaching. Copper in the groundwater dataset collected lies in the range from 0.03 to 1.7 mg/L. Copper in drinking water results in stomach and intestinal pain, liver and kidney harm, paleness in high measurements.
- The presence of zinc in groundwater is generally due to industrial waste, metal plating, and plumbing. Also zinc is a major component of sludge. Zinc in the groundwater dataset collected lies in the range from 0.007 to 4.7 mg/L. Zinc has no major impact on human health with the exception of in high measurements gives an undesirable taste to water.
- Iron is available in huge sums in soils and rocks, primarily as insoluble in nature. Iron content of groundwater data ranges from 0.001 to 7.3 mg/L. Despite the fact that there is regularly no unsafe impact on people using waters with noteworthy measures of iron, the issues are fundamentally aesthetic, resulting in colour or turbidity
- Manganese in the groundwater dataset ranges from 0.005 to 26 mg/L. Its presence affects the taste of water and causes aesthetic and economic damage.
- Lead enters the system from industries, mining activities and so on. Influences red platelet science; defers typical physical and mental advancement in infants and young kids. The range of lead in the groundwater dataset lies between 0.005 to 0.393 mg/L.

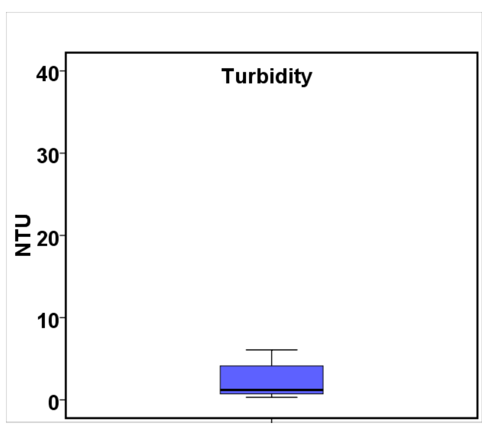
- Cadmium in the water is the consequence of release from metal plating, water funnels, batteries, paints and shades, plastic stabilizers, and leachate occurring from landfill. The range of cadmium in the groundwater dataset is from 0.001 to 0.04.
- The Total Chromium content of groundwater dataset lies in the range from 0.001 to 78.2 mg/L. Chromium is utilized as a part of metal plating and as a cooling-tower water added substance. At higher levels chromium causes liver and kidney harm, respiratory harm, dermatitis, and ulcers on the skin.

The basic statistical analysis revealed that five groundwater quality parameters (turbidity, total hardness, iron, manganese chromium) considered for the study were exceeding permissible limit ,especially chromium and manganese. The heavy metal concentration indicated pollution from anthropogenic source, especially chromium which is known to be human carcinogen.

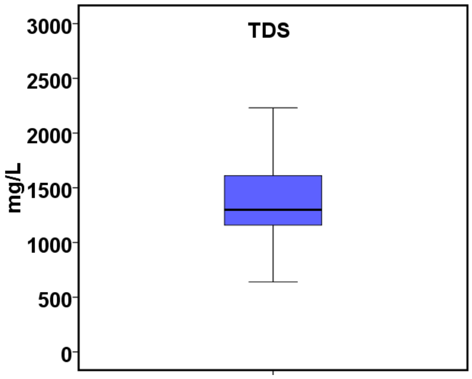
5.2.1 Box Plots

Box plots were constructed for all the 20 parameters to visualize the variation in concentration of groundwater quality concentration and to examining key statistical properties of a parameters (Fig. 5.1).

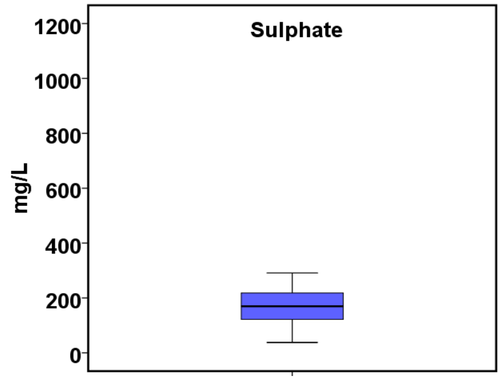




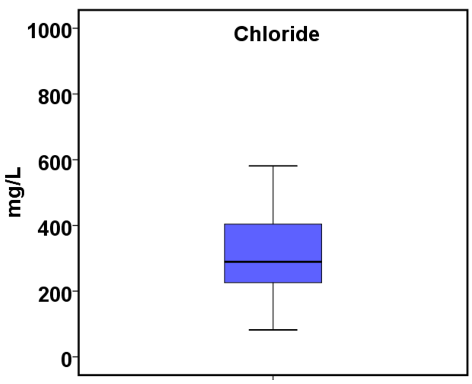
5.1(a) **5.1(b)**



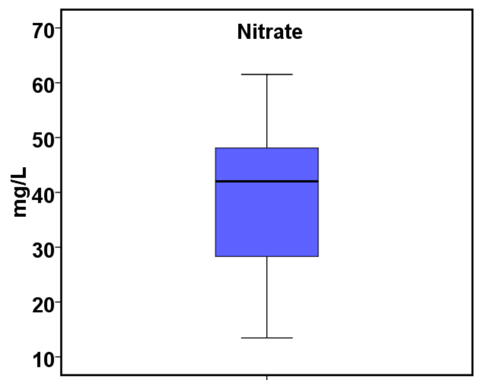
5.1(c)



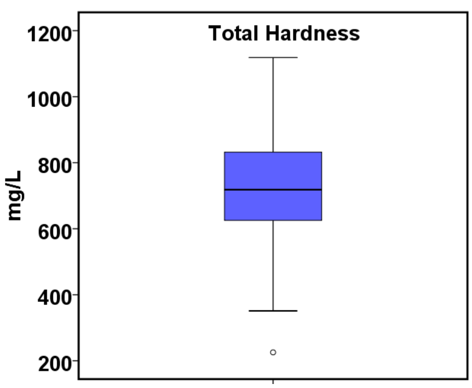
5.1(d)



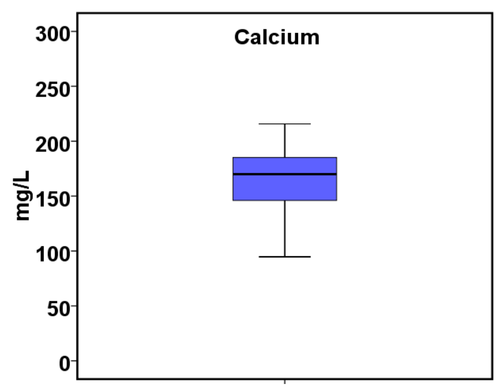
5.1(e)



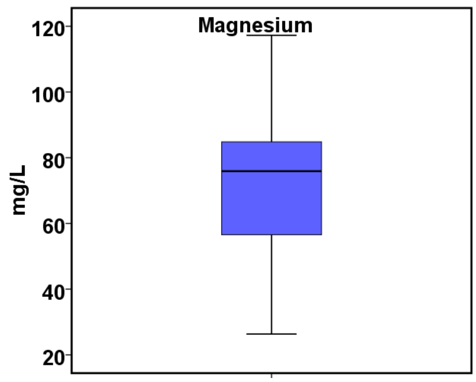
5.1(f)



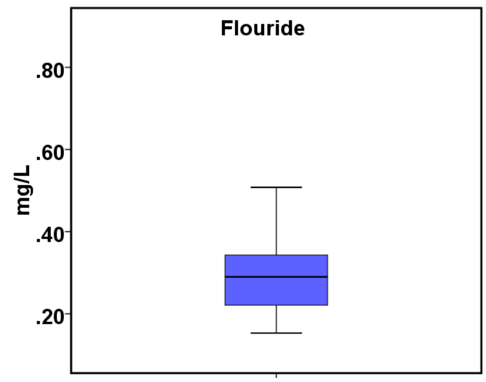
5.1(g)



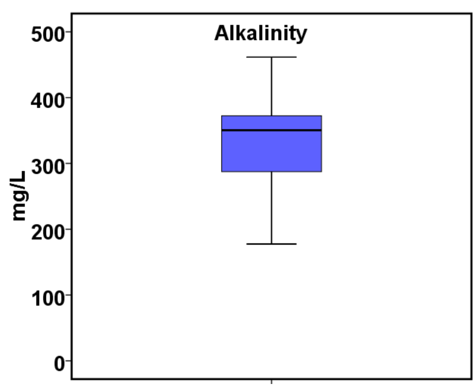
5.1(h)



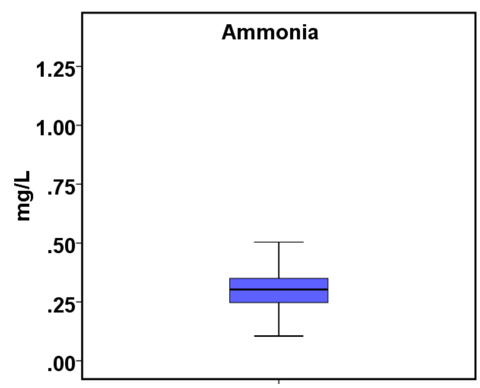
5.1(i)



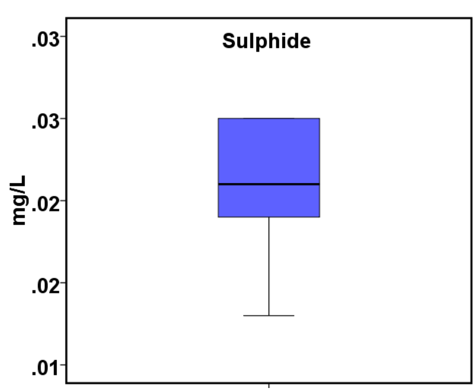
5.1(j)



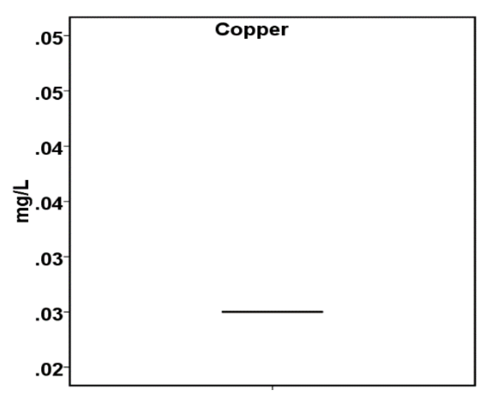
5.1(k)



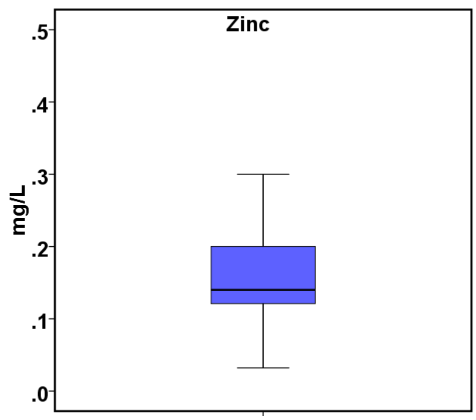
5.1(l)



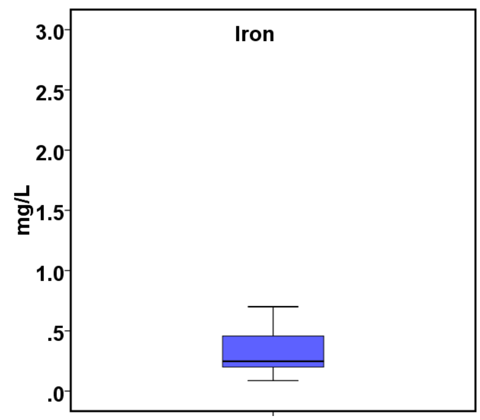
5.1(m)



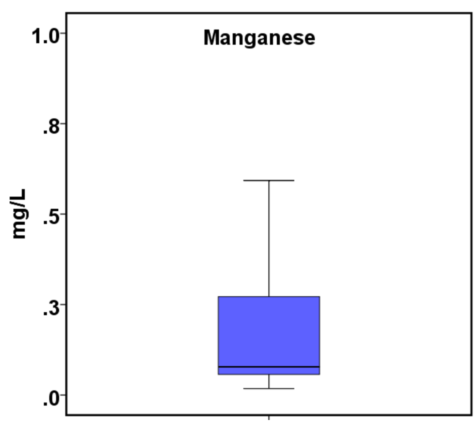
5.1(n)



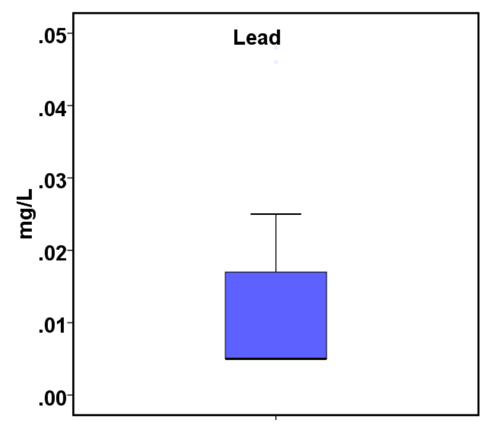
5.1(o)



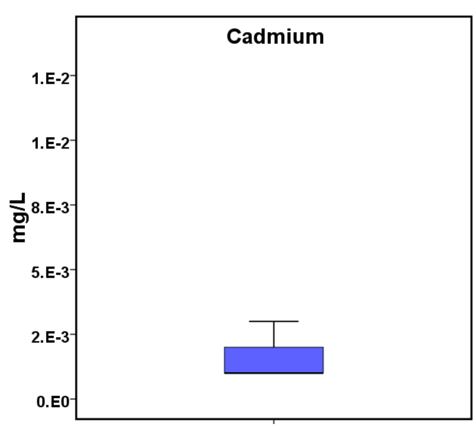
5.1(p)



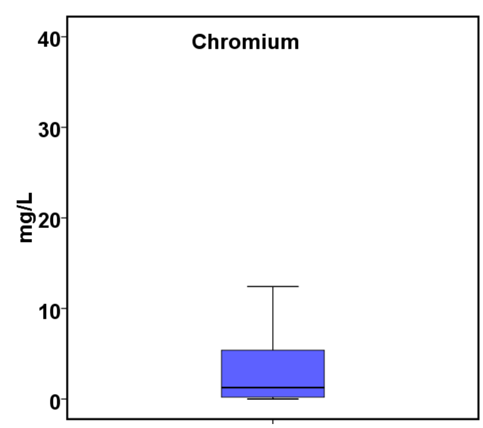
5.1(q)



5.1(r)



5.1(s)



5.1(t)

Fig. 5.1(a) – 5.1 (t). Box plots of Groundwater quality parameters

From the box plots it was observed that in the original data parameters sulphate, chloride, total harness, calcium, fluoride, ammonia and alkalinity were seen to be normally distributed as the box was divided in to two equal halves by the median. However, the distribution of other parameters was far from normal, which may be due to outliers causing the data to be skewed. Therefore the box plots were useful in assessing the symmetry of the data.

5.2.2 Correlation Matrix of Ground Water Quality Parameters

Correlation is the shared connection between two factors. Direct correlation exists when increment or lessening in the estimation of one parameter is related with a comparing increment or decline in the estimation of the other. Favorable position of the technique is that we can influence expectations about things, when we have an idea about correlations. The relationship is positive when increment in one parameter causes the expansion in the other parameter and it is negative when increment in one parameter causes the diminishing in other parameter.

The value of correlation coefficient (r) ranges between +1 and -1. For strong correlation between the parameters the range of “r” should be +0.8 to 1.0 and -0.8 to -1.0.Correlation is said to be moderate if the value of “r” is in the range of +0.5 to 0.8 and -0.5 to -0.8 and weak when range of “r” is between +0.0 to 0.5 and-0.0 to -0.5 (Liu et al., 2003).

A correlation matrix for the groundwater quality dataset is shown in Table 5.2.

Table 5.2: Correlation matrix of ground water quality parameters

	pH	Tur	TDS	SO_4	Cl	NO_3	TH	Ca	Mg	F	HCO_3	NH_3	S^{--}	Cu	Zn	Fe	Mn	Pb	Cd	Cr
pH	1.00																			
Tur	0.14	1.00																		
TDS	-0.16	-0.29	1.00																	
SO_4	-0.05	**0.82**	0.09	1.00																
Cl	-0.13	-0.20	**0.81**	0.03	1.00															
NO_3	0.10	-0.20	0.57	0.02	0.43	1.00														
TH	-0.17	0.20	**0.79**	0.55	0.74	0.53	1.00													
Ca	-0.09	0.44	0.43	**0.68**	0.55	0.34	**0.81**	1.00												
Mg	-0.27	-0.12	**0.78**	0.20	**0.81**	0.48	**0.86**	0.55	1.00											
F	-0.07	0.20	-0.11	0.01	0.01	-0.24	-0.13	-0.12	-0.01	1.00										
HCO_3	-0.02	-0.48	0.51	-0.46	0.55	0.29	0.29	0.12	0.42	0.01	1.00									
NH_3	-0.29	-0.05	-0.18	-0.15	-0.19	-0.24	-0.29	-0.33	-0.17	0.35	-0.17	1.00								
S^{--}	0.28	0.12	-0.02	0.10	0.07	-0.01	0.09	0.13	0.03	-0.39	-0.06	-0.05	1.00							
Cu	0.13	0.05	0.39	0.18	-0.06	0.26	0.21	0.03	0.04	-0.16	-0.05	-0.04	-0.06	1.00						
Zn	0.02	-0.05	0.23	-0.02	0.09	0.19	0.18	0.08	0.12	-0.25	0.06	-0.12	0.21	0.39	1.00					
Fe	-0.23	0.24	-0.24	0.04	-0.17	-0.35	-0.22	-0.24	-0.17	**0.59**	-0.28	**0.66**	-0.15	-0.08	0.01	1.00				
Mn	0.00	0.18	0.35	0.27	0.12	0.07	0.26	0.16	0.10	-0.09	-0.07	0.31	0.07	**0.64**	0.19	0.23	1.00			
Pb	-0.27	-0.10	0.13	-0.01	0.17	0.13	0.16	0.08	0.26	-0.05	0.05	0.19	-0.16	-0.01	-0.06	0.19	0.02	1.00		
Cd	-0.21	0.05	0.09	0.11	0.15	0.03	0.24	0.20	0.30	-0.06	0.12	-0.08	-0.10	-0.04	-0.15	-0.10	0.00	0.18	1.00	
Cr	**-0.64**	-0.12	0.13	0.04	0.08	0.16	0.19	0.14	0.26	-0.16	0.13	0.21	-0.10	-0.01	0.06	0.15	-0.01	0.48	0.18	1.00

- The strong positive correlation between turbidity and sulphate (0.82), total dissolved solids and chloride (0.81), calcium and hardness (0.81), magnesium and chloride (0.81), hardness and magnesium (0.86), were found which are responsible for water mineralization as they share a common origin source and their tendency to follow a similar trend (e.g., due to concentration by water-rock interaction and ion exchange).
- Moderate correlation was observed between hardness and total dissolved solids (0.79), calcium and sulphates (0.68), magnesium and total dissolved solids (0.78), iron and fluoride (0.59), iron and nitrate (0.66), manganese and copper (0.64), chromium and pH (0.64).
- The correlation of Calcium, magnesium, alkalinity, copper, iron and manganese was found to be weak with the other parameters.

After basic statistical analysis of the groundwater quality data, to further process it and to report overall trends from it, multivariate statistical techniques were applied for the interpretation and presentation.

5.3 MULTIVARIATE STATISTICAL ANALYSIS

Multivariate statistical techniques are helpful for examination and translation of water quality datasets and in water quality evaluation, finding the contamination sources and understanding temporal/spatial varieties for able water quality administration.

5.3.1 Cluster Analysis

Cluster analysis was performed using SPSS to find out the spatial similarity between sampling sites and to group them into meaningful clusters. The objective was to extricate information regarding the similarities or dissimilarities among the sampling sites.

5.3.1.1 Dendrogram

The cluster analysis of data generated a dendrogram, grouping the 41 sampling sites into three statistically significant clusters from the study area. Fig. 5.2 shows the dendrogram from the hierarchical cluster analysis for the groundwater quality dataset.

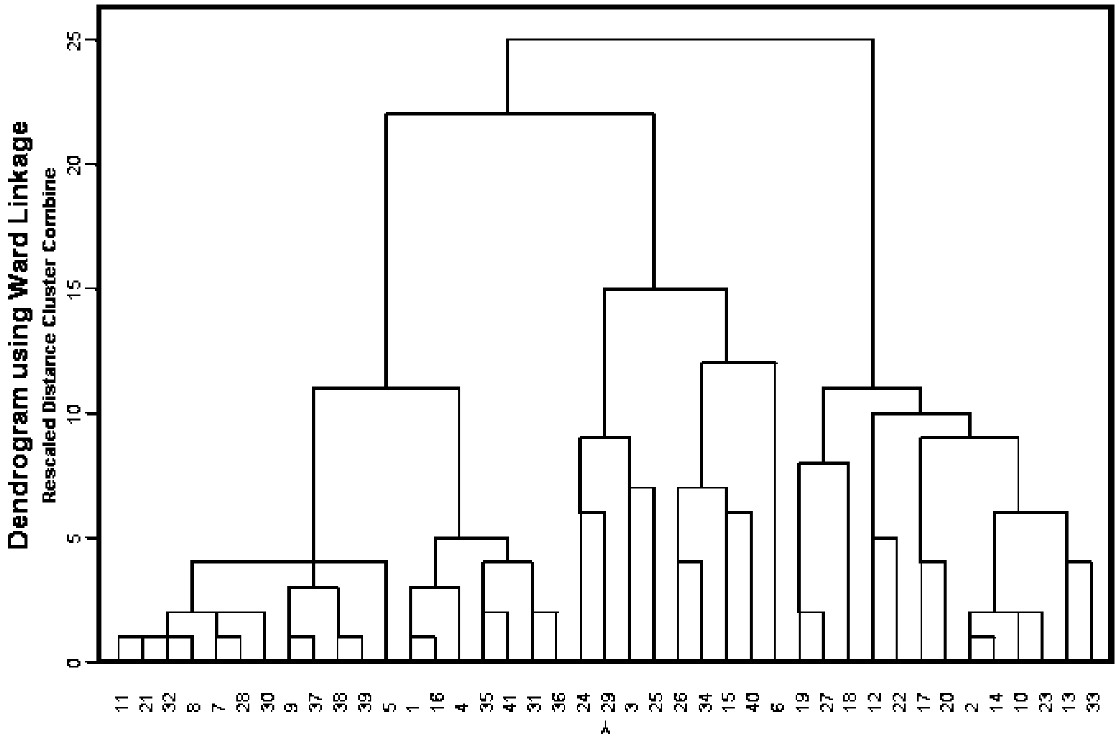


Fig 5.2 Dendrogram showing spatial clustering of monitoring sites

- Spatial cluster analysis produced a dendrogram as shown in Fig. 5.2. It can be seen from the Fig. 5.2 that there are three early clusters at (Dlink/Dmax) X 100 < 25. Cluster 1 includes 19 stations (47 %), Cluster 2 includes 09 stations (22 %) and the Cluster 3 includes the remaining 13 stations (32 %) for the groundwater quality dataset.
- The three clusters produced by the dendrogram were used to find the average concentration of different parameters in each of the cluster as shown in Table 5.3.

Table 5.3: Average concentration for each Cluster

Parameter	Cluster 1	Cluster 2	Cluster 3
pH	6.83	6.70	6.69
Turbidity (NTU)	3.11	**35.33**	1.81
TDS (mg/L)	1182.85	1166.78	1910.09
Sulphate (mg/L)	144.40	322.61	207.46
Chloride (mg/L)	275.25	229.34	519.56
Nitrate (mg/L)	39.27	30.80	**47.93**
Total Hardness (mg/L)	**632.05**	**656.91**	**898.07**
Calcium (mg/L)	153.19	150.05	195.46
Magnesium (mg/L)	63.75	60.28	94.92
Fluoride (mg/L)	0.28	0.36	0.30
Total Alkalinity(mg/L)	337.74	208.47	384.39
Ammonia (mg/L)	0.29	**0.57**	0.33
Sulphide (mg/L)	0.02	0.02	0.02
Copper (mg/L)	0.03	0.07	0.03
Zinc (mg/L)	0.25	0.25	0.36
Iron (mg/L)	0.22	**1.07**	**0.40**
Manganese (mg/L)	0.09	**0.83**	**0.28**
Lead (mg/L)	0.01	0.01	**0.02**
Cadmium (mg/L)	0.00	0.00	0.00
Chromium (mg/L)	**2.09**	**5.15**	**10.14**
Number of sites	**19**	**09**	**13**

- From Table 5.3 it can be noted that, 13 borewells (Number 19, 27 18 12 22 17, 20, 02, 14, 10, 23, 13, 33) belonging to cluster 3 have high concentration values and exceeds the permissible limit for 6 parameters (nitrate, total hardness, iron, manganese, lead and chromium). It can be also noted that the 09 borewells (Number 06, 40, 15, 34, 26, 25, 03, 29, 24) belonging to cluster 2 have 6 parameters (turbidity, total hardness, ammonia, iron, manganese and chromium) exceeding the permissible limit. Whereas parameters (total hardness & chromium) in cluster 1 are exceeding the permissible limits from 19 borewells (Number 36, 31, 41, 35, 04, 16, 01, 05, 39, 38, 37, 09, 30, 28, 07, 08,, 32, 21, 11).
- The high level of industrial and residential activity in the Peenya region is bound to have a significant impact on the environment. Thus the spatial distribution of cluster analysis results indicates the poor quality of groundwater in the region and can be attributed significantly to anthropogenic activities as it can be observed that the industries are spread out throughout the 4 phases (Phase1, 2, 3 and 4) of the study area.
- It can be observed that the clustering analysis procedure is valuable in offering sound characterization of groundwater quality in the entire area and will facilitate a future spatial sampling methodology in an ideal way which can decrease the quantity of monitoring stations and related expense.

There are other studies, where similar approach has successfully been applied in water and groundwater quality programs. Vega et al., (1998) used cluster analysis and accomplished a significant grouping of stream water samples in light of seasonal and spatial criteria. Kowalkowski et al., (2006) decided common clusters and groups of observing locations with alike contamination character and recognizing essential discriminant in the dataset using cluster analysis. Andrade et al., (2008) minimized the sample collection and analysis, with regard to space and time and minimal loss of data.

Singh et al., (2005) used the river water quality to group the river stretch in to relatively low pollution, very high pollution and moderate pollution regions.

Various studies in their cluster analysis (Adar et al,. 1992; Schot and van der Wal, 1992; Gu¨ler et al., 2002) also found that using the Euclidean distance as a distance measure and Ward's method as a linkage rule produced the most distinctive group.

Further to distinguish the groundwater quality variables which are responsible for spatial and temporal variations in groundwater quality, discriminant analysis was performed on the groundwater quality data.

5.3.2 Discriminant Analysis

Discriminant analysis was used to recognize the most significant parameters influencing the spatial and temporal variations in groundwater quality.

- In the present study, discriminant analysis (DA) was carried out on raw data using three different modes, standard, forward stepwise and backward stepwise to form discriminant functions (DFs). The objective was to evaluate both temporal and spatial variations in groundwater quality. Temporal discriminant analysis was carried out taking the monitoring period (winter, summer and monsoon) as the grouping variable and the 20 measured groundwater quality parameters as the independent variables. Summer season was considered from March to May, monsoon from June to November and winter from December to February (IMD).
- Discriminant functions (DF) and classification matrix (CM) for temporal variation relative to standard mode, forward stepwise mode and backward stepwise mode are shown in Table 5.4 and 5.5.
- The standard discriminant analysis mode constructed the discriminant functions using 20 parameters are shown in Table 5.4. Both standard mode and forward stepwise mode constructed DFs using 20 and 14 discriminant parameters respectively.

Table 5.4: **Classification Functions for discriminant analysis of temporal variation**

Variable	Standard mode			Forward stepwise mode			Backward stepwise mode		
	Winter coefficient	Summer coefficient	Monsoon coefficient	Winter coefficient	Summer coefficient	Monsoon coefficient	Winter coefficient	Summer coefficient	Monsoon coefficient
pH	117.36	116.6	114.36	96.47	94.70	90.69			
Turbidity	2.99	−3.43	−2.06						
TDS	0.63	0.52	0.47	0.41	0.38	0.31	0.32	0.29	0.19
SO_4	−0.50	−0.32	0.46	−0.61	−0.38	0.40			
Cl	−3.31	−3.26	−3.15	−2.83	−2.44	−2.463			
NO_3	−0.64	0.44	1.22	2.38	3.74	3.58	4.02	5.36	7.24
TH	0.04	0.05	0.07	−0.02	0.05	0.03	−0.12	−0.17	−0.11
Ca	0.22	0.48	0.19	−0.94	−0.53	−0.41	-0.74	-0.42	-0..51
Mg	−9.15	−9.4	−9.76	−8.78	−8.44	−8.27	-6.52	-6.21	-6.05
F	0.28	−0.01	−0.10						
HCO_3	0.97	1.91	0.83	0.98	1.13	0.91	0.71	0.97	0.82
NH_3	0.06	0.08	−0.01						
S^{--}	10.05	9.29	9.05						
Cu	−42.91	−28.11	−18.45	−30.36	−21.27	−14.62			
Zn	3.91	2.82	2.01	0.90	−0.76	−0.33			
Fe	−1.09	−1.44	−1.19	−0.400	−0.96	−0.63			
Mn	1.10	1.01	0.92	1.21	1.13	0.86			
Pb	25.93	17.66	69.34						
Cd	1264.81	1300.58	1365.88	1124.56	1214.66	1302.34			
Cr	3.67	3.30	2.28	1.36	1.12	0.60			
Constant	**−777.26**	**−714.44**	**−738.65**	**−625.40**	**−658.53**	**−688.76**	**−62.43**	**−69.51**	**−72.58**

Table 5.5: Classification Matrix for Discriminant analysis of temporal variation

Monitoring Season	% Correct	Winter	Summer	Monsoon
Standard DA mode				
Winter	94.12	84	4	3
Summer	95.59	3	85	4
Monsoon	95.89	1	3	85
Total	95.05	87	92	92
Forward stepwise DA mode				
Winter	93.32	85	3	1
Summer	94.35	3	86	3
Monsoon	94.42	1	2	86
Total	94.03	89	91	90
Backward stepwise DA mode				
Winter	92.18	92	1	3
Summer	91.06	2	93	2
Monsoon	91.00	1	2	93
Total	94.42	95	96	98

- The corresponding CMs assigned 95% cases correctly (Table 5.4 and 5.5). But, in backward stepwise mode DA awarded CMs with 94% correct assignations by making use of only 5 discriminant parameters (Table 5.4 and 5.5) with a little different match for each season compared with the forward stepwise mode. Forward stepwise DA revealed that T-Hard, NO_3, Ca, Mg, HCO^3 and TDS were accompanied by another group of parameters i,e pH, SO_4, Cl, Cu, Zn, Fe, Mn, Cd and Cr. Also, a quite little significant group of 4 parameters which were remaining was observed from the standard DA mode assignations.
- In forward stepwise mode, variables are included step-by-step beginning with the more significant until no significant changes are obtained, whereas, in backward stepwise mode, variables are removed step-by-step beginning with the less significant until no significant changes are obtained.

- Hence, the results of temporal DA suggested that T-Hard, NO_3, Ca, Mg, HCO^3 and TDS are the most critical parameters to segregate between three different seasons thereby accounting for the vast majority of the anticipated temporal variations in the groundwater quality.
- Spatial discriminant analysis was performed in the same way as temporal discriminant analysis, by taking the spatial clusters obtained in cluster analysis (Clusters 1,2 & 3) as the grouping variable. The 20 measured groundwater quality parameters were considered as the independent variables. The discriminant functions and classification matrices obtained from the standard; forward stepwise and backward stepwise modes of DA are shown in Tables 5.6 and 5.7.
- Standard discriminant analysis mode constructed discriminant functions including all the 20 parameters assigning 97% of the cases correctly. This can be attributed to the fact that the initial clustering was also done taking all the parameters into consideration in cluster analysis.
- Both the standard and forward stepwise mode constructed DFs by making use of 20 and 14 discriminant parameters respectively. The corresponding CMs assignations was than 94% correct (Table 5.6 and 5.7).The backward stepwise mode DA provided CMs with quite less than 92% correct assignations by making use of 6 discriminant parameters (Table 5.6 and 5.7). From the Backward stepwise discriminant analysis it was found that Fe, Cr, Cl, Mn, Cu and Cd are the discriminating parameters spatially.
- The correct assignations (94%) by DA for three different clusters obtained from cluster analysis (Cluster 1, 2 and 3) also confirms the appropriateness of discriminant analysis. Both CA and DA projected relevant variations in groundwater quality, resulting from the impact of the Peenya industrial area and also due to the effect of seasonal variation in groundwater quality. DA also showed that there are noteworthy differences between these three clusters Cluster 1, 2 and 3, that were demonstrated by 6 discriminating parameters.

Table 5.6: **Classification Functions for discriminant analysis of Spatial variation**

Variable	Standard mode			Forward stepwise mode			Backward stepwise mode		
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 1**	**Cluster 2**	**Cluster 3**
pH	97.24	96.16	94.23	86.15	84.78	80.29			
Turbidity	2.22	−2.45	−1.96						
TDS	0.51	0.46	0.38	0.33	0.25	0.19			
SO_4	−0.40	−0.29	0.26	−0.53	−0.34	0.30			
Cl	−3.21	−3.19	−3.09	−2.75	−2.62	−2.43	-2.22	-2.01	-1.92
NO_3	−0.54	0.39	0.92	1.98	2.52	2.76			
TH	0.03	0.04	0.06	−0.01	0.03	0.02			
Ca	0.19	0.38	0.14	−0.74	−0.57	−0.39			
Mg	−8.23	−8.38	−8.68	−7.36	−7.51	−7.18			
F	0.28	−0.01	−0.10						
HCO_3	0.85	1.05	0.78	0.86	1.20	0.96			
NH_3	0.06	0.08	−0.01						
S^{--}	10.05	9.29	9.05						
Cu	−39.14	−28.24	−17.34	−28.28	−19.16	−16.44	-21.23	-15.34	-12.24
Zn	2.82	1.93	1.41	0.70	−0.62	−0.21			
Fe	−0.92	−1.22	−1.02	−0.24	−0.85	−0.42	-0.19	-0.98	-0.35
Mn	0.98	0.79	0.64	1.01	1.23	0.74	1.12	1.32	0.92
Pb	25.93	17.66	69.34						
Cd	1198.81	1200.45	1256.23	1094.24	1164.32	1178.22	998.62	1034.26	1078.42
Cr	2.96	2.45	2.14	0.96	1.02	0.52	0.78	0.94	0.46
Constant	**−698.34**	**−701.78**	**−718.34**	**−595.12**	**−618.98**	**−654.49**	**−54.89**	**−59.23**	**−52.22**

Table 5.7: Classification Matrix for Discriminant analysis of Spatial variation

Monitoring Season	% Correct	Cluster 1	Cluster 2	Cluster 3
Standard DA mode				
Cluster 1	98.42	94	0	0
Cluster 2	97.35	2	97	1
Cluster 3	96.64	2	1	98
Total	97.54	98	98	99
Forward stepwise DA mode				
Cluster 1	95.22	95	2	0
Cluster 2	93.74	0	96	3
Cluster 3	93.68	1	0	96
Total	94.44	96	98	98
Backward stepwise DA mode				
Cluster 1	92.34	95	1	0
Cluster 2	91.66	0	96	2
Cluster 3	91.20	1	1	96
Total	91.12	96	97	98

Further the groundwater quality dataset was subjected to principal component analysis for pattern recognition and to explain the variance of a large set of inter-correlated parameters

5.3.3 Principal Component Analysis

Principal Component Analysis was carried out on the standardized data in order to differentiate between the configuration patterns of the analyzed groundwater samples and to locate the factors that influence each one.

- Principal component analysis of the entire data set evolved 7 principal components with eigen values explaining around 74% of the overall variance in the groundwater quality data set.

- To test the credibility of principal component analysis, Kaiser-Meyer-Olkin (KMO) and Barlett's tests were carried out. The Kaiser Rule was used as the sole cut-off criterion for estimating the number of factors. The Kaiser rule drops all components with eigenvalues under 1.
- For a good principal component analysis/factor analysis, KMO values close to 1 are required. KMO > 0.5 indicates the model is correct (Monteiro and Pinheiro 2004). The Closer the KMO values are to 1, demonstrates a sizeable examining sufficiency (0.8 and higher are incredible, 0.7 is satisfactory, 0.6 is average, under 0.5 is unsatisfactory). Sensibly huge values are required for a decent principal component analysis. Smaller KMO signifies that a principal component analysis of the factors may not be a smart thought.
- The results obtained from the KMO and Bartlett's sphericity test were 0.712 (Table 5.8) with Approx. Chi-square of 3895.420 respectively, in the current study, implying that factor analysis (FA) would be effective in reducing dimensionality. Hence it was concluded that the sampling size is sufficient and the correlation between parameters can be accepted.

Table 5.8 KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy	**0.712**
Approx. Chi-Square	3895.420
Bartlett's Test of Sphericity	190
Sig.	**.000**

Bartlett's test of sphericity indicates whether a correlation matrix is an identity matrix, which would indicate that variables are unrelated. Here the probability related with Bartlett's Test of Sphericity should be smaller than the level of significance, the probability associated with the Bartlett test (Sig = 0) is <0.001, which meets this requirement indicating that there is significant relationships among the variables.

5.3.3.1 Scree plot

The observation of the scree plot from the Fig. 5.3, gives a visual of overall variance linked with each of the factors. The scree plot was utilized to recognize the number of principal components to be kept with a specific end goal to understand the fundamental information structure.

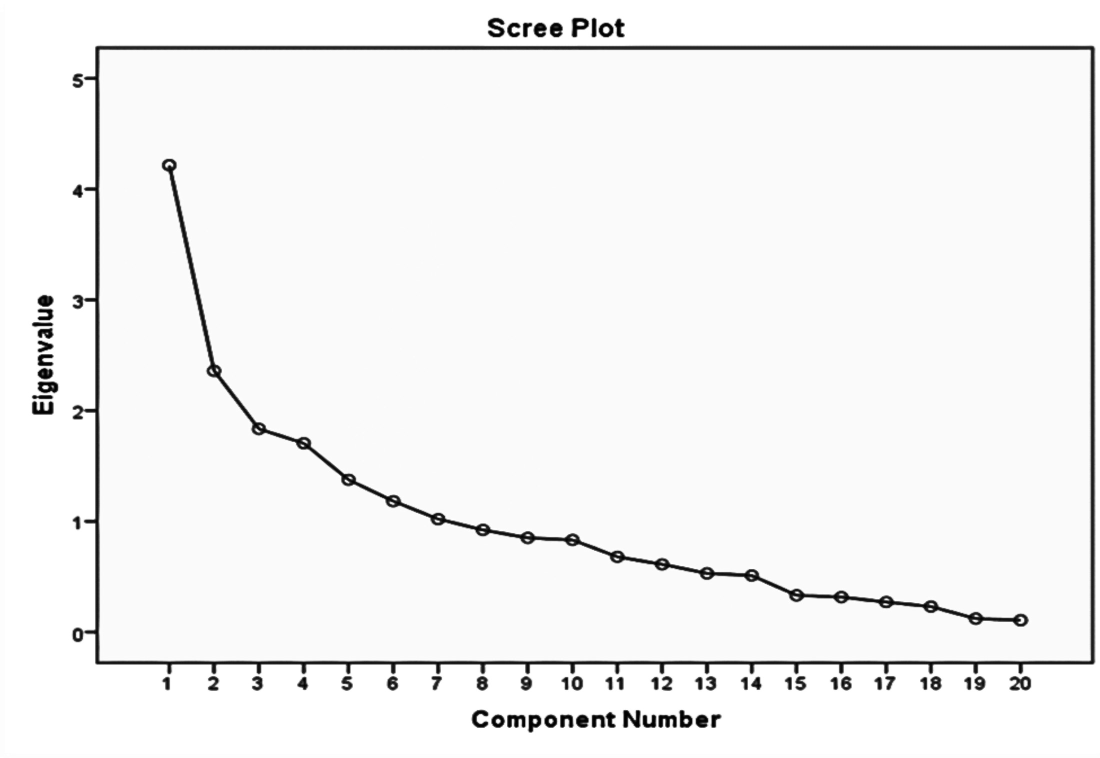


Fig 5.3 Scree plot

For the current study, the scree plot demonstrated an articulated difference in the slope after the fourth eigen value, but we have retained seven eigen values greater than unity and explaining 73.42% of the variance. The fact that a principal component has as eigen score higher than 1 implies it contains more data than the primal factor, so the lessening of dimensionality is guaranteed.

5.3.3.2 Eigen values, Percent of Variance

Table 5.9 presents the eigen values, the percentage of variance and the cumulative percentage of variance linked with one another. It can be observed that the initial seven factors explain 73.42% of total variance.

Table 5.9 Eigen values, percent of variance, cumulative eigen value, cumulative percent of variance for the Principal Component analysis

Factor	Eigenvalue	Cumulative Eigen Value	% of variance	Cumulative percent of variance
1	4.21	4.21	21.07	21.07
2	2.35	6.57	12.79	33.86
3	1.83	8.40	10.17	44.03
4	1.70	10.11	9.52	53.55
5	1.37	11.48	7.87	61.42
6	1.18	12.67	6.90	68.32
7	1.02	13.69	5.10	73.42
8	0.92	14.61	4.61	78.03
9	0.85	15.46	4.25	82.28
10	0.83	16.29	3.05	85.33
11	0.67	16.97	2.89	88.22
12	0.61	17.58	2.23	90.45
13	0.53	18.11	1.95	92.40
14	0.51	18.62	1.65	94.05
15	0.33	18.95	1.32	95.37
16	0.31	19.27	1.16	96.53
17	0.27	19.54	1.05	97.58
18	0.23	19.77	0.94	98.52
19	0.12	19.89	0.86	99.38
20	0.10	20	0.62	100.00

5.3.3.3 Component Matrix

Table 5.10 shows the loading associated with each variable with respect to each of the seven factors. The first factor of principal component analysis often represents the most important process with highest eigen value and has the highest variance among the factors.

Table 5.10: Varimax rotated factor loading

Parameter	Varimax Rotated Component						
	VF 1	VF2	VF3	VF4	VF5	VF6	VF7
pH	-0.09	0.00	0.05	**0.77**	-0.03	0.18	0.00
Turbidity	**0.86**	-0.12	0.00	-0.11	0.16	-0.01	-0.00
TDS	**0.82**	-0.10	0.16	0.10	-0.09	0.04	-0.04
SO_4	0.17	**0.92**	0.10	0.04	-0.03	-0.00	0.00
Cl	-0.07	**0.87**	-0.00	0.04	0.00	0.03	0.05
NO_3	-0.005	**0.64**	0.05	-0.08	-0.06	0.40	-0.28
TH	**0.85**	0.35	0.06	0.09	-0.10	0.06	0.07
Ca	**0.64**	0.26	-0.02	0.05	-0.15	0.08	0.01
Mg	**0.84**	0.05	0.05	0.12	0.01	0.00	0.14
F	0.06	0.07	-0.13	-0.22	**0.70**	-0.31	-0.04
HCO_3	**0.60**	-0.50	-0.01	-0.06	0.00	-0.05	-0.04
NH_3	-0.11	-0.10	0.03	0.24	**0.59**	0.29	-0.00
S^{--}	-0.01	0.07	-0.05	-0.10	0.02	-0.01	0.10
Cu	0.02	0.04	**0.91**	-0.03	-0.06	0.00	-0.02
Zn	0.07	-0.03	0.35	0.04	-0.06	0.38	-0.01
Fe	-0.10	0.09	0.08	0.09	0.08	**0.84**	0.05
Mn	0.10	0.07	**0.88**	-0.03	0.11	0.02	0.02
Pb	0.04	0.00	0.02	0.45	0.04	-0.03	**0.86**
Cd	0.12	-0.00	-0.01	-0.06	-0.01	0.06	0.36
Cr	0.11	-0.01	0.00	**0.79**	0.03	0.09	-0.03
Eigen Value	**4.21**	**2.35**	**1.83**	**1.70**	**1.37**	**1.18**	**1.02**
Percentage Variance	**21.07**	**12.79**	**10.17**	**9.52**	**7.87**	**6.90**	**5.10**
Cumulative Percentage Variance	**21.07**	**33.86**	**44.03**	**53.55**	**61.42**	**68.32**	**73.42**

Factor loadings are represented by the terms ‘strong’, ‘moderate’, and ‘weak’, which denotes the absolute loading values of >0.75, 0.75–0.5 and 0.5–0.3, respectively (Liu et al.,2003).

5.3.3.4 Varimax rotation

The values of PCA can be enhanced by using varimax rotation of eigen values. By doing so varifactor’s are acquired in which the participation of the original variables is more clearer. The rotation of principal components can accomplish a less complex and more important portrayal of basic factors by diminishing the contribution to PCs of factors with minor importance and expanding the more critical ones. Rotation gives another arrangement of components, every one including essentially a subset of the original factors with as meager overlap as could be allowed, so that the first factors are partitioned into groups to some degree independent of each other. 73.42 % of the total variance in the data set was explained by the seven factors, while the percentage of variability ranged between 5.10 and 21.07 %.

- The first varifactor accounting for 21.07% of the total variance, had strong positive loadings with, Magnesium, Total dissolved solids, Turbidity and Total hardness and moderate loading with Calcium and Alkalinity. Turbidity in groundwater is mostly inorganic and caused by natural geological factors. The weathering of crystalline basement rock hosting the groundwater aquifer may be responsible for the high concentration of these ionic species. It can be seen that varifactor 1 contains typical hydro chemical variables originating from mineralization of geological components of the soil. Thus it represents the mineral group and the source can be attributed as “natural or bedrock”, as reported by earlier researchers also (Drever, 1997; Kumar et al., 2006, Subba Rao et al., 2006). Also 70% of the mechanism controlling the chemical composition of groundwater of Bangalore city is controlled by rock-water interaction (DMG 2011).

- Varifactor 2 explains 12.79% of the total variance and has strong positive loadings with and Chloride, Nitrate and Sulphate. Nitrate in groundwater of the study area can be attributed to contamination from unlined drains and sewage effluent as there is no agricultural activity nor application of nitrogenous fertilizers as it is an urban area. The moderate loadings for sulphates and chloride can be attributed to seepage from sewers, septic tanks and industrial effluents. Thus the factor can be attributed as “sewage”.
- Varifactor 3 had strong positive loadings with Copper and Manganese and accounted for explains 10.17% of the total variance. Copper is used in alloys, as a catalyst, in anti-fouling paints. Thus, it basically represents a toxic metals group which is due to the industrial activity in the area. Hence this factor can be related as “paint shipping”.
- Varifactor 4 contributed to 9.52% of the total variance and showed strong positive loadings with pH and Chromium and indicated that it may have come from unique source. Presently around 60 electroplating industries located in Peenya Industrial area. Many of these industries were engaged in hard chrome/chrome plating and zinc plating, few units are engaged in copper, tin and nickel plating. The contamination of ground water due to illegal discharge of electroplating effluent into drains or due to seepage of effluent from underground storage tanks of these industries has contributed to chromium content in the ground water (CPCB 2014). Therefore this factor can be attributed as “chromium electroplating”.
- Varifactor 5 with 7.87% of the total variance exhibited moderate loading with Fluoride and Ammonia. High fluoride content is present for most part in gnessic and granitic territories. Granitic gneisses are the most seasoned arrangements in the state and have experienced greatest weathering. The joints, breaks, faults and vertical openings in the formations are possessed by fluoride-bearing minerals. The leachable fluoride in these minerals is reflected in the upper aquifer system

(DMG 2011). The geology of the study area is predominated by granites/gneisses with intensive presence of pegmatites which contributes to the occurrence of fluoride in bore wells. Hence the source can be attributed as "geologic".

- Varifactor 6 clarified 6.90% of the overall variance and showed strong positive loadings with iron. The presence of iron represents metal pollution derived from industrial effluents most probably from large steel processing plants, which are present in the study area. Further groundwater samples located near to the steel processing plants, also showed high iron concentrations, further confirming this speculation. So the source can be called as "steel Processing industry". Chucking away of remnant iron in open areas because of industrial activity is one of the factor causing higher values of iron in groundwater in the region (Basappa Reddy 2003).
- The last varifactor (VF7) with strong positive loadings with cadmium and moderate loading with lead explains 5.1% of the total variance. Manufacturing units generating lead battery units which dispose off acid directly into the environment, contaminates the groundwater with lead. A study conducted by Ramesh, A (2014), in order to determine the remediation of heavy metal contaminated soil and groundwater also found the presence of lead in groundwater as well as soil sample in the region. This similarity points out that the factor can be termed as "lead acid Battery manufacturing unit"

The results of principal component analysis reveal that there are both natural and anthropogenic sources in groundwater with the anthropogenic sources being the major ones. This shows that anthropogenic activities have a significant impact on groundwater quality and form a major source of groundwater pollution. Thus, the sources affecting water quality in the study region may be broadly classified as dissolution of minerals from rock water interactions in the aquifer, effect of anthropogenic activities, and ion exchange processes in water.

5.4 Conclusions

- From the basic statistical analysis it was observed that five groundwater quality parameters (turbidity, total hardness, iron, manganese chromium) considered for the study were exceeding permissible limit ,especially chromium whose average concentration was 5.21 mg/L.
- A correlation matrix of variables was calculated to distinguish several relevant hydro chemical relationships. A strong positive correlation between turbidity and sulphate (0.82), total dissolved solids and chloride (0.81), calcium and hardness (0.81), magnesium and chloride (0.81), hardness and magnesium (0.86), were found which are responsible for water mineralization.
- Cluster Analysis was useful in classifying the 41 sampling sites into three main clusters as high pollution and low pollution areas. This helps in the identification of problematic zones in the area where remedial actions need to be focused.
- DA was useful in identifying a few indicator parameters responsible for large variations (spatial and temporal) in groundwater quality in the study area. T-Hard, NO_3, Ca, Mg, HCO^3 and TDS were the most significant parameters to discriminate between three different seasons and accounted for 94% assignation of seasonal cases, thereby causing the temporal variations in the groundwater quality.
- Fe, Cr, Cl, Mn, Cu and Cd as the most important parameters discriminating between the 3 clusters and accounting for 92% spatial assignation of cases. Therefore, discriminant analysis caused lessening in the dimensionality of the vast dataset, portraying a few marker parameters in charge of expansive variations in the groundwater quality.

- Principal component analysis was useful in recognizing the seven factors/sources explaining 73.42 % of the total variance. Varifactors obtained from principal component analysis showed that the groundwater quality variations are primarily explained by mineralization, sewage and industrial activity in the area especially the electroplating industries which are responsible for high heavy metal concentration in the groundwater content.

These results were further used to calculate the percentage source contributions using receptor oriented source apportionment techniques which are discussed in the next chapter.

CHAPTER 6

SOURCE APPORTIONMENT OF GROUNDWATER QUALITY

6.1 INTRODUCTION

Post qualitative determination of number and characteristics of possible sources by principal component analysis. The source contributions were calculated using different receptor oriented source apportionment techniques;

- Absolute Principle Component Scores-Multiple Linear Regression (APCS-MLR)
- Unmix
- Positive Matrix Factorization (PMF).

6.2 ABSOLUTE PRINCIPLE COMPONENT SCORES-MULTIPLE LINEAR REGRESSION (APCS-MLR)

Source apportionment of the parameters by receptor modeling through multiple linear regression on absolute principal component scores (APCS-MLR) was carried out. The Principal Components (PCs) determined using PCA were considered as the sources in APCS-MLR.

6.2.1 Source Contribution to groundwater quality

Once the number of sources and their characteristics were determined by principal component analysis, the source contributions were then computed using multiple linear regression on absolute principal component scores (APCS–MLR). APCS-MLR is a proven and effective technique for getting quantitative information regarding the contributions of each source type. The absolute factor score was calculated for each sample and for each identified factor. The contributing concentration for each sample was estimated by a multiple regression analysis using the absolute factor scores as predictor variables, as shown in Table 6.1.

Table 6.1: APCS-MLR Modelled Source contribution to groundwater quality

Parameter	Source type								Estimated mean (E)	Measured mean (M)	Ratio (E/M)	R^2
	Source 1	Source 2	Source 3	Source 4	Source 5	Source 6	Source 7	UIS				
pH	-	-	0.35	5.46	-	-	1.29	-	7.10	6.75	1.05	0.85
Turbidity	5.92	-	-	-	0.96	-	-	-	6.89	9.79	0.70	0.81
TDS	1385	-	86.00	173.25	-	-	-	88.21	1632.45	1404.15	1.23	0.75
SO_4	-	203.74	25.47	-	-	5.09	-	7.64	237.68	202.43	1.26	0.90
Cl	7.45	253.30	-	11.92	-	-	20.86	4.47	298.00	340.28	0.88	0.78
NO_3		28.00	2.50	-	-	1.50	-	-	32.00	40.31	0.79	0.59
TH	501.50	-	-	-	40.00	-	-	49.00	590.00	718.76	0.82	0.89
Ca	187.50	37.50	-	12.50	-	-	-	12.50	179	165.20	1.51	0.77
Mg	99.45	-	-	11.70	-	-	-	5.85	76	72.54	1.61	0.75
F	0.04	-	-	-	0.20	-	-	-	0.24	0.30	0.78	0.77
HCO_3	444.60	-	-	-	-	29.64	19.76	-	384.00	322.86	1.53	0.82
NH_3	-	-	-	-	0.18	0.02	-	0.01	0.29	0.36	0.57	0.83
S^{--}	-	0.01	-	-	0.008	-	0.008	-	0.026	0.02	0.72	0.74
Cu	-	-	0.02	-	-	0.00	-	-	0.02	0.03	0.57	0.74
Zn	-	-	0.06	-	-	0.05	-	0.04	0.19	0.25	0.58	0.69
Fe	-	-	-	-	-	0.64	-	0.11	0.55	0.46	1.63	0.68
Mn	-	-	0.21	-	-	-	0.01	-	0.22	0.31	0.69	0.61
Pb	-	-	-	0.01	-	-	0.01	-	0.013	0.01	1.61	0.62
Cd	-	-	-	-	-	0.0008	0.002	-	0.0028	0.002	1.64	0.78
Cr	-	-	-	3.483	-	-	-	0.39	3.87	5.21	0.74	0.64

After calculating the estimated to measured ratios of the parameters, the accuracy of the APCS–MLR can be tested. APCS-MLR provides information on the contribution from each source to the concentration of each element in the same sample. The (E/M) ratio of the parameters varied between 0.57(Nitrate) to 1.64(Cadmium).

From the correlation coefficients it was observed that there was reasonable adequacy between the measured and estimated values from multiple regression. The ratio of mean estimated and measured values of most of the groundwater quality parameters indicated the compatibility of receptor modeling approach to the source apportionment of groundwater quality. Based on R^2 values the accuracy of the model is very high for pH, turbidity, sulphates, total hardness, alkalinity and nitrates with R^2 between 0.8- 1, high for total dissolved solids, chloride, calcium, magnesium, fluoride, sulphide, copper, zinc, iron, manganese, lead, cadmium and total chromium with R^2 between 0.6-0.8, moderate for nitrate with R^2 between 0.4-0.6 as shown in Table 6.1.

6.2.2 Percentage Source Contribution to groundwater quality

Using APCS-MLR model the percentage contribution from the different sources to each parameter was then calculated and is shown in Table 6.2.

From the table 6.2, it can be inferred that most of the groundwater quality parameters were primarily influenced by natural/bedrock source (accounting for 85.92, 79.94, 85, 75, 85 and 65% of variations in turbidity, TDS, T-Hard, Ca, Mg, HCO_3 concentrations respectively), sewage (80, 85, 87.50 and 40 % of SO_4, Cl, NO_3 and S^{--}, respectively), geologic (85 and 85% of F and NH_3) and various industrial wastewater pollution sources (92, 95, 76.90, 90, 80, 96, 80 and 96% of Cu, Zn, Mn, pH, Cr, Fe, Pb and Cd, respectively). Apart from these, unidentified sources in the study area also attributed to groundwater pollution in the case of most water quality variables (between 3 to 25%).

Table 6.2: APCS-MLR Modelled Percentage Source contribution to groundwater quality

Parameter	Source type								R^2
	Source 1	Source 2	Source 3	Source 4	Source 5	Source 6	Source 7	UIS	
pH	-	-	4.93	76.90	-	-	18.17	-	0.85
Turbidity	85.92	-	-	-	13.93	-	-	-	0.81
TDS	79.94	-	4.96	10	-	-	-	5.09	0.75
SO_4	-	80	10.00	-	-	7	-	3	0.90
Cl	2.50	85	-	4	-	-	7	1.5	0.78
NO_3	-	87.50	7.81	-	-	4.69	-	-	0.59
TH	85	-	-	-	6.78	-	-	8.31	0.89
Ca	75	15	-	5	-	-	-	5	0.77
Mg	85	-	-	10	-	-	-	5	0.75
F	15	-	-	-	85	-	-	-	0.77
HCO_3	65	-	-	-	20	6	4	11	0.82
NH_3	-	-	-	-	85	10	-	5	0.83
S^{--}	-	40	-	-	20	-	10	15	0.74
Cu	-	-	92	-	-	8	-	-	0.74
Zn	-	-	40	-	-	35	-	25	0.69
Fe	-	-	-	-	-	85	-	15	0.68
Mn	-	-	95	-	-	-	5	-	0.61
Pb	-	-	-	20	-	-	80	-	0.62
Cd	-	-	-	-	-	4	96	-	0.78
Cr	-	-	-	90	-	-	-	10	0.64

Source 1= Natural or Bedrock; Source 2= Sewage; Source 3= Industrial; Source 4 = Electroplating; Source 5= geologic; Source 6: Steel Processing industry; Source 7: Lead manufacturing; UIS; Unidentified source.

Significant percentage of contribution to zinc (25%) was observed from unidentified sources which may be due to the heavy zinc galvanizing activity in the area. Also 15% contribution to iron from these sources may be attributed to corrosion of casing pipes of the bore wells as majority of the borewells in the area are older than a decade or two (Basappa Reddy 2003). The 11% contribution to alkalinity can be related to the effluents from the pharmaceutical and drug industries in the area (Pius et al., 2012). 15% of sulphide contribution can be attributed to fecal contamination of the aquifer impacted by septic tank leaching (Roser 2005).

After quantifying the source contributions using APCS-MLR model, the groundwater quality data was subjected to Unmix model to find the source contribution.

6.3 UNMIX MODEL

6.3.1 Source Identification

Unmix utilizes some strict criteria to decide the number of sources. For example it doesn't prescribe to use every one of the factors into the investigation when running the model, as a part of the information may have a huge inherent error corrupting the S/N ratio. The information was screened utilizing the signal-to-noise (Min S/N ratio) criteria higher than 2, assessed by Unmix. The model identified 6 sources using 16 groundwater quality parameters (EC, TDS, Ca, Mg, Na, HCO^3, Cl, SO_4, NO_3 and TH).
The parameters that were discarded by the model according to suggest exclusion were pH, NH_3, turbidity, TDs due to specific variances SV > 0.5. It is suggested that species having variance of more than 50 percent because of error, or specific variance (SV) be excluded further Unmix modelling (Paatero 2007), which is not considered in APCS-MLR model.

"Typical" level of Auto Unmix was chosen and the model was run which gave a six source solution. Using the Unmix model the percentage contribution from the different sources to each parameter was calculated as shown in Table 6.3 and are represented in the pie charts in Fig. 6.2. Unmix produced essentially the same results as that of APCS-MLR with very little differences among them. This is beacause UNMIX assumes that factor compositions are approximately constant and that all observations are not affected by all factors. The fact that both models extracted similar sources is not surprising, as the UNMIX algorithm is based on PCA and both models operate on correlations in the data. Five out of six sources coincided with the sources identified by the APCS-MLR model. These are Natural source, Chromium Electroplating, Sewage, Geologic, Lead acid Battery manufacturing. The only noticeable discrepancy is in the significant contribution of Cu, Zn, Fe and Mn in the third source (phosphating) resolved by Unmix.

- Due to high composition of magnesium, total hardness, calcium and alkalinity which are typical hydrochemical variables originating from oxidation of geological constituents of the soil, Source 1 was inferred as the “natural/bedrock”. The weathering of crystalline basement rock hosting the groundwater aquifer might be responsible for high concentration of these ionic species. Thus it represents the mineral group and the source can be attributed as “natural or bedrock”, as reported by earlier researchers also (Drever, 1997; Kumar et al., 2006, Subba Rao et al., 2006). Also 70% of the mechanism controlling the chemical composition of groundwater of Bangalore city is controlled by rock-water interaction (DMG 2011).
- Higher concentration of chromium and zinc in source 2 indicated that this source is originated from “chromium electroplating”. Presently more than 60 electroplating industries located in Peenya Industrial area and many illegally functioning workshops also. Many of these industries and workshops are engaged in hard chrome/chrome plating and zinc plating. The contamination of ground water due to illegal discharge of electroplating effluent into drains or due to seepage of effluent from underground storage tanks of these industries has contributed to chromium and zinc content in the ground water (CPCB 2014).

Table 6.3: Unmix Modelled Percentage Source contribution to groundwater quality

Parameter	Source type					
	Source 1	**Source 2**	**Source 3**	**Source 4**	**Source 5**	**Source 6**
SO_4	2	3	-	**82**	10	3
Cl	6	< 1	< 1	**92**	-	-
NO_3	3	-	4	**89**	2	2
TH	**92**	-	4	2	< 1	< 1
Ca	**89**	-	-	9	2	-
Mg	**84**	-	4	12	-	-
F	10	1	2	5	**78**	4
HCO_3	**75**	1	1	18	5	-
S^{--}	-	-	-	-	-	-
Cu	-	25	**75**	-	-	-
Zn	-	**40**	**60**	-	-	-
Fe	15	35	**40**	-	-	10
Mn	-	6	**88**	3	2	1
Pb	-	20	-		-	**80**
Cd	-	2	-	6	-	**92**
Cr		**92**	5	1	< 1	< 1

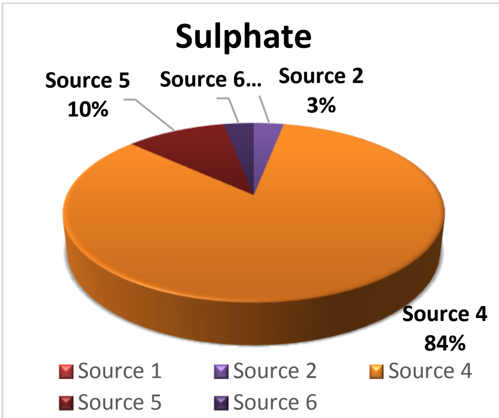
Source 1= Natural; Source 2= Chromium Electroplating; Source 3= Phosphating; Source 4 = Sewage; Source 5= Geologic; Source 6: Lead acid Battery manufacturing;

- Source 3 corresponds to "Phosphating" as copper, zinc, iron and manganese were enriched in Unmix results. Phosphating is practiced in the study area wherein, the metal surface is coated with a layer of insoluble phosphates by treating it with an acidic phosphate containing solution containing zinc, copper, iron and manganese, which are commonly used in the region.
- Sulphates, chlorides and nitrates contributed significantly in Source 4. Nitrate in groundwater of the study area can be attributed to contamination from unlined drains and sewage effluent as there is no agricultural activity nor application of nitrogenous fertilizers as it is an urban area. The moderate loadings for sulphates

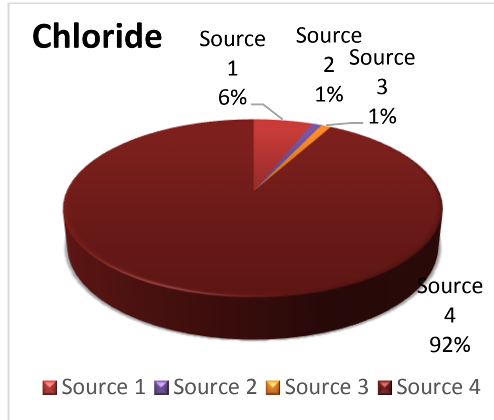
and chloride can be attributed to seepage from sewers, septic tanks and industrial effluents. Thus the factor can be attributed as “sewage”.

- Large compositions with fluoride was characterized with Source 5. High fluoride content is present for the most part in gnessic and granitic territories. Granitic gneisses are the most seasoned arrangements in the state and have experienced greatest weathering. Since the geology of the study area is predominated by granites/gneisses with intensive presence of pegmatites which contributes to the occurrence of fluoride in bore wells (DMG 2011). Hence the source can be attributed as “geologic”.
- An industrial source was identified as Source 6 because of high compositions by the heavy metal elements lead and cadmium. Lead acid battery manufacturing units use raw materials which are alloys of lead calcium, lead antimony, lead tin, pure lead and sulphuric acid in the study region, disposing off acid directly into the environment have contaminated the groundwater with lead and cadmium. A study conducted by Ramesh, A (2014), in order to determine the remediation of heavy metal contaminated soil and groundwater also found the presence of lead in groundwater as well as soil sample in the region. This similarity points out that the factor can be termed as “lead acid Battery manufacturing unit”.

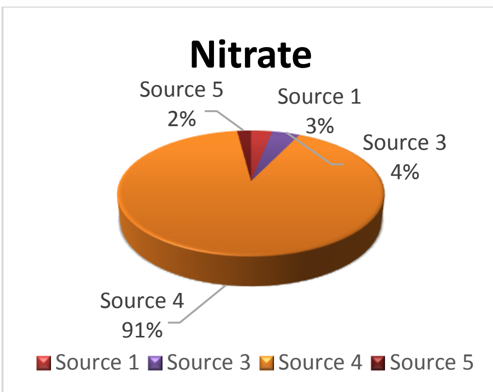
6.3.2 Pie Charts showing Percentage Source Contribution to groundwater quality



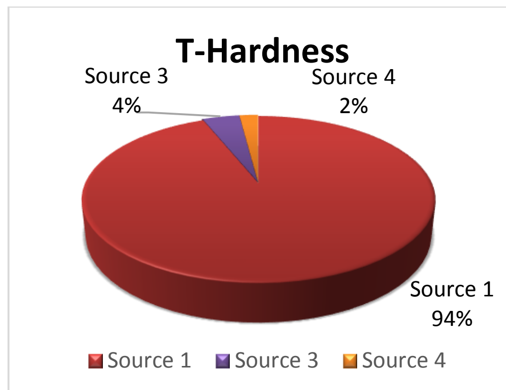
6.1 (a)



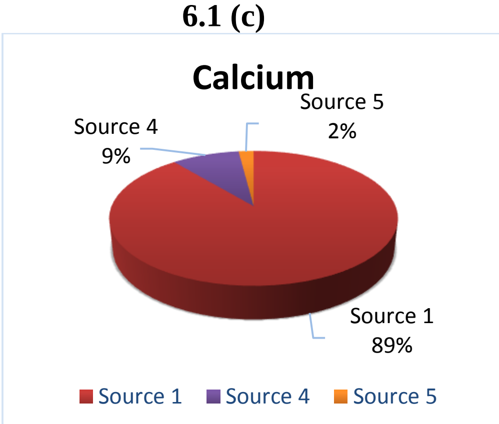
6.1 (b)



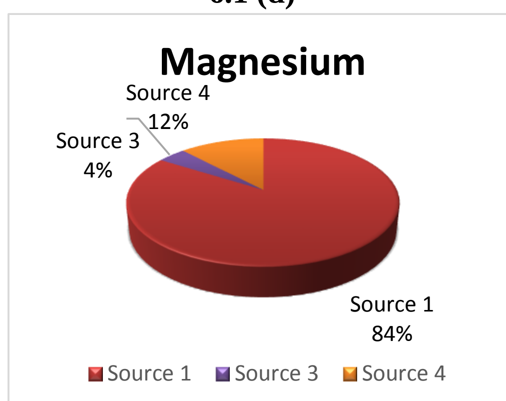
6.1 (c)



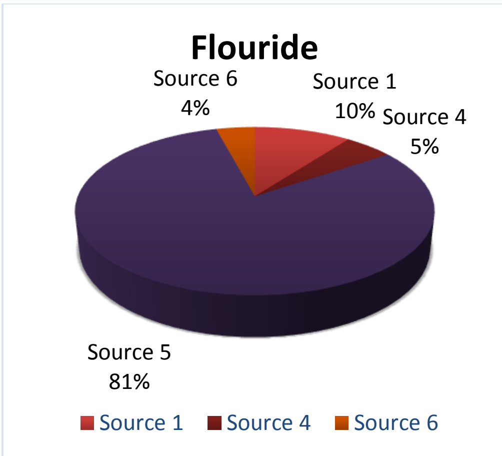
6.1 (d)



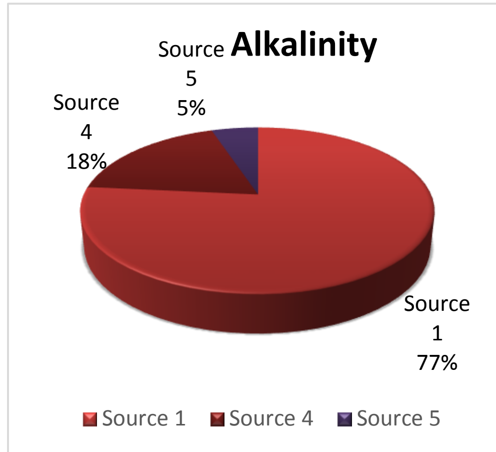
6.1 (e)



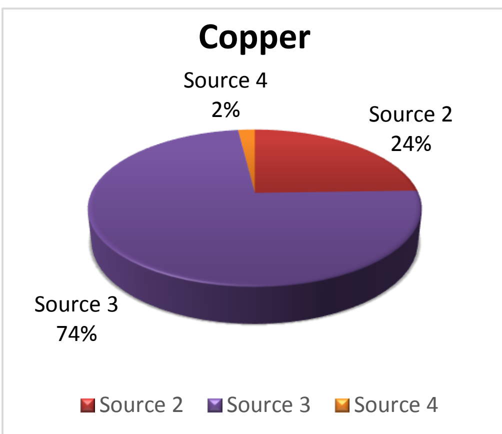
6.1 (f)



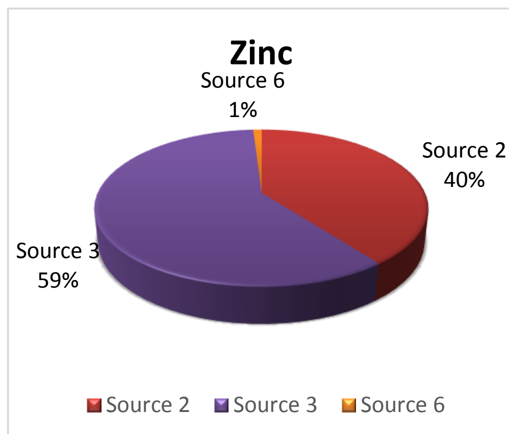
6.1 (g)



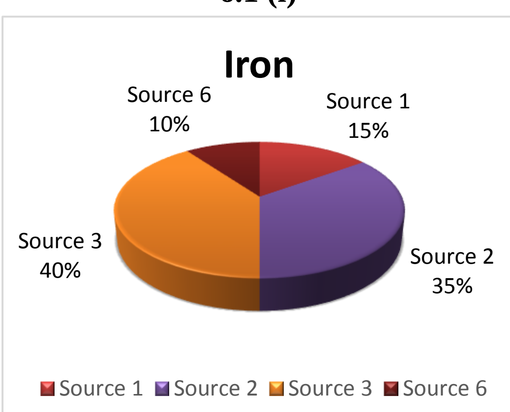
6.1 (h)



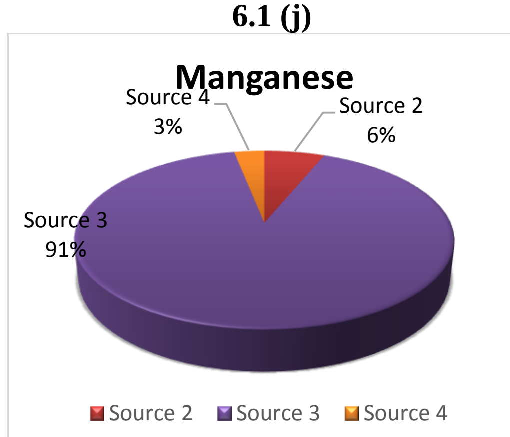
6.1 (i)



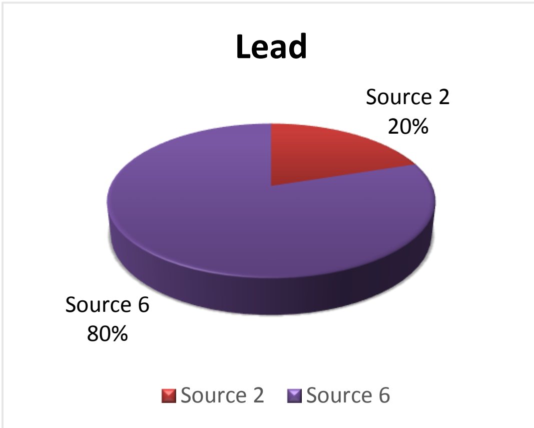
6.1 (j)



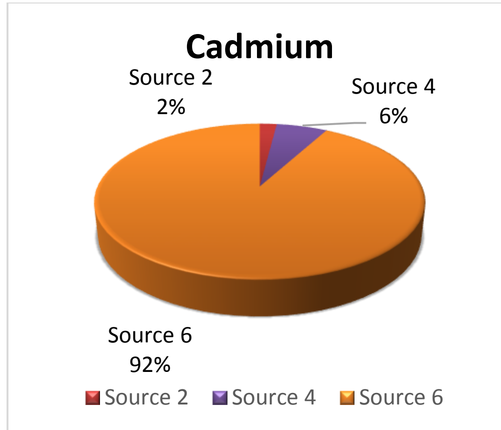
6.1 (k)



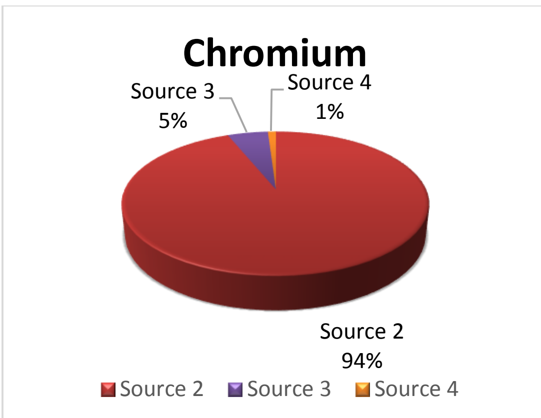
6.1 (l)



6.1 (m)



6.1 (n)



6.1 (o)

Fig. 6.1(a) – 6.2 (o): Unmix modelled Source contributions

The pie charts indicate that majority of the groundwater quality parameters were primarily influenced by natural/bedrock source (accounting for 92, 89, 84 and 75% of variations in T-Hard, Ca, Mg, HCO_3 concentrations respectively), sewage (82, 92 and 89 % of SO_4, Cl, and NO_3, respectively), geologic (78% of F) and various industrial wastewater pollution sources (92, 75, 60, 88, 80 and 92% of Cu, Zn, Mn, Cr, Pb and Cd, respectively). Presence of iron was attributed to two sources i,e chromium electroplating and phosphating with contributions of 35 to 40% respectively.

6.3.3 Model Performance Evaluation

The accuracy of the Unmix can be tested by comparing the modelled values with the measured concentrations for each parameter. The mean difference indicates that majority of the Unmix modelled results were closer to the measurements. Larger uncertainty was observed for Mg (37.33), NO_3 (-35.50), Mn (-35.58), Cu (30) and Fe (30) as shown in Table 6.4.

Table 6.4: Unmix comparison of measured and calculated concentrations

Parameters	Measured Mean (M)	Unmix Modelled (E)	Ratio (E/M)	R^2	% error
SO_4	202.43	170.46	0.84	0.89	15.79
Cl	340.28	368.46	1.08	0.92	-8.28
NO_3	40.31	54.62	1.35	0.78	-35.50
TH	718.76	620.45	0.86	0.62	13.68
Ca	165.20	205.36	1.24	0.70	-24.31
Mg	72.54	45.46	0.62	0.69	37.33
F	0.30	0.24	0.8	0.74	20.00
HCO_3	322.86	350.85	1.09	0.97	-8.67
S^{--}	0.02	0.018	0.90	0.98	10.00
Cu	0.03	0.021	0.70	0.86	30.00
Zn	0.25	0.31	1.24	0.68	-24.00
Fe	0.46	0.32	0.69	0.81	30.43
Mn	0.31	0.42	1.35	0.65	-35.58
Pb	0.01	0.012	0.83	0.74	-20.00
Cd	0.002	0.0017	0.85	0.80	15.00
Cr	5.21	4.25	0.81	0.78	18.43

As observed from the correlation coefficients, the measured and predicted values exhibited good adequacy between them. Additionally, the ratio of mean Unmix modelled and measured values of most of the groundwater quality variables suggest goodness of the receptor modelling approach to the source apportionment of groundwater. Based on the R^2 values, the accuracy of the model is very high for sulphate, total hardness, alkalinity, ammonia and zinc with R^2 between 0.8-1, high for chloride, calcium, magnesium, fluoride, sulphide, copper, iron, manganese, lead, cadmium and total chromium with R^2 between 0.6-0.8, moderate for nitrate with R^2 between 0.4-0.6 as shown in Table 6.6.

6.3.4 Observed vs predicted Line Plots

The Line plots of observed vs modelled values for the parameters are shown below.

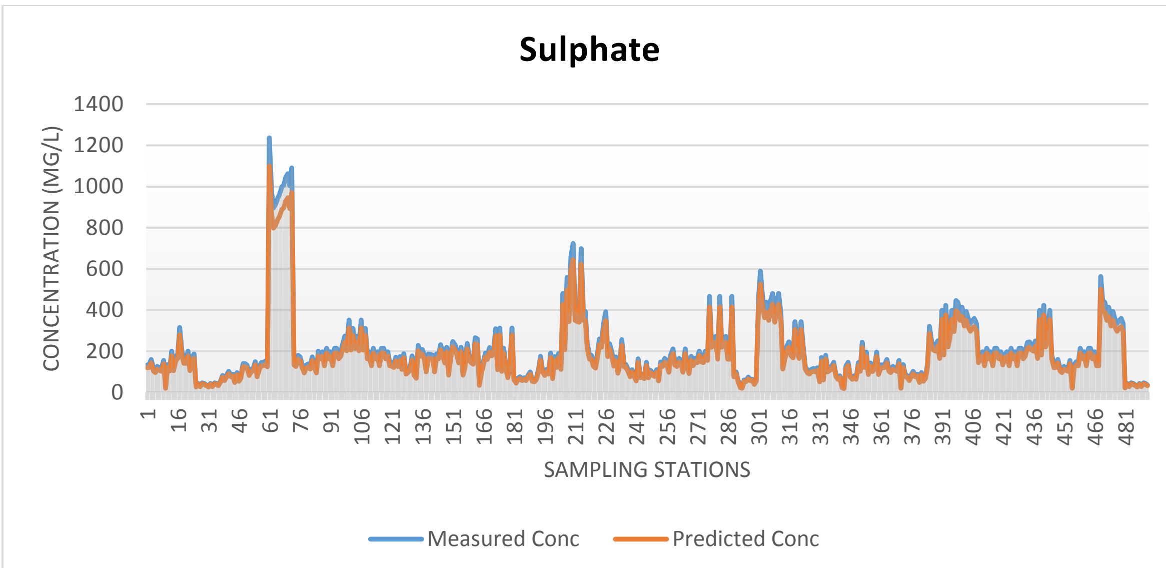


Fig 6.2 (a)

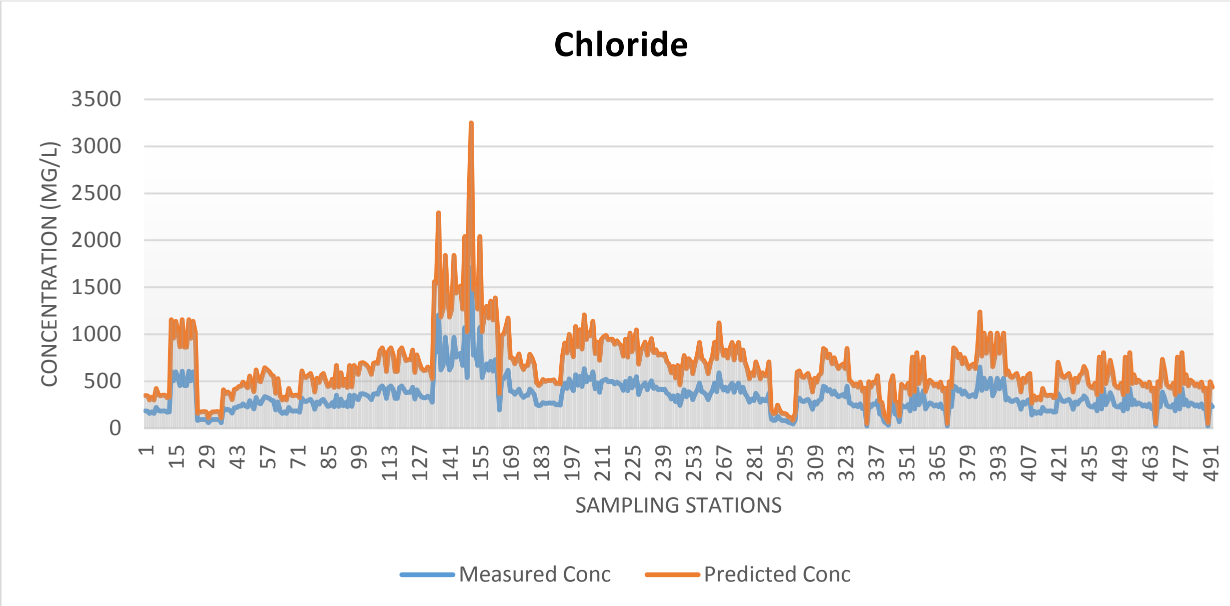


Fig 6.2 (b)

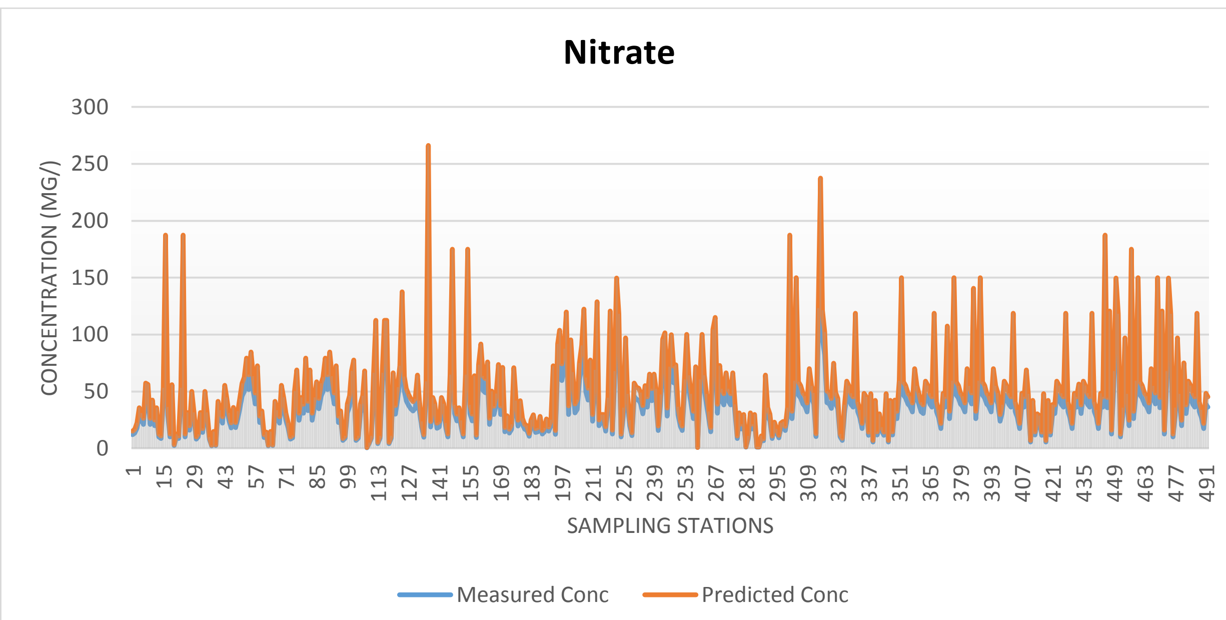


Fig 6.2 (c)

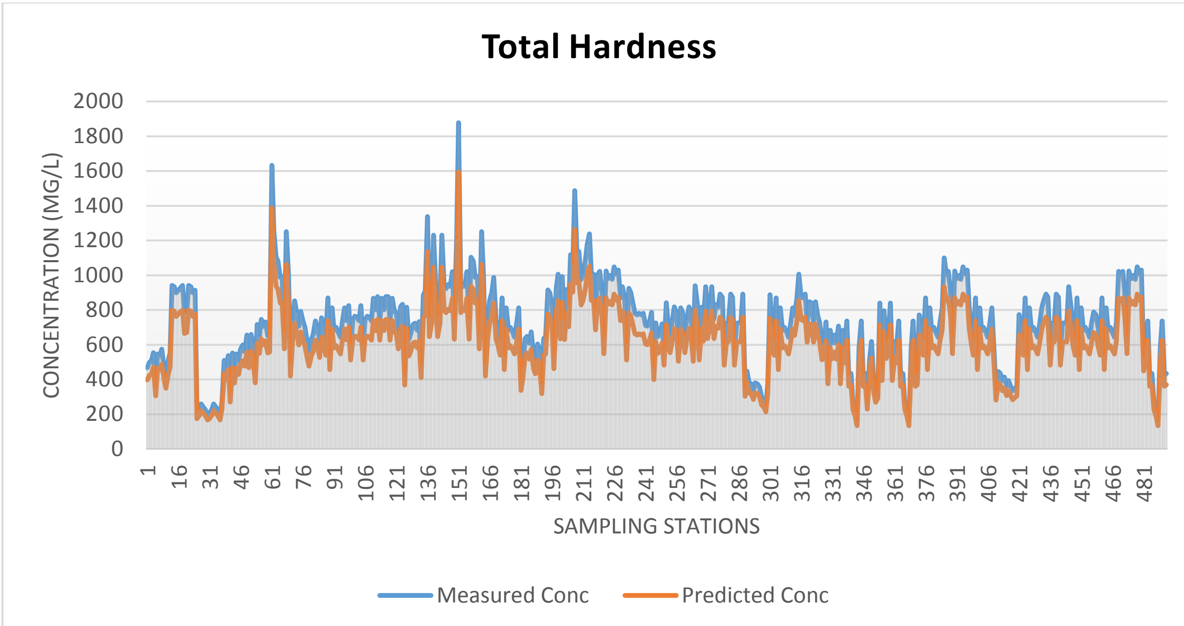


Fig 6.2 (d)

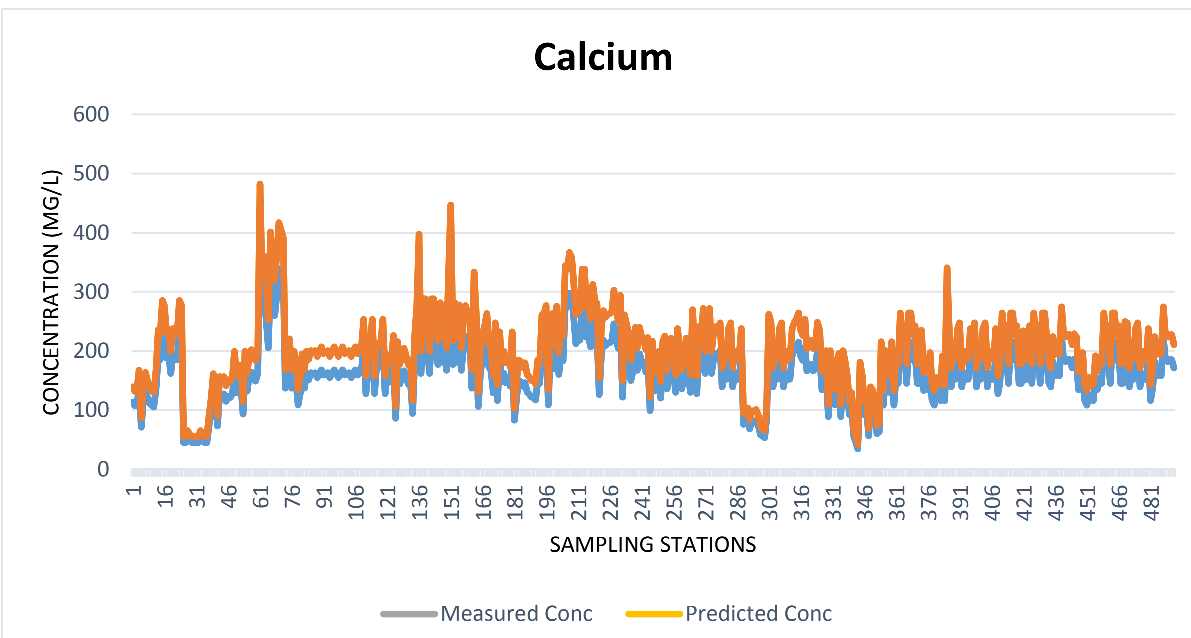


Fig 6.2 (e)

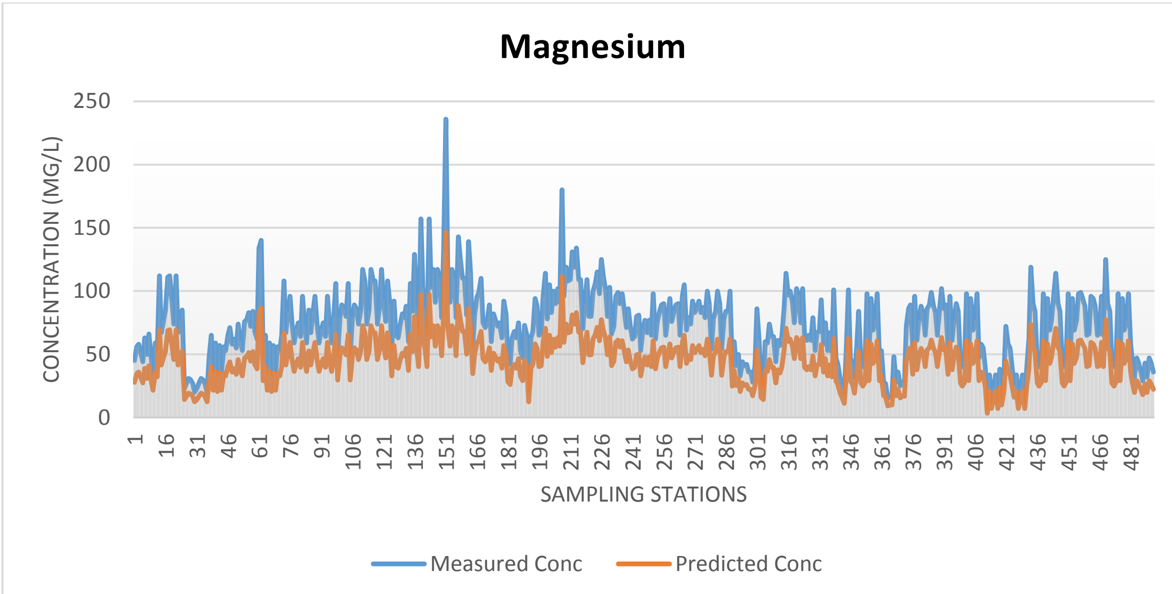


Fig 6.2 (f)

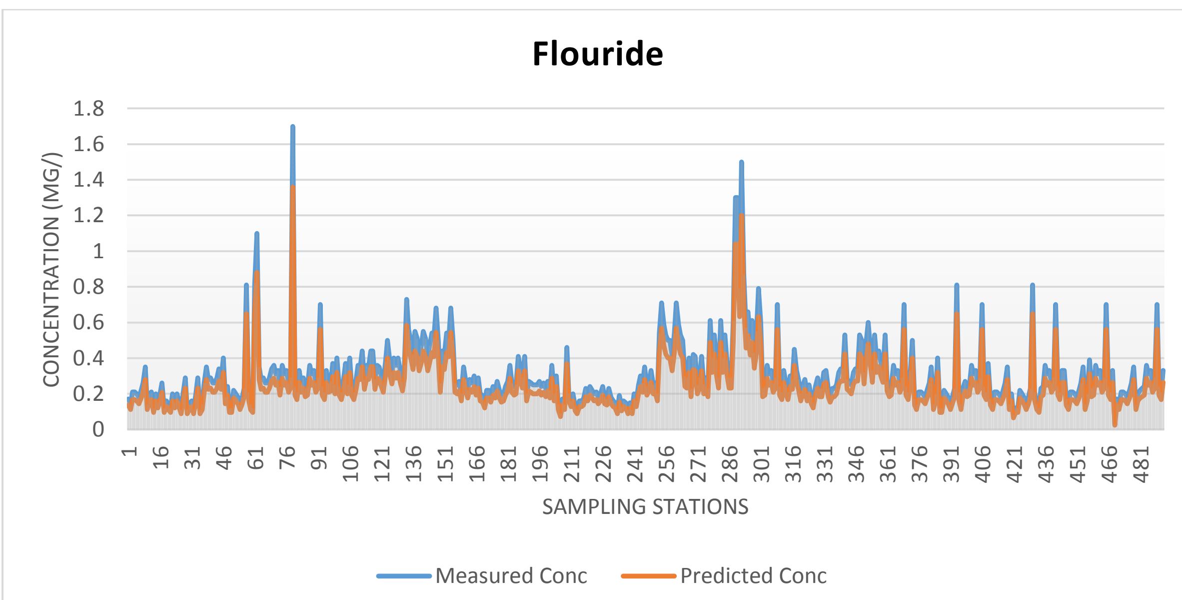


Fig 6.2 (g)

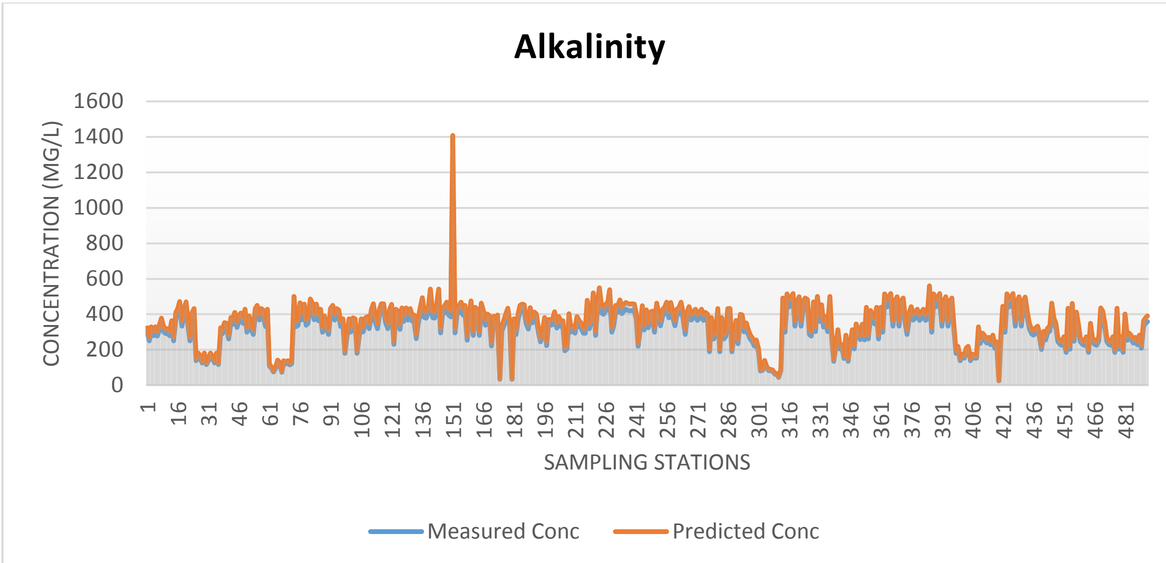


Fig 6.2 (h)

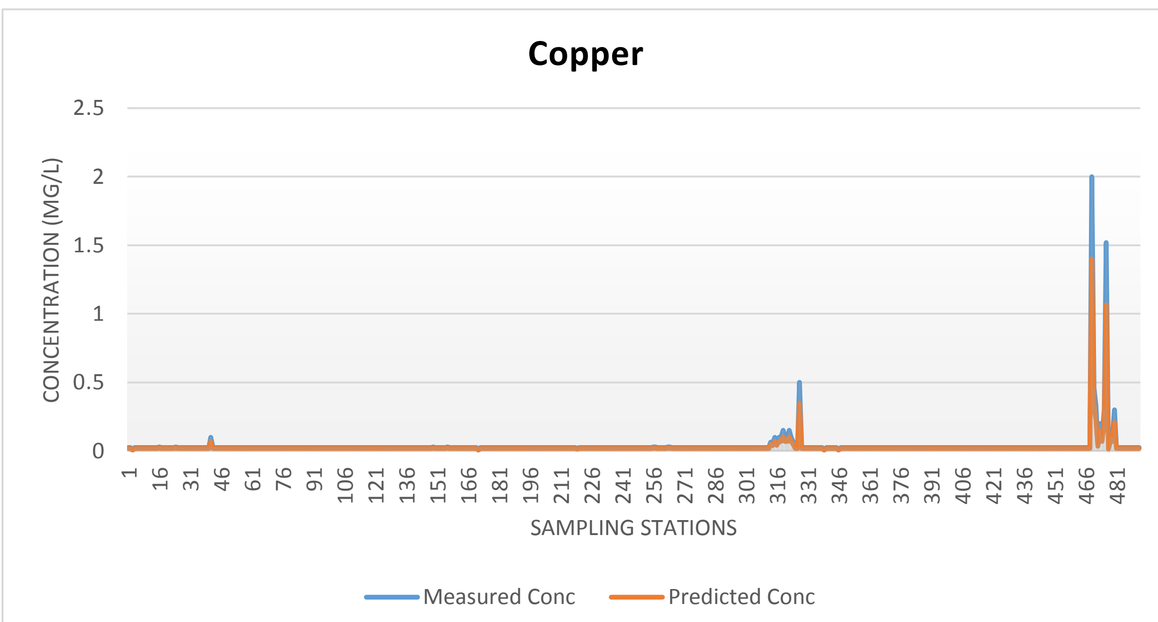


Fig 6.2 (i)

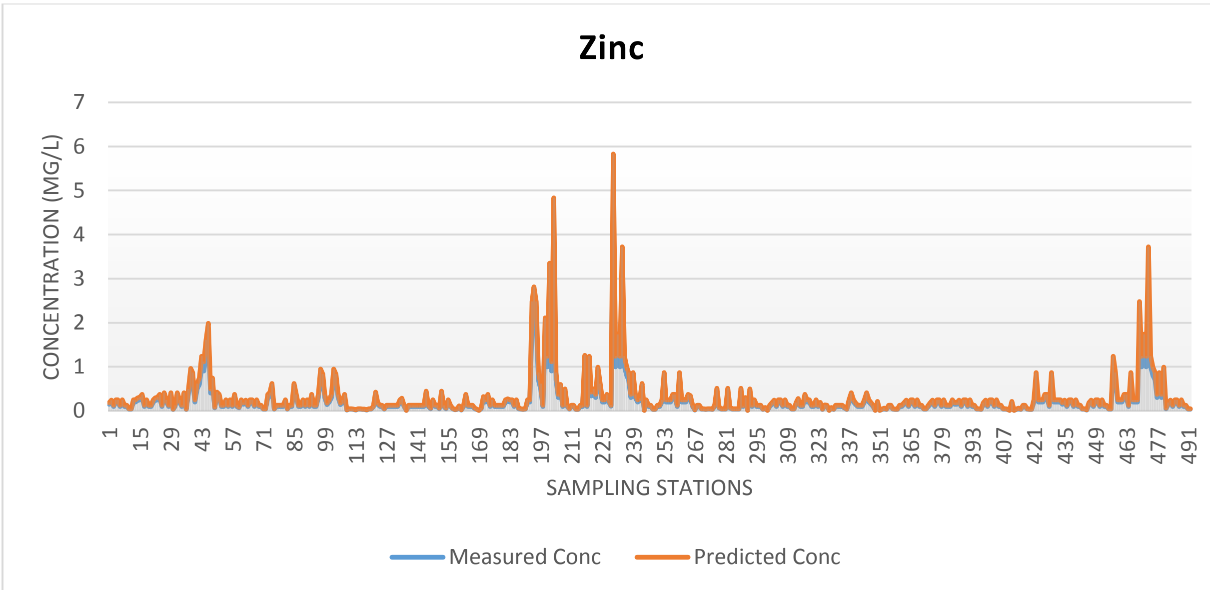


Fig 6.2 (j)

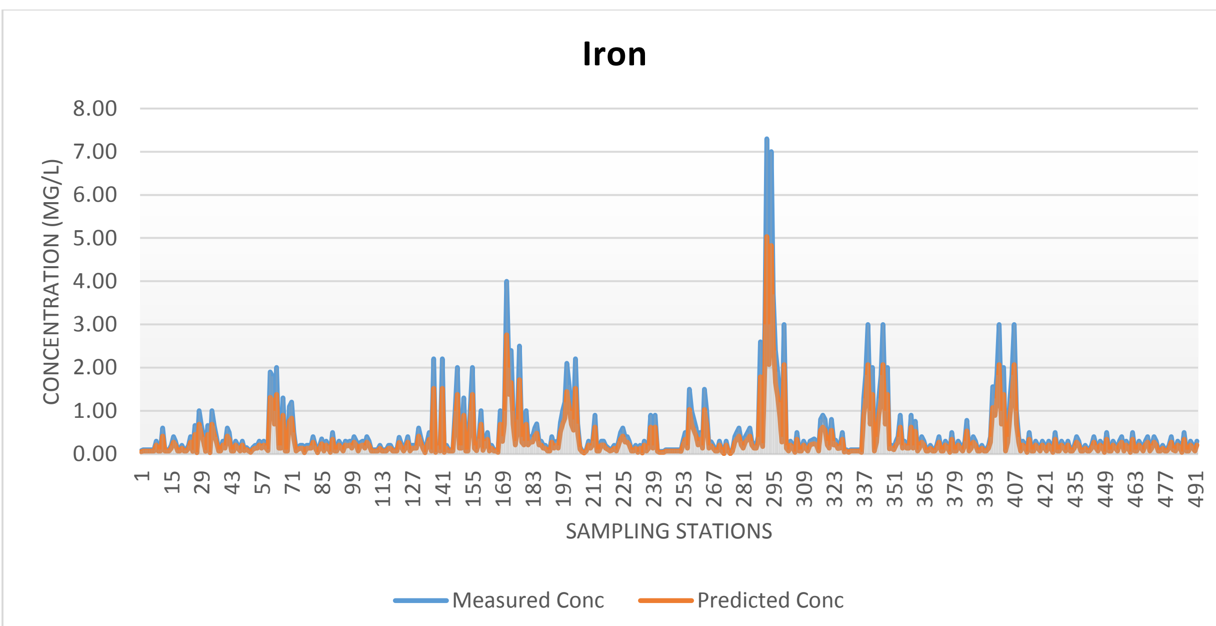


Fig 6.2 (k)

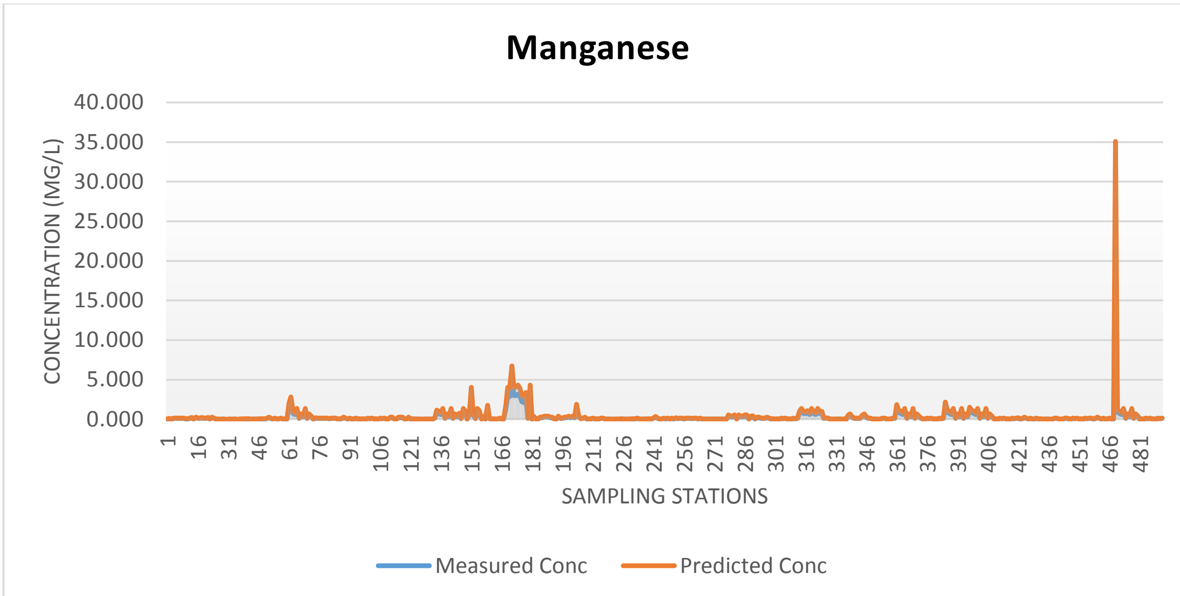


Fig 6.2 (l)

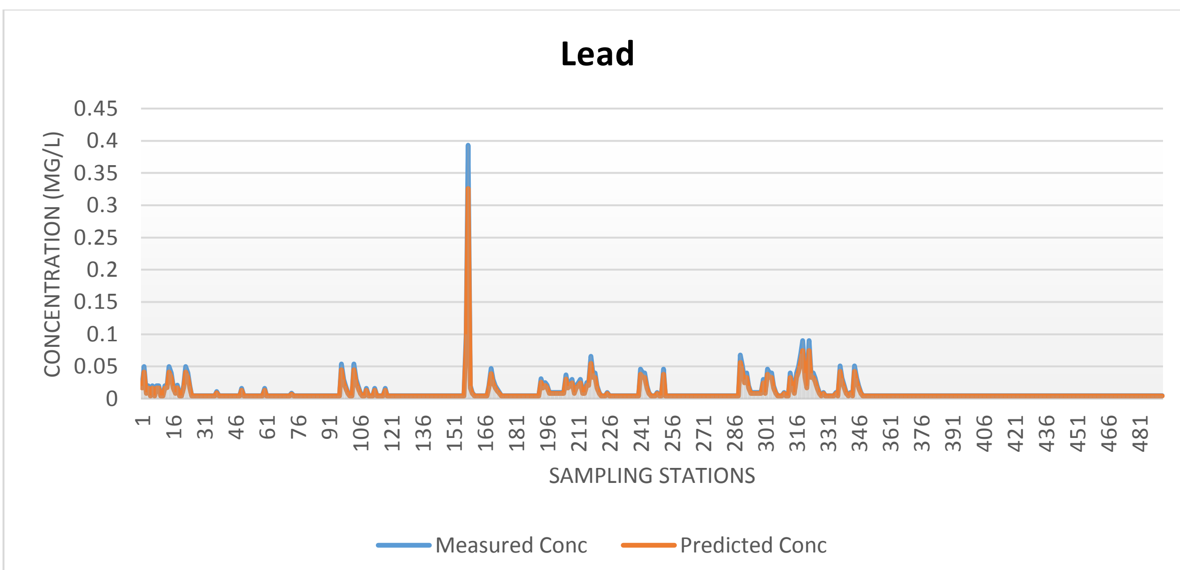


Fig 6.2 (m)

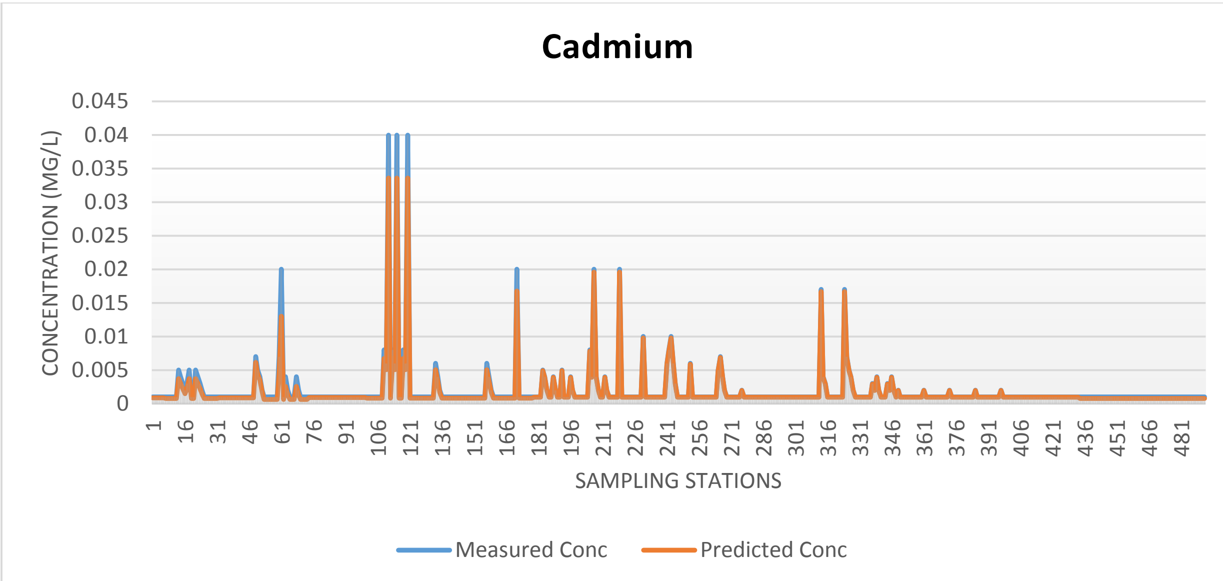


Fig 6.2 (n)

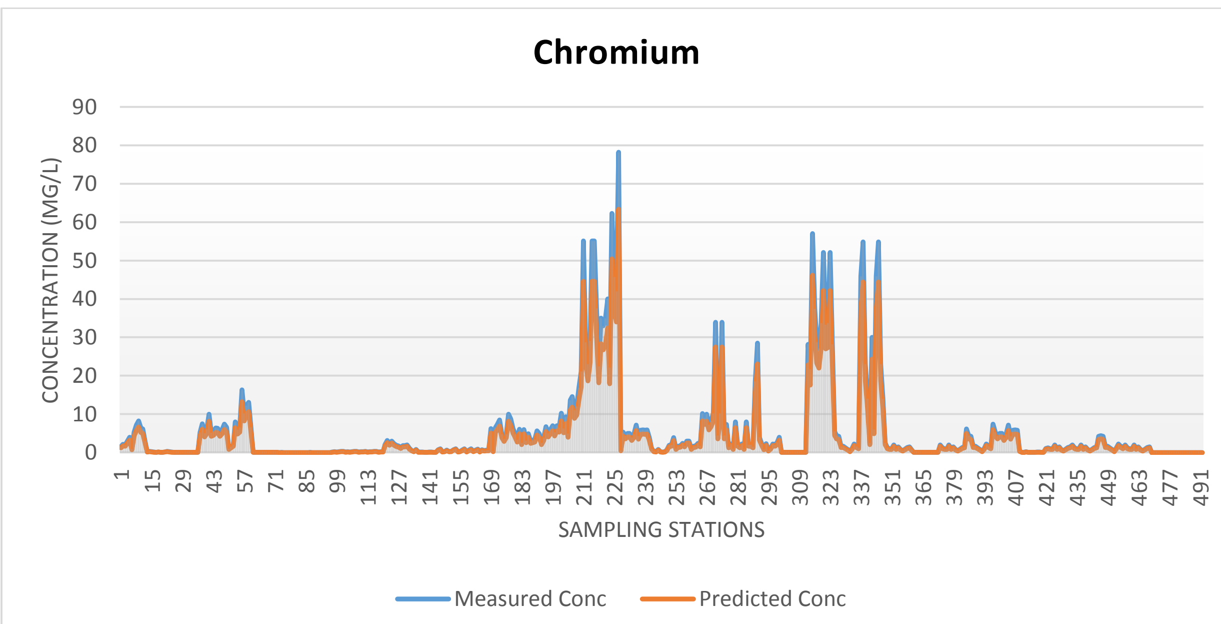


Fig 6.2 (o)

Fig 6.2(a)-6.2(o): Unmix Modelled Observed vs Modelled Line Plots of Groundwater parameters

From the line plots of observed vs predicted concentrations it was observed that the model overestimated the concentrations for the parameters sulphide, total hardness, magnesium, fluoride, iron and chromium.

However for parameters chloride, nitrate and calcium the model appears to have underestimated the concentrations. Also it is evident from the plots that at peak values the model has overestimated the concentrations.

After calculating the source contributions using Unmix model, the groundwater quality data was subjected to Positive matrix factorization model to find the source contribution.

6.4 POSITIVE MATRIX FACTORIZATION (PMF)

The EPA PMF (version 5.0) model was run on a total of 20 parameters of groundwater samples. Two input files, one with concentrations of the observed parameters and the other with the uncertainties associated with these concentrations, were used for data analysis. Detection limits of each variable and the error were considered to calculate the matrix of measurement uncertainties for. The model was set to keep running in the robust mode to decrease the impact of outliers. Signal-to-noise ratio (S/N), decided by the model for each component, took into consideration grouping of the component/pollutant as “strong”, “bad” or “weak”. Diverse runs were completed by down weighting the species classification from "strong" to "weak" to those species with low signal/noise ratio, however there was no improvement in the results. Since all the variables exhibited good signal to noise ratio, all variables were included by considering them“strong”. All the elements of groundwater samples had scaled residuals within 3.0 and −3.0, indicating the well modeled and thus the data classification was not changed (Bhuiyan et al. 2015).

The ‘‘optimal’’ solution was considered to have a Q value near the theoretical Q value and a solution that did not depend on the initial seed suggesting that a stable solution was obtained. The robust Q values were close to the true Q values in this study which entails the reasonable fitness of the model with the outlier. Additionally, it is important that the span of Q values should be sufficiently less from the random runs (100 runs in the present study). This will corroborate the attainment of a similar global minimum thus confirming the fitness of outliers (Guo et al. 2017).

6.4.1 Source Identification

The results given by seven factors gave a sensible elucidation of the groundwater quality information suggesting pollution sources. Profiles and contributions of base run 16 were inspected to recognize components of every parameter. While, the seven factors in groundwater tests gave a comprehensive picture of source identification of the polluted groundwater samples (Table 6.5).

- Strong positive loadings with chromium, cadmium and zinc were observed in Source 1, pointing towards unique source. Around 60 electroplating industries are located in Peenya Industrial area. Many of these industries were engaged in hard chrome/chrome plating and zinc plating, few units are engaged in copper, tin and nickel plating. The contamination of ground water due to illegal discharge of electroplating effluent into drains or due to seepage of effluent from underground storage tanks of these industries has contributed to chromium, zinc and cadmium content in the ground water (CPCB 2014). Therefore this factor can be attributed as “chromium electroplating”.
- Source 2 was identified as “geologic” since it had strong contributions with fluoride and moderate contribution with ammonia and alkalinity. High fluoride content is present for most part in gnessic and granitic territories. Granitic gneisses are the most seasoned arrangements in the state and have experienced greatest weathering. The joints, breaks, faults and vertical openings in the formations are possessed by fluoride-bearing minerals. The leachable fluoride in these minerals is reflected in the upper aquifer system (DMG 2011). However the leachability of fluoride is

governed by dissolved carbon dioxide in the soil (bicarbonates) which confirms this speculation.

Table 6.5: PMF Modelled Percentage Source contribution to groundwater quality

Parameter	Source type							R^2
	Source 1	Source 2	Source 3	Source 4	Source 5	Source 6	Source 7	
pH	71.2	0.10	0.60	4.70	23.40	-	-	0.85
Turbidity	4.10		93.80	-	2.10	-	-	0.81
TDS	-	22.30	75.20	-	0.30	2.11	-	0.75
SO_4	-	16.60	-	76.30	1.60	5.50	-	0.90
Cl	-	7.50	6.00	76.80	1.30	8.40	7	0.78
NO_3	-	4.20	10.70	81.70	-	3.40	-	0.59
TH	0.48	11.12	80.30	0.50	0.40	7.20	-	0.89
Ca	0.60	11.60	76.50	1.10	1.10	9.10	-	0.77
Mg	1.60	9.60	83.20	0.70	0.20	4.70	-	0.75
F	7.5	80.90	2.10	3.30	3.80	2.40	-	0.77
HCO_3	3.1	43.50	46.30	3	2.20	1.81		0.82
NH_3	19.70	56.20	-	7.10	14.60	2.31	-	0.83
S^{--}	5.10	8.30	24.30	51.80	4.50	6.09		0.74
Cu	15.06			4.60			80.34	0.74
Zn	30.00	23.10	4.70		3.60	3	35.60	0.84
Fe	10	6.30	5.60	0.80	6.50	70.8	-	0.68
Mn	10.8			-	10.4	5.70	73.10	0.61
Pb	6.50	1	2.20	6.80	81.90	1.60		0.62
Cd	45.50	2	4.30	3.80	6.20	38.20		0.78
Cr	87.60	-			8.30	-	4.1	0.64

Source 1= Chromium Electroplating; Source 2= Geologic; Source 3= Natural source; Source 4 = Sewage; Source 5= Lead acid Battery Manufacturing; Source 6: Steel Processing industry; Source 7: Paint Shipping

- High compositions with turbidity, total dissolved solids, total hardness, calcium and magnesium hence Source 3 was interpreted as 'natural source'. It can also be seen that Source 3 contains typical hydro chemical variables originating from mineralization of geological components of the soil. Thus it represents the mineral group and the source can be attributed as "natural or bedrock", as reported by earlier researchers also (Drever, 1997; Kumar et al., 2006, Subba Rao et al., 2006). Also 70% of the mechanism controlling the chemical composition of groundwater of Bangalore city is controlled by rock-water interaction (DMG 2011).
- Elevated compositions with and Chloride, Nitrate and Sulphate were seen in Source 4. Nitrate in groundwater of the study area can be attributed to contamination from unlined drains and sewage effluent as there is no agricultural activity nor application of nitrogenous fertilizers as it is an urban area. The moderate loadings for sulphates and chloride can be attributed to seepage from sewers, septic tanks and industrial effluents. Thus the factor can be attributed as "sewage".
- The presence of heavy metal element lead with high composition in Source 5 suggested an industrial source. Lead acid battery manufacturing units use raw materials which are alloys of lead calcium, lead antimony, lead tin, pure lead and sulphuric acid in the study region, disposing off acid directly into the environment have contaminated the groundwater with lead and cadmium. A study conducted by Ramesh, A (2014), in order to determine the remediation of heavy metal contaminated soil and groundwater also found the presence of lead in groundwater as well as soil sample in the region. This similarity points out that the factor can be termed as "lead acid Battery manufacturing unit"
- Large percentage contribution from iron was noted in Source 6. The presence of Fe represents metal pollution derived from industrial effluents most probably from large steel rolling mills and steel forging industries that are located in the study area. Disposal of scrap iron in open areas due to industrial activity is one of the

factor causing higher values of iron in groundwater in the region (Basappa Reddy 2003).

- Strong representations with Copper and Manganese were observed in Source 7. Copper is used in alloys, as a catalyst, in anti-fouling paints. Thus, it basically represents a toxic metals group which is due to the spray painting and paint shipping activity in the area.

6.4.2 Model Performance Evaluation

- The accuracy of the PMF can be tested by comparing the calculated values with the measured concentrations for each parameter. The PMF comparison of measured and calculated concentrations are shown in Table 6.6.
- As observed from the correlation coefficients, the measured and predicted values exhibited good adequacy between them. Additionally, the ratio of mean PMF modelled and measured values of most of the groundwater quality parameters indicate goodness of the receptor modelling approach to the source apportionment of groundwater quality. Based on the R^2 values the accuracy of the model is very high for pH, chloride, total hardness, alkalinity, and zinc with R^2 between 0.8- 1, high for turbidity, total dissolved solids, calcium, magnesium, ammonia, sulphide, iron, manganese and total chromium with R^2 between 0.6-0.8, moderate for fluoride, copper, lead and cadmium with R^2 between 0.4-0.6 as shown in Table 6.6 below.

Table 6.6: PMF comparison of measured and calculated concentrations

Parameters	Measured Mean (M)	PMF Modelled (E)	Ratio (E/M)	R^2	% error
pH	6.75	6.42	0.95	0.89	4.89
Turbidity	9.79	7.60	0.77	0.69	22.37
TDS	1404.15	1584.84	1.12	0.72	-12.87
SO_4	202.43	251.98	1.24	0.64	-24.48
Cl	340.28	294.74	0.86	0.81	13.38
NO_3	40.31	47.62	1.18	0.78	-18.13
TH	718.76	798.45	1.11	0.82	-11.09
Ca	165.20	198.36	1.20	0.70	-20.07
Mg	72.54	98.46	1.30	0.71	-35.73
F	0.30	0.19	0.63	0.55	36.67
HCO_3	322.86	350.85	1.09	0.97	-8.67
NH_3	0.36	0.21	0.58	0.63	41.67
S^{--}	0.02	0.029	1.45	0.68	-45.00
Cu	0.03	0.017	0.56	0.49	43.33
Zn	0.25	0.19	0.76	0.88	24.00
Fe	0.46	0.25	0.53	0.61	45.65
Mn	0.31	0.25	0.80	0.75	19.35
Pb	0.01	0.015	1.50	0.44	-50.00
Cd	0.002	0.0011	0.55	0.40	45.00
Cr	5.21	4.68	0.89	0.76	10.17

6.4.3 Observed vs predicted Line Plots

The Line plots of observed vs modelled values for the parameters are shown below.

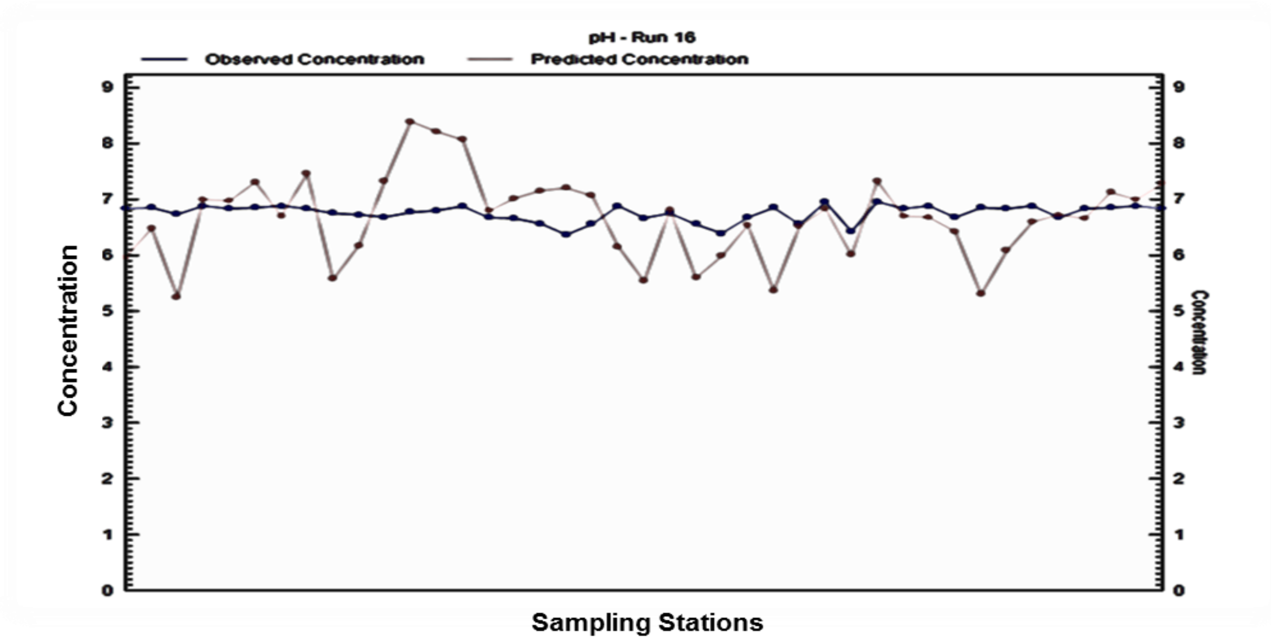


Fig 6.3 (a) pH

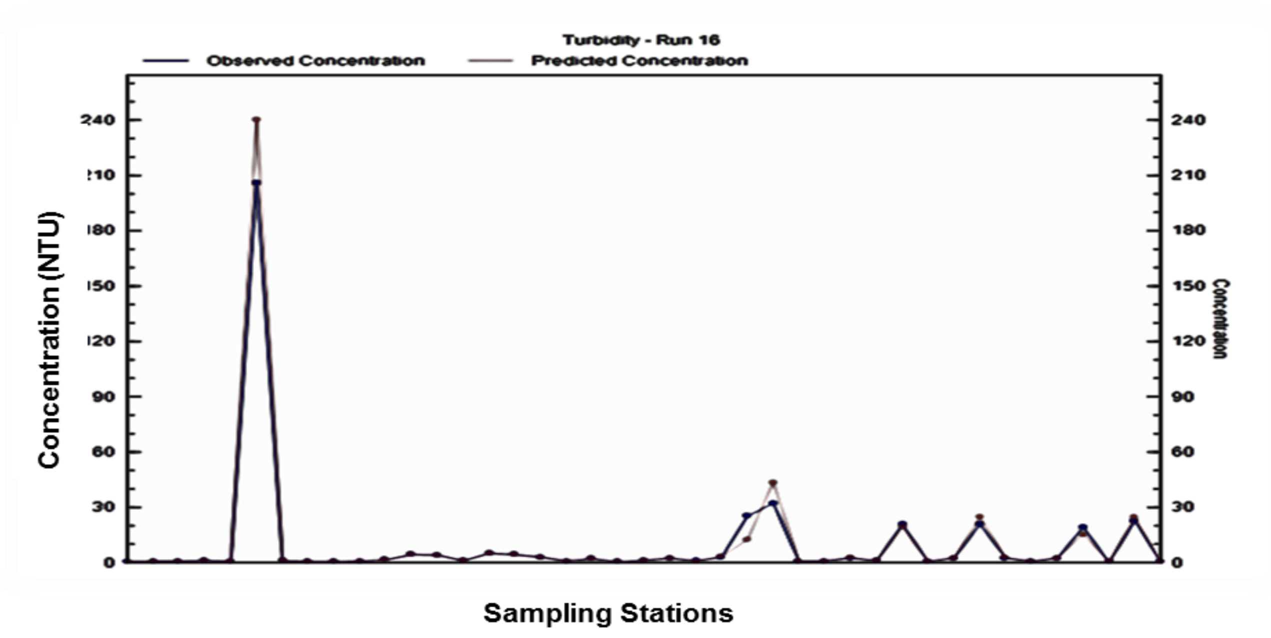


Fig 6.3 (b) Turbidity

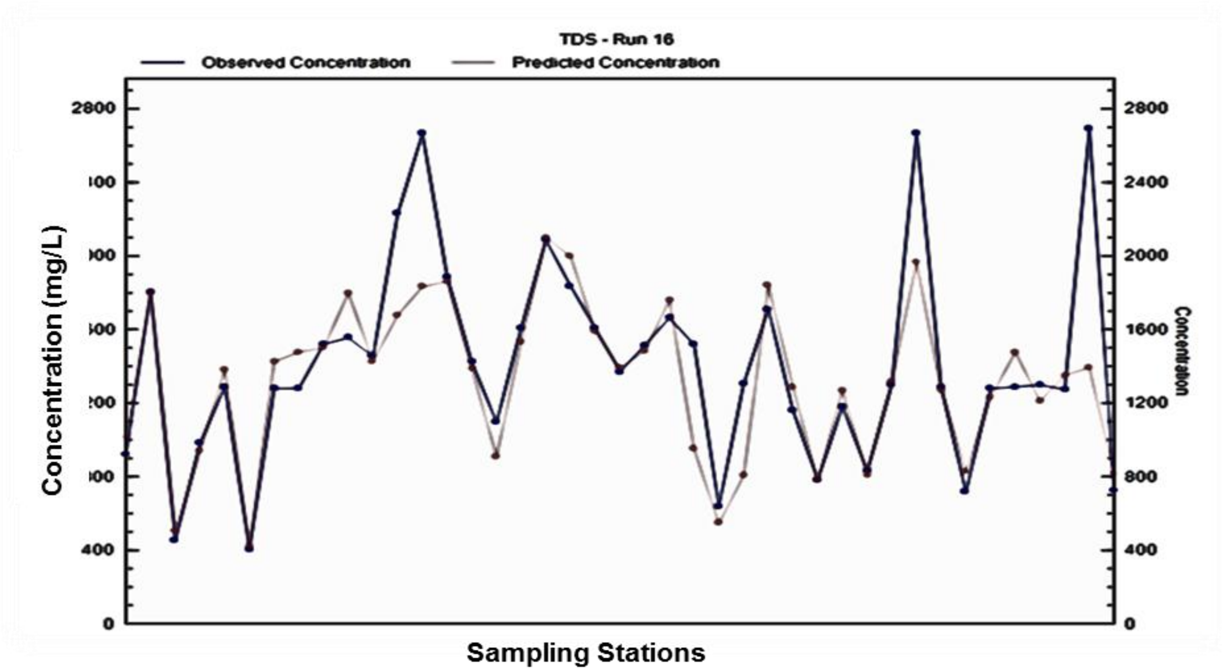


Fig 6.3 (c) Total Dissolved Solids

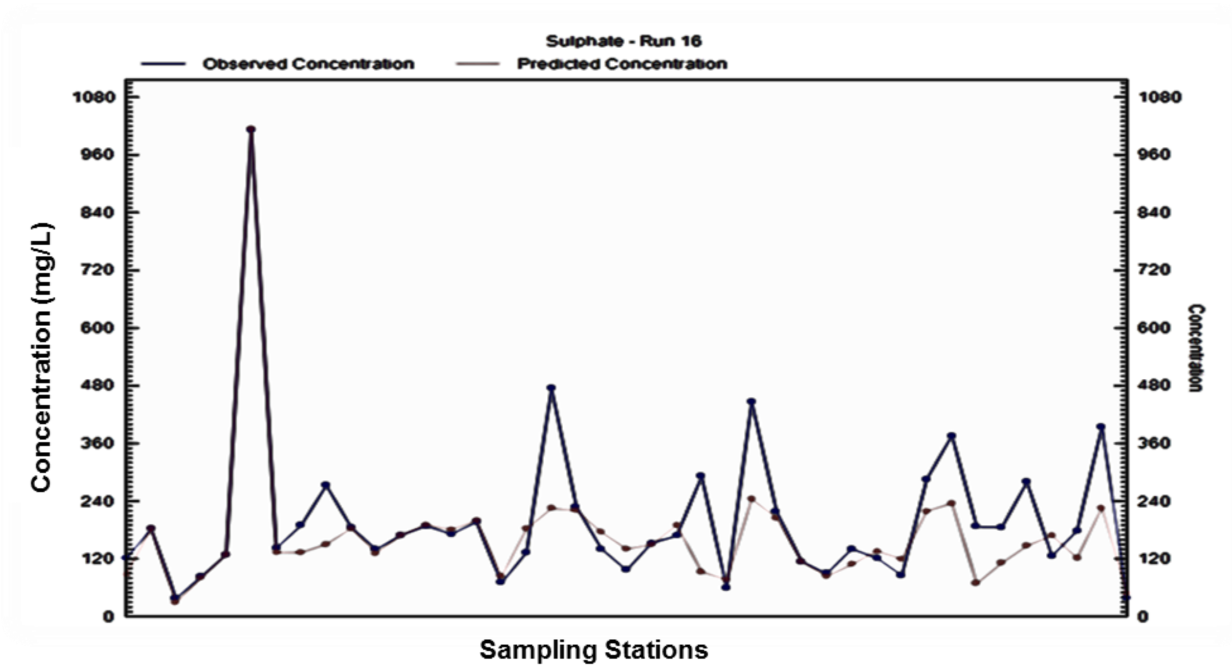


Fig 6.3 (d) Sulphate

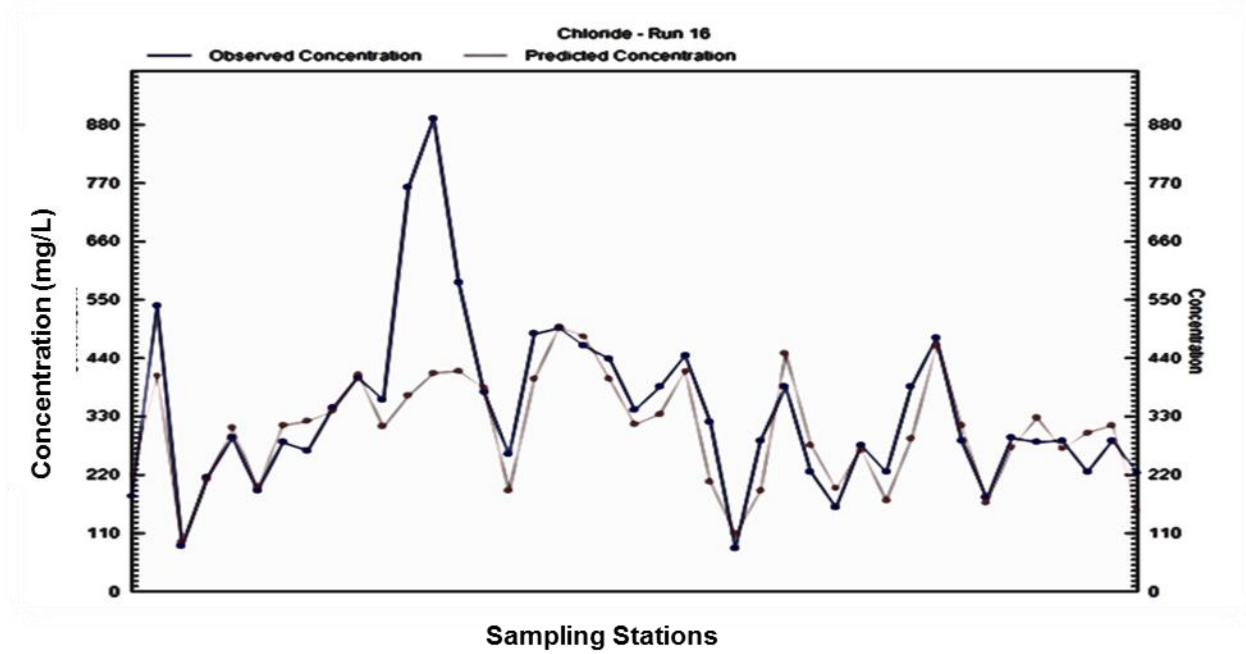


Fig 6.3 (e) Chloride

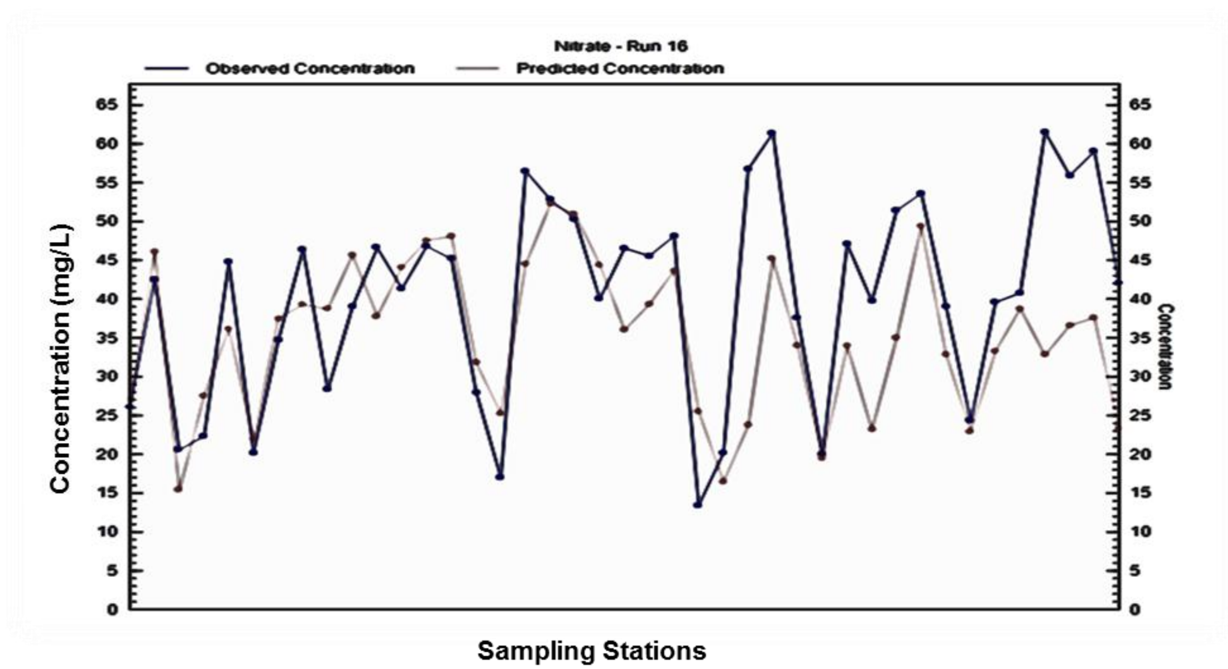


Fig 6.3 (f) Nitrate

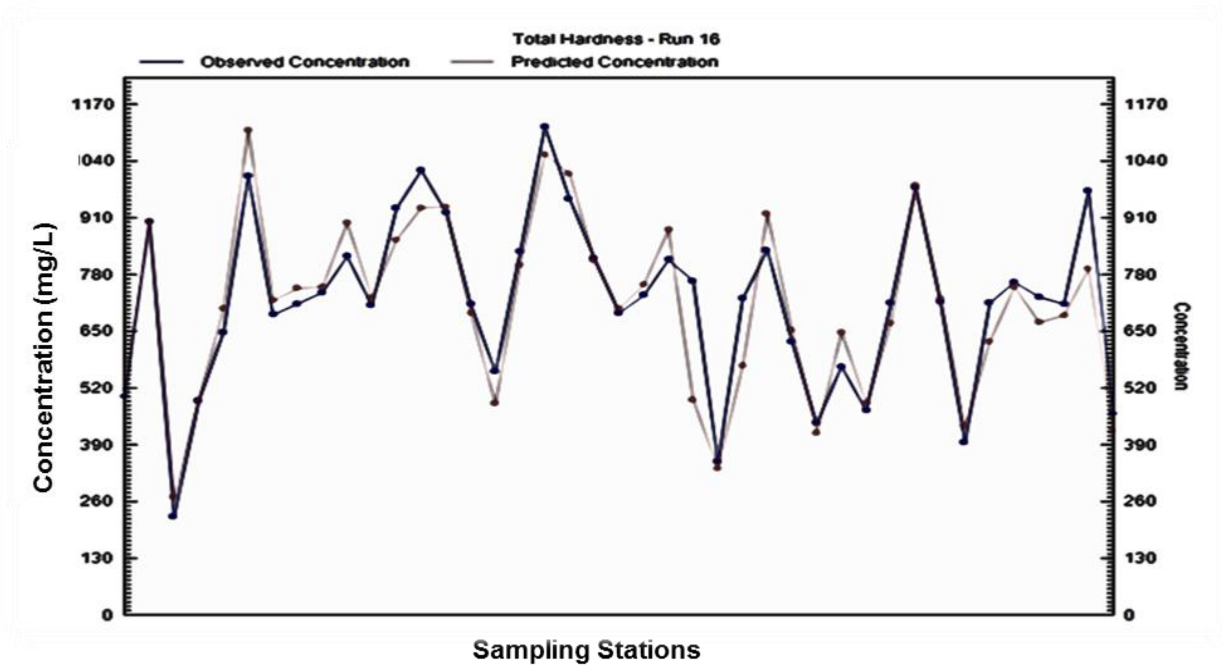


Fig 6.3 (g) Total Hardness

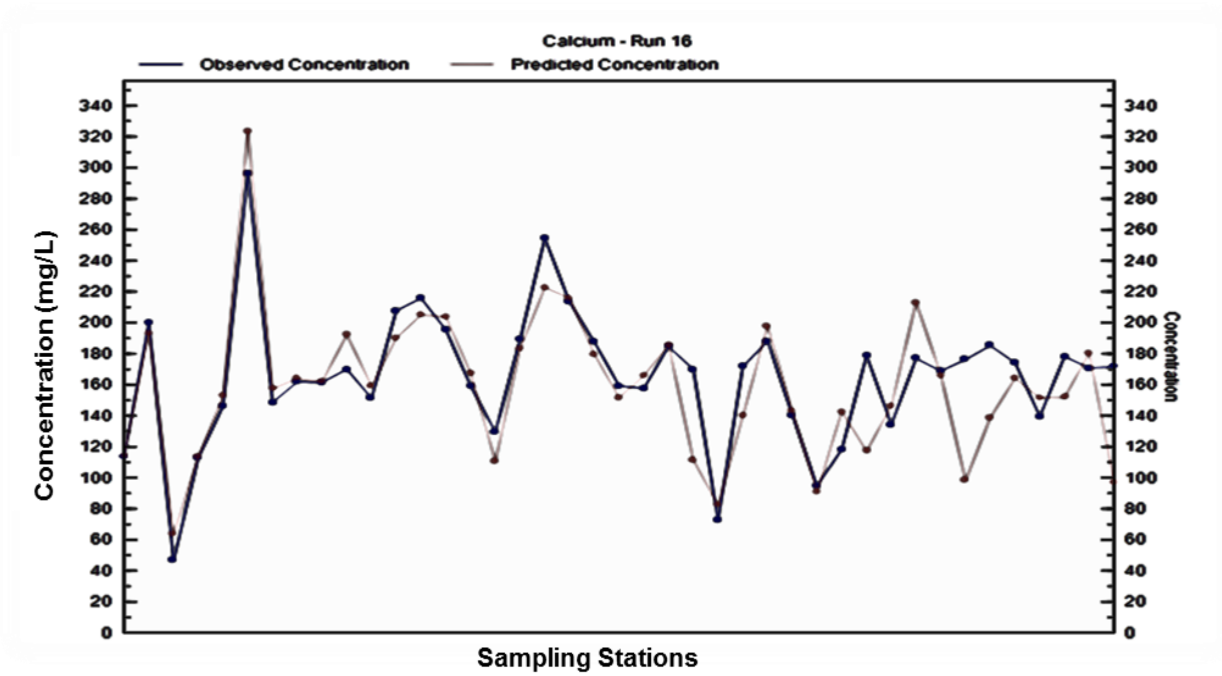


Fig 6.3 (h) Calcium

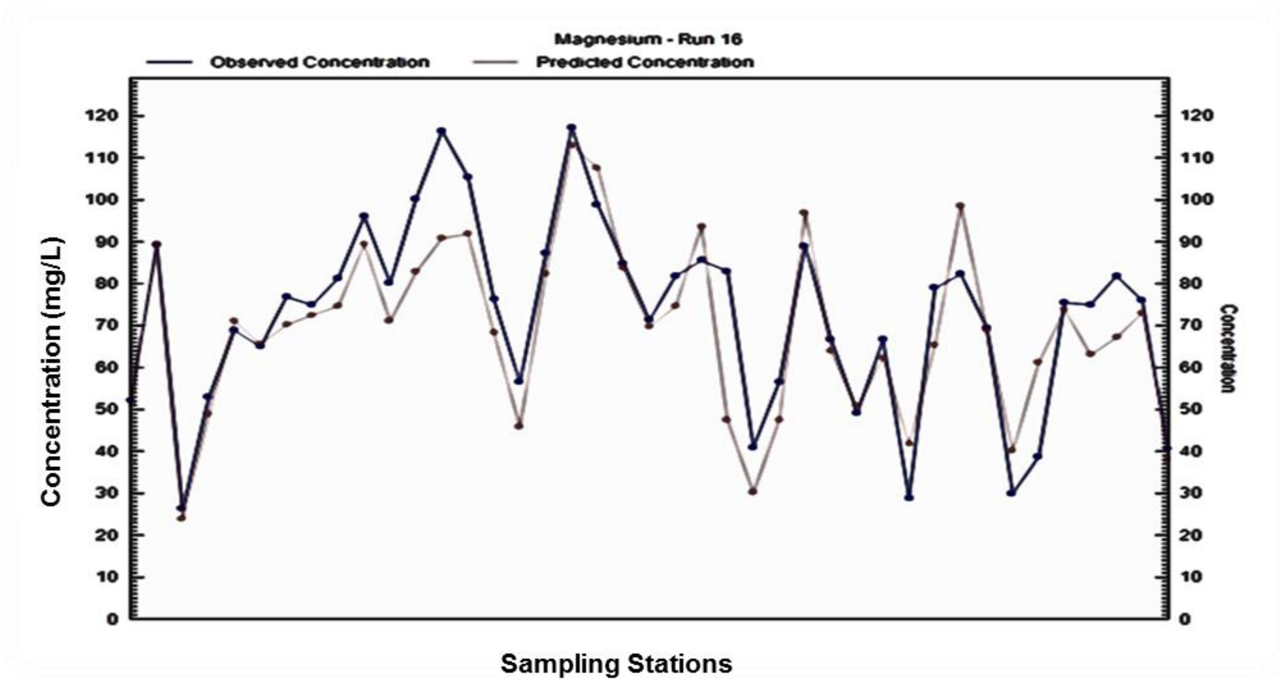


Fig 6.3 (i) Magnesium

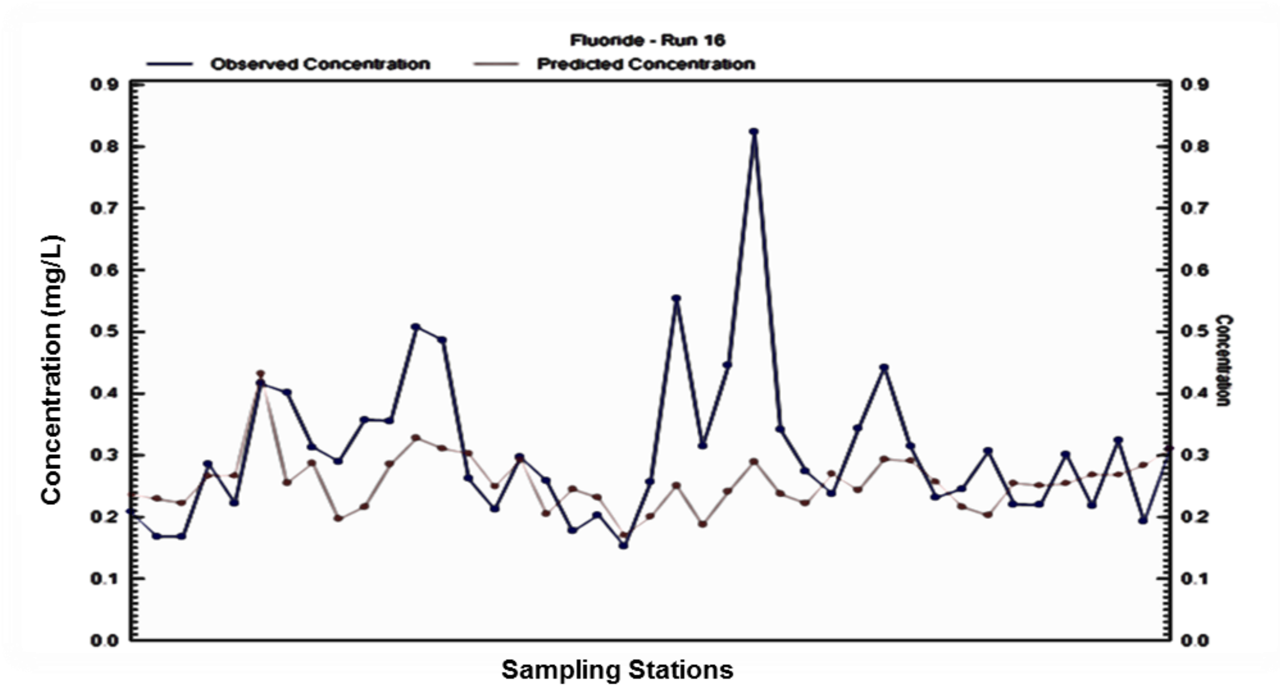


Fig 6.3 (j) Flouride

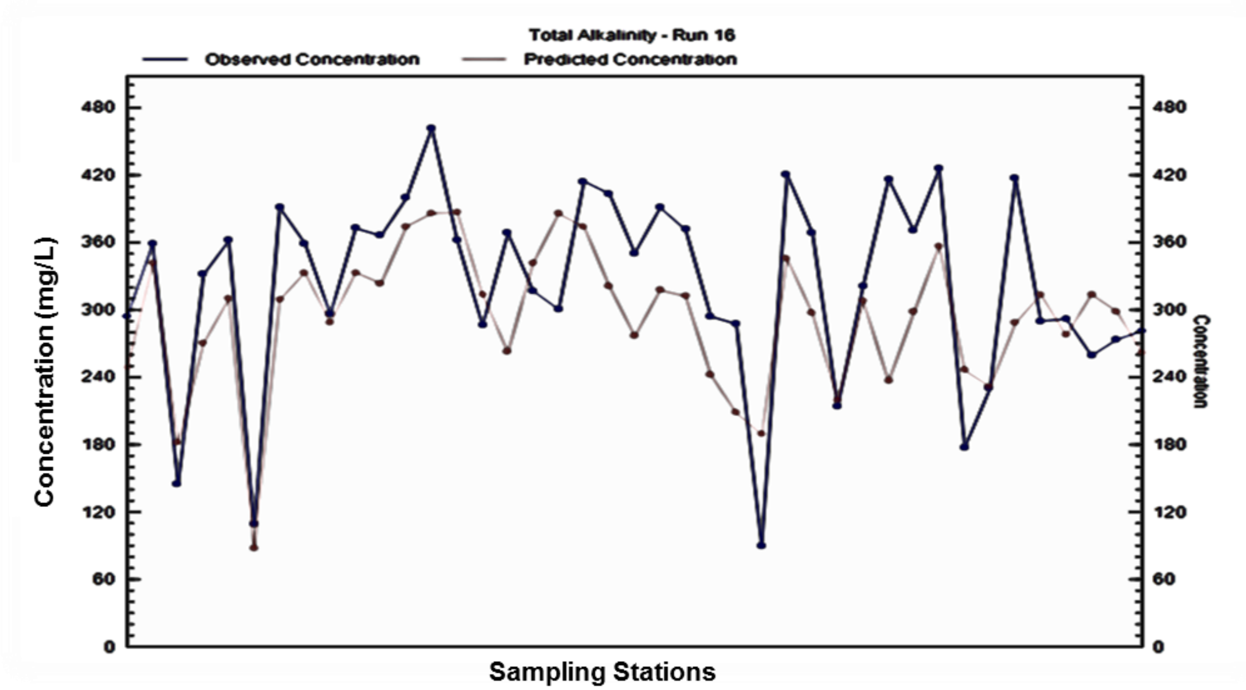


Fig 6.3 (k) Alkalinity

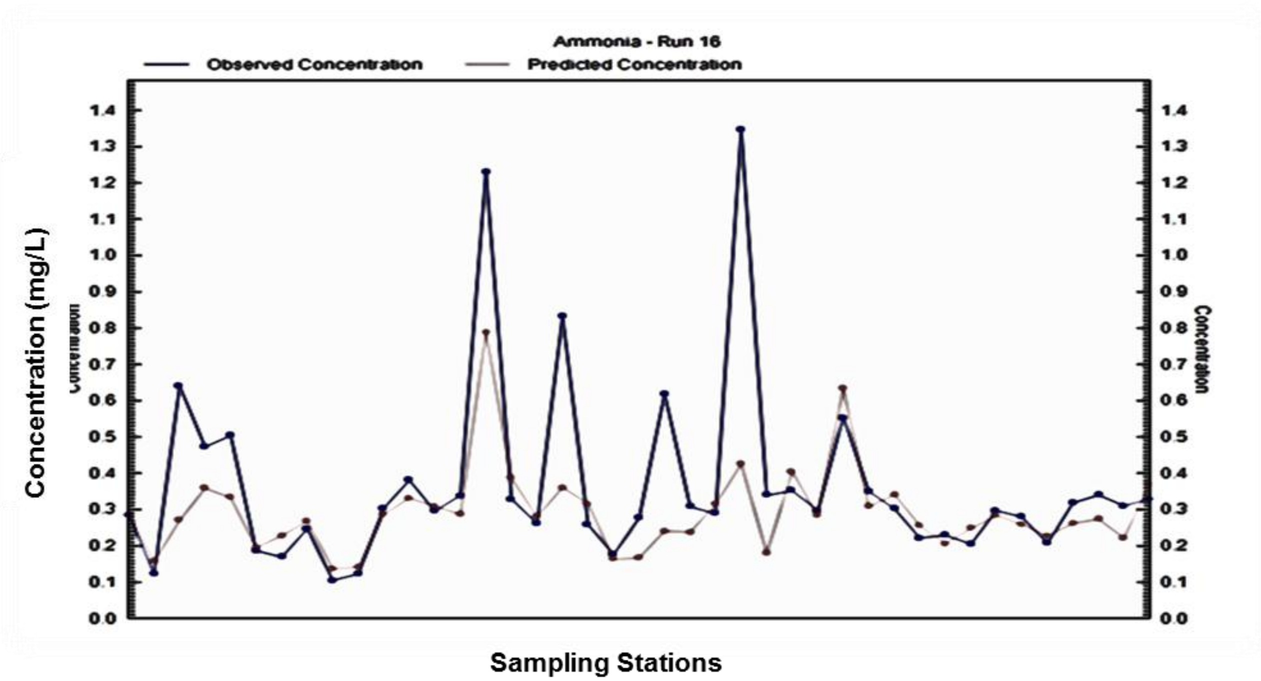


Fig 6.3 (l) Ammonia

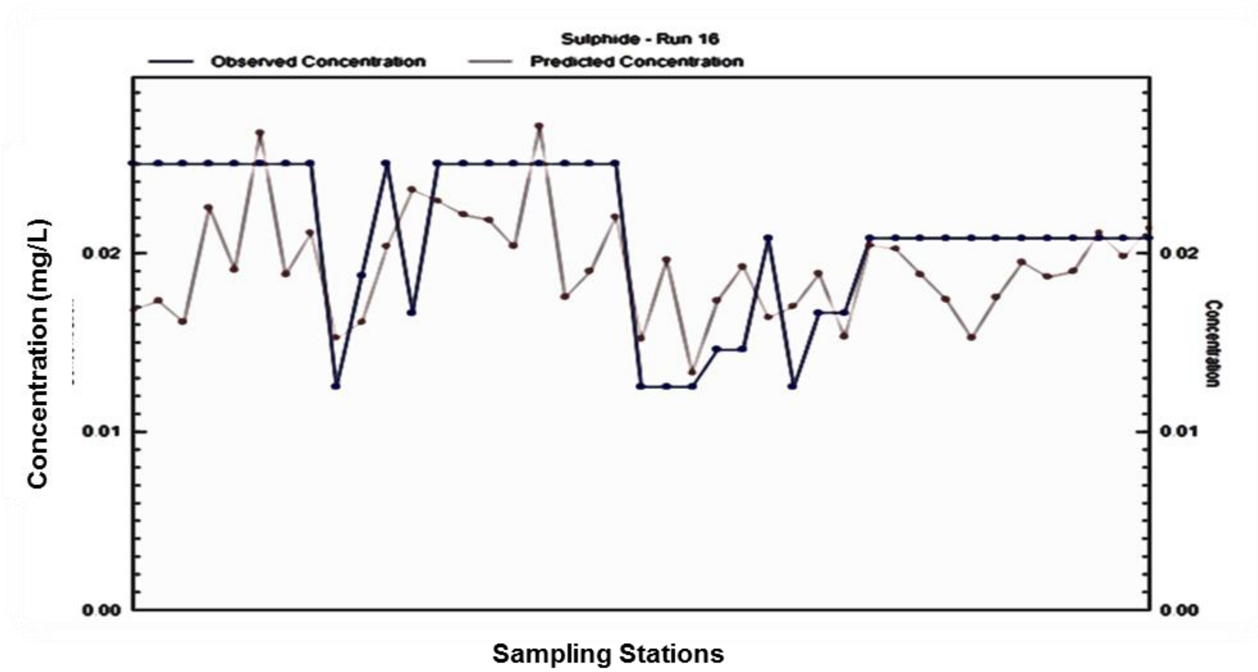


Fig 6.3 (m) Sulphide

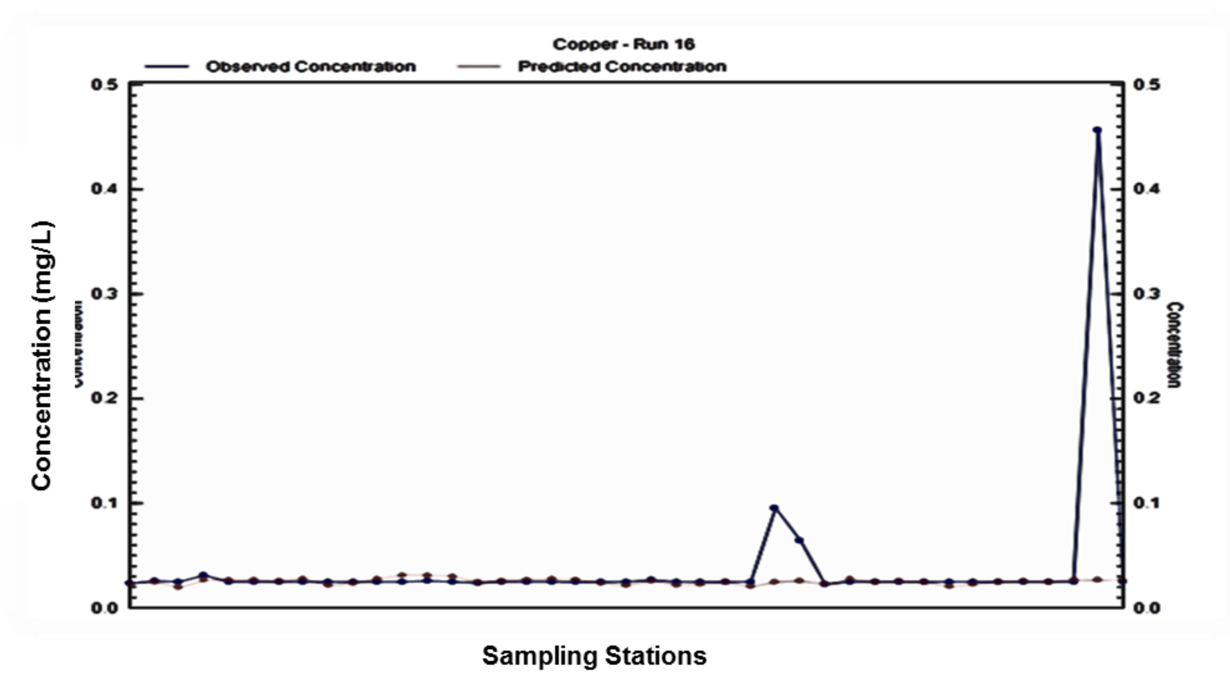


Fig 6.3 (n) Copper

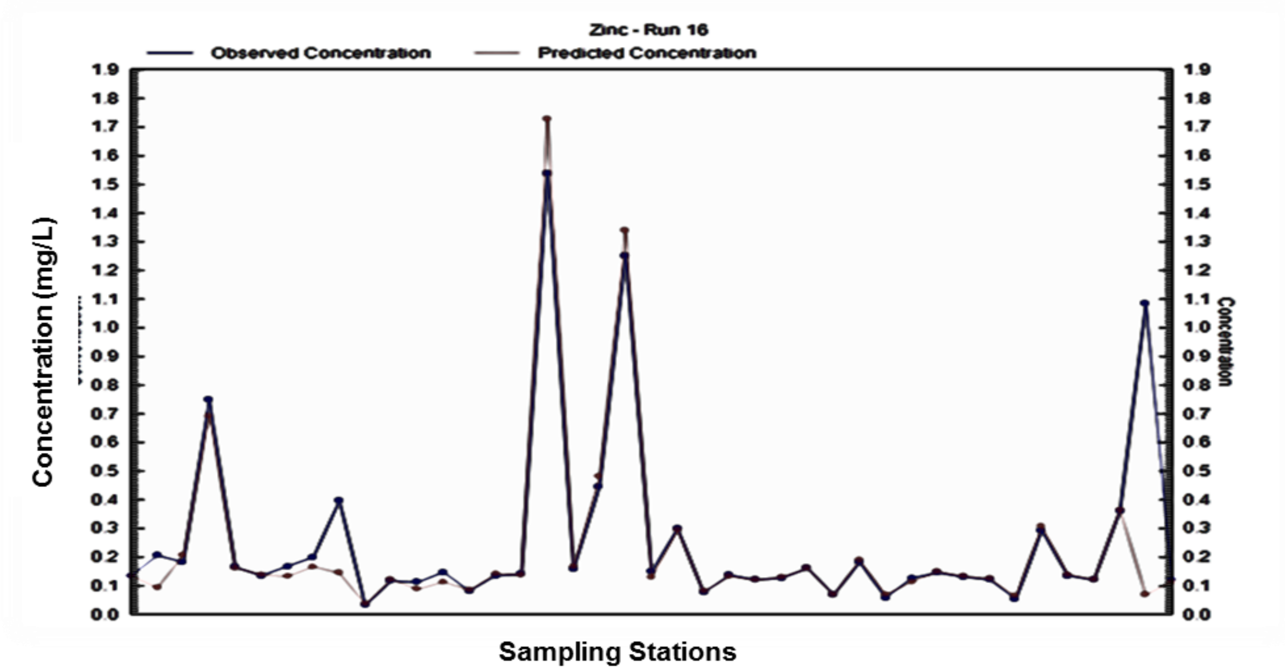


Fig 6.3 (o) Zinc

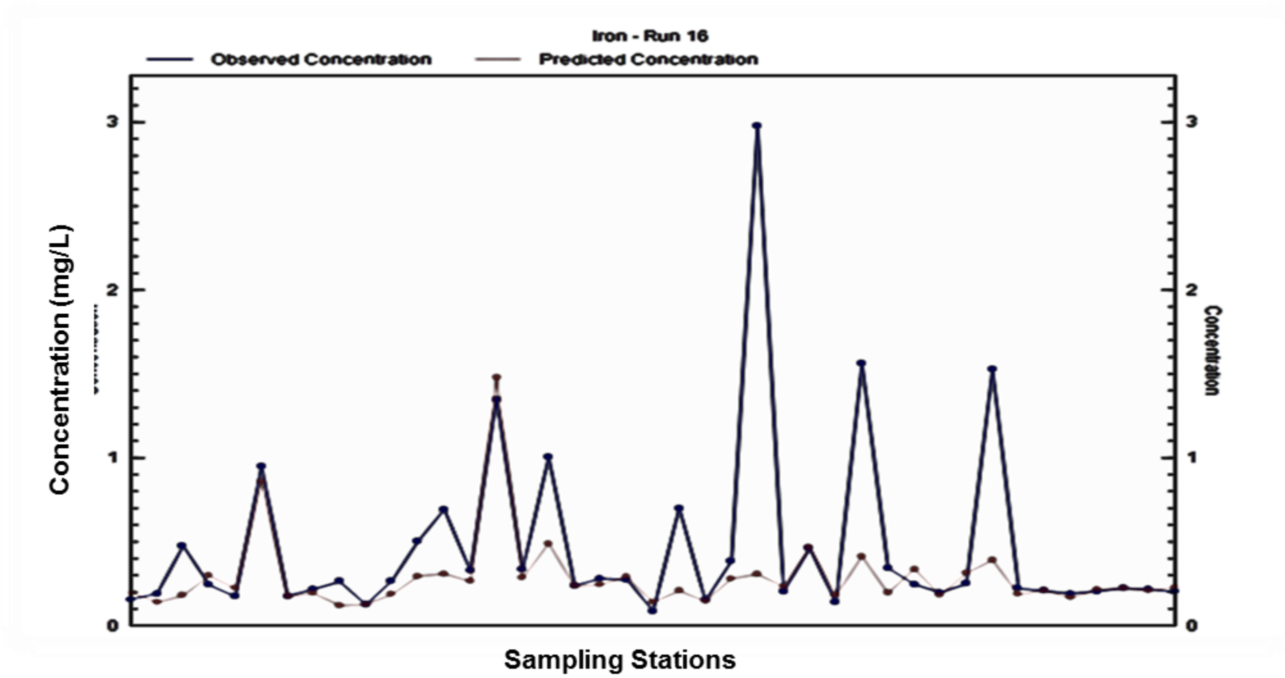


Fig 6.3 (p) Iron

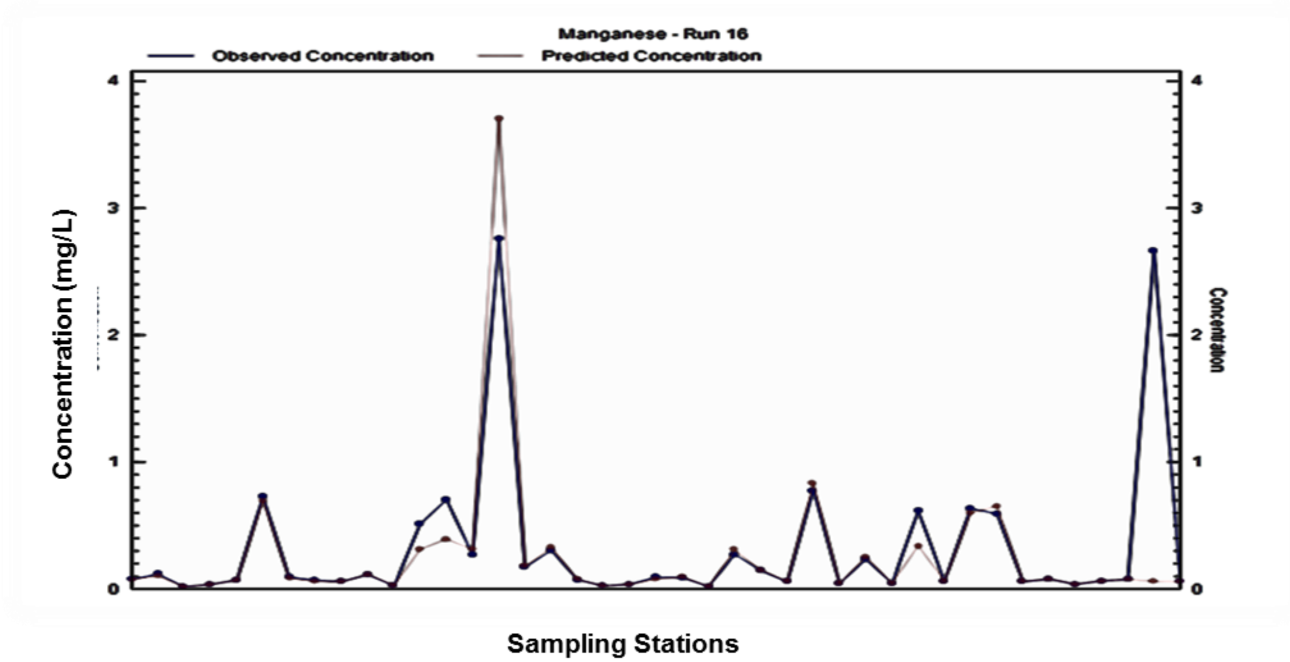


Fig 6.3 (q) Manganese

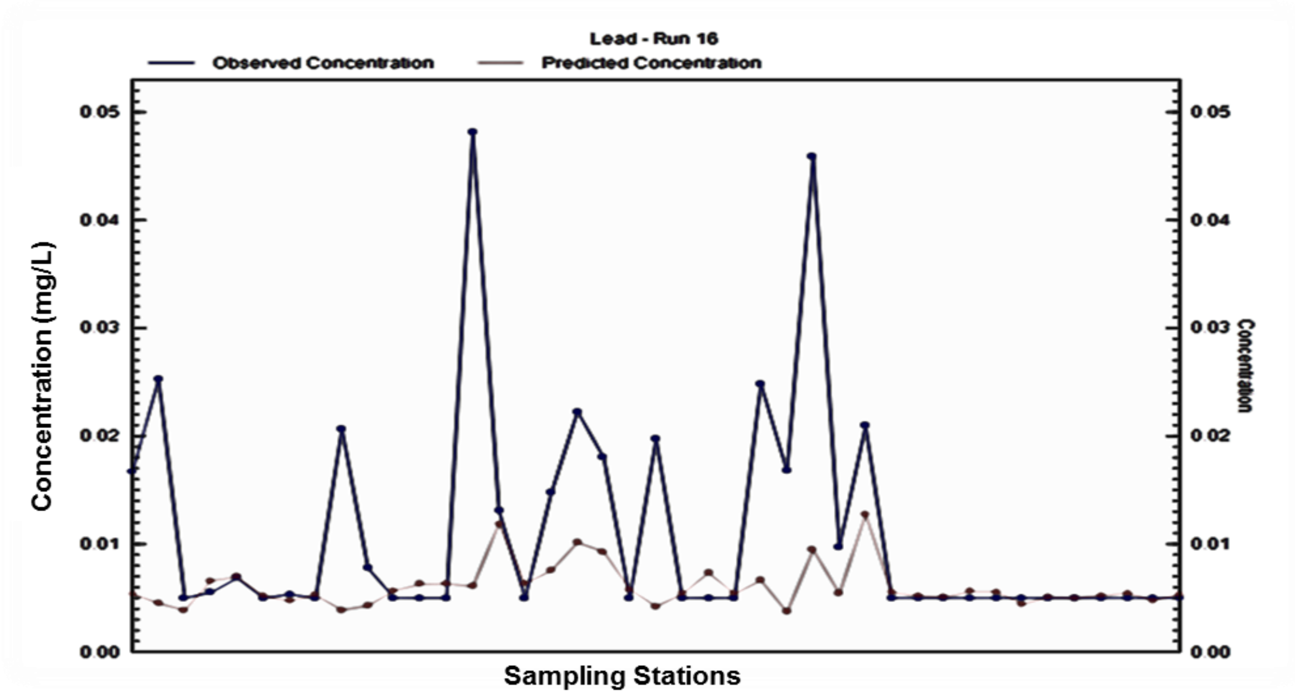


Fig 6.3 (r) Lead

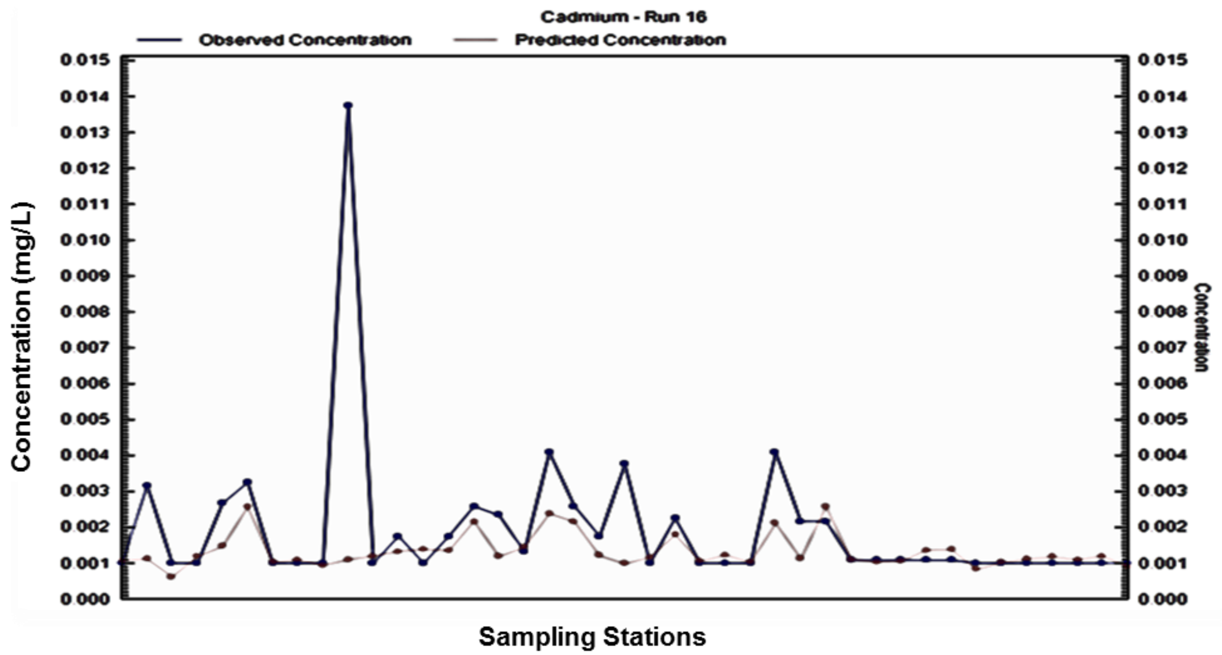
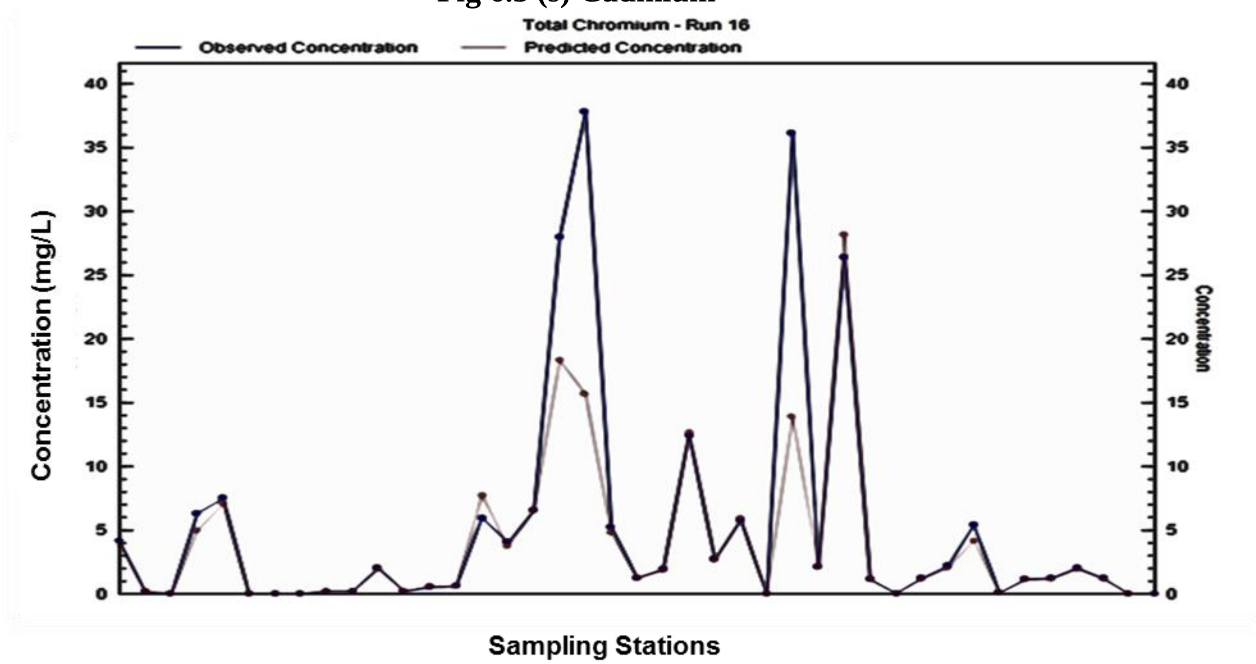


Fig 6.3 (s) Cadmium



6.3(t): Total Chromium

Fig 6.3(a) – 6.3(t): PMF Modelled Observed vs Modelled Line Plots of Groundwater parameters

The line plots of observed vs predicted concentrations indicate that the model underestimated the concentrations for the majority of the parameters except turbidity and calcium. Also it is evident from the plots that at peak values the model has overestimated the concentrations which may be because the method relies on many parameters, model input, missing values and values below MDL which have an impact on the quality of results. Due to this uncertainty the estimated values is higher than the one of measured ones.

6.5 PERFORMANCE EVALUATION OF APCS-MLR, UNMIX AND PMF MODELS

- Three different receptor models (APCS-MLR, Unmix and PMF) were applied to apportion the groundwater quality in Peenya industrial region of Bangalore city in with a view to compare the outcomes and the adequacy of models. The comparison of the three models was performed by considering distinctive features: the fitting quality between the measured and the modeled ones, the number and nature of the recognized sources and finally the contribution of each source
- The nature of the models was portrayed by regressing the parameters modeled for each model against the one measured. It was discovered that all three models gave great outcomes in regards to their capacity to recreate measured concentrations with fundamentally alike sources in most of the cases but with the APCS-MLR model showing the best correlation and the closest slope to the unity as APCS-MLR looks for variations in the concentrations of different components and identifies a minimal set of “unobservable factors” which can explain these variations.
- Unmix model showed comparatively higher error 37% which may be due to uncertainties, whereas 20% to 25% error was observed for APCS-MLR and PMF models. Unmix was quite conservative as it could not differentiae few sources.

Table 6.7: Comparison among measured and APCS-MLR, PMF and Unmix model calculated concentrations

Parameter	Measured	APCS-MLR			UNMIX			PMF		
		Modeled	r^2	% error	Modeled	r^2	% error	Modeled	r^2	% error
pH	6.75	7.10	0.65	-5.19	-	-	-	6.42	0.89	4.89
Turbidity	9.79	6.89	0.81	29.62	-	-	-	7.60	0.69	22.37
TDS	1404.15	1632.45	0.75	-16.26	-	-		1584.84	0.72	-12.87
SO_4	202.43	237.68	0.90	-17.41	170.46	0.89	15.79	251.98	0.64	-24.48
Cl	340.28	298.00	0.78	12.43	368.46	0.92	-8.28	294.74	0.81	13.38
NO_3	40.31	32.00	0.59	20.62	54.62	0.78	-35.50	47.62	0.78	-18.13
TH	718.76	590.00	0.89	17.91	620.45	0.62	13.68	798.45	0.82	-11.09
Ca	165.20	179	0.77	-8.35	205.36	0.70	-24.31	198.36	0.70	-20.07
Mg	72.54	76	0.75	-4.77	45.46	0.69	37.33	98.46	0.71	-35.73
F	0.30	0.24	0.77	20	0.24	0.74	20.00	0.19	0.55	36.67
HCO_3	322.86	384.00	0.82	-18.94	350.85	0.97	-8.67	350.85	0.97	-8.67
NH_3	0.36	0.29	0.83	19.44		-		0.21	0.63	41.67
S^{--}	0.02	0.026	0.74	30.00	0.018	0.98	10.00	0.029	0.68	-45.00
Cu	0.03	0.02	0.74	33.33	0.021	0.86	30.00	0.017	0.49	43.33
Zn	0.25	0.19	0.69	24.00	0.31	0.68	-24.00	0.19	0.88	24.00
Fe	0.46	0.55	0.68	-19.57	0.32	0.81	30.43	0.25	0.61	45.65
Mn	0.31	0.22	0.61	29.03	0.42	0.65	-35.58	0.25	0.75	19.35
Pb	0.01	0.013	0.62	-30.00	0.012	0.74	-20.00	0.015	0.44	-50.00
Cd	0.002	0.0028	0.78	-40.00	0.0017	0.80	15.00	0.0011	0.40	45.00
Cr	5.21	3.87	0.64	25.72	4.25	0.78	18.43	4.68	0.76	10.17

- The correlation coefficients very often high, suggesting acceptable results. A reasonable agreement between PCA and PMF was found, with both models identifying the same sources and with good correlations for the same identified sources.
- Although APCS-MLR, Unmix and PMF identified similar sources the Unmix model was not able to identify the steel processing industry and was little limited in recognizing the sources, six versus seven from the PMF model.
- EPA Unmix and PMF can provide source contributions in the output window. The source contributions and comparison from APCS-MLR, Unmix and PMF are shown in Table 6.7.
- Receptor models have followed a logical evolution which has derived models from basic assumptions, defined measurements intended to fulfill the requirements of the models, recognized that certain of the assumptions are not met, and modified the models and measurements to accommodate that recognition.
- Unmix does not make any assumptions as to the number and composition of the sources, relying instead on the correlations of the observed species. The species concentrations are apportioned by a principal components analysis using constraints to assure non-negative and realistic sources compositions and contributions.
- Whereas PMF assumes that emission sources is constant over the period of sampling at the receptors, chemical species used in PMF do not interact with each other and their concentrations are linearly additive, source profiles are linearly independent of each other and the numbers of species is greater than or equal to the number of sources.
- Several assumptions of existing models are not met in real world applications, and the evolution continues by determining the effects of deviations from them on model results and by modifying the assumptions, the models, and the measurements to better represent reality.

6.6 Conclusions

- Receptor modeling through APCS-MLR provided apportionment of various sources/factors responsible the groundwater quality characteristics of the study area. The percentage contribution of the identified sources to each parameter was determined with respect to the sources identified by PCA.It was also found out that some parameters received the significant contribution from the unidentified sources.
- Unmix model identified six sources using 16 groundwater quality parameters. A total of 4 variables were excluded by both because of low signal to noise ratio.
- Through comparison of Unmix and PMF, we found that five of the seven sources, including natural source, chromium electroplating, sewage, geologic, lead acid battery manufacturing, have good Unmix-derived counterparts. However PMF appointed two anthropogenic sources namely paint shipping and steel processing plants.
- Overall, three different receptor models (APCS-MLR, Unmix and PMF) were applied to apportion the groundwater quality in of Peenya industrial area, Bangalore with an intention to compare the outcomes and the adequacy of models. The nature of the models was portrayed by regressing the parameters modeled for each model against the one quantified gave great outcomes in regards to their capacity to recreate measured concentrations with fundamentally alike sources in most of the cases but with the, yet with the APCS-MLR model demonstrating the best correlation and the nearest slope to the unity.

CHAPTER 7

SUMMARY AND CONCLUSIONS

7.1 INTRODUCTION

The primary objective of this research was identification and apportionment of pollution sources to groundwater of Peenya industrial area, Bengaluru using multivariate statistical techniques. In general, the study confirmed the usefulness of multivariate statistical techniques such as cluster analysis, discriminant analysis, principal component analysis and APCS-MLR, Unmix and Positive Matrix Factorization in handling and interpreting complex environmental data to draw meaningful conclusions from it. In fact studies conducted earlier on this region came up with general prediction of sources causing pollution to groundwater but information regarding quantification of pollution sources was not available. In this study, using multivariate statistical techniques, different aspects of groundwater quality were studied. Information regarding spatial and temporal variation, most significant parameters, underlying factors and source contributions from the different factors, was gathered from the analyses.

IMPORTANT CONCLUSIONS DRAWN OUT FROM THIS STUDY ARE:

Objective: Basic Statistical Analysis

- Basic statistical analysis revealed that five groundwater quality parameters (turbidity, total hardness, iron, manganese chromium) considered for the study were exceeding permissible limit ,especially chromium whose average concentration was 5.21 mg/L. The heavy metal concentration indicated pollution from anthropogenic source.

- A correlation matrix of variables was calculated to distinguish several relevant hydro chemical relationships. A strong positive correlation between turbidity and sulphate (0.82), total dissolved solids and chloride (0.81), calcium and hardness (0.81), magnesium and chloride (0.81), hardness and magnesium (0.86), were found which are responsible for water mineralization.

Objective: Application of Multivariate statistical techniques to groundwater quality data.

- Cluster Analysis was useful in classifying the 41 sampling sites into 3 main clusters as high pollution and low pollution areas. This helps in the identification of problematic zones in the area where remedial actions need to be focused. Also, grouping the areas having similar groundwater condition may be used to determine the number of sampling sites required for regular monitoring of groundwater quality.
- DA was useful in identifying the indicator parameters which were causing large variations (spatial and temporal) in groundwater quality in the study area. T-Hard, NO_3, Ca, Mg, HCO^3 and TDS were found to be the most important parameters to discriminate between three different seasons and accounted for 94% assignation of seasonal cases, thereby causing the temporal variations in the groundwater quality. Fe, Cr, Cl, Mn, Cu and Cd as the most important parameters discriminating between the 3 clusters and accounting for 92% spatial assignation of cases. Therefore, discriminant analysis permitted lessening in the dimensionality of the substantial huge dataset, portraying a couple of marker parameters causing spatial and temporal in the groundwater quality.

- Principal component analysis aided in finding the 7 factors/sources explaining 73.43% of the overall variance. Varifactors attained from principal component analysis pointed out that the groundwater quality variations are mainly explained by mineralization, sewage and industrial activity in the area especially the electroplating industries which are responsible for high heavy metal concentration in the groundwater content.

- Most of the units in the study area were found engaged in hard chrome plating, copper plating, zinc plating and nickel plating and same is contributing pollution load in terms respective heavy metal in the effluent.

Objective: Application of Receptor oriented source apportionment models to quantify source contribution

- Receptor modeling through APCS-MLR provided apportionment of various sources/factors responsible the groundwater quality characteristics of the study area. The percentage contribution of the identified sources to each parameter was determined with respect to the sources identified by PCA.It was also found out that some parameters received significant contribution from unidentified sources.
- Unmix model identified six sources using 16 groundwater quality parameters. A total of 4 variables were excluded by both because of low signal to noise ratio. For Unmix model the ratio of mean observed and measured values of most of the water quality variables. Based on the R^2 values the accuracy of the model is very high for SO_4, Cl, HCO_3, S^{--}, Cu, Fe and Cd with R^2 between 0.8- 1, high for NO_3, TH, Ca, Mg, F, Zn, Mn and Pb suggesting goodness of the receptor modelling approach to the source apportionment of groundwater.
- In PMF modeling all the elements of groundwater samples had scaled residuals within 3.0 and −3.0, indicating the well modeled and thus the data classification was not changed. Based on the R^2 values the accuracy of the model is very high for

pH, Cl, TH, HCO_3, and Zn with R^2 between 0.8- 1, high for Turbidity, TDS, SO_4, NO_3, Ca, Mg, Fe, Mn and Cr suggesting goodness of the receptor modelling approach to the source apportionment of groundwater.

- The contrast with respect to the distinguished sources by every receptor model was basically because of the contemplations of the models to pick the species chosen as variable. Although APCS-MLR, Unmix and PMF identified similar sources, the Unmix model did not recognize the steel processing industry and spray painting sources. It was conservative and recognized a lesser number of sources, six versus seven from the PMF and APCS-MLR model. This limitation can be identified with the model since it doesn't consider the uncertainty in the dataset, yet it is extremely delicate to this one, barring a few factors.
- Overall, three different receptor models (APCS-MLR, Unmix and PMF) were applied to apportion the groundwater quality in of Peenya industrial area, Bangalore with an intention to compare the outcomes and the adequacy of models. The nature of the models was portrayed by regressing the parameters modeled for each model against the one quantified gave great outcomes in regards to their capacity to recreate measured concentrations with fundamentally alike sources in most of the cases but with the, yet with the APCS-MLR model demonstrating the best correlation and the nearest slope to the unity.
- Receptor models are regularly used to distinguish source contributions. To date, a few such models have been well known in light of their physical assumptions and limitations. Be that as it may, the contrast among the outcomes or results of various models are essential to better comprehend source apportionment.
- In light of the information obtained from PCA and in this way the contributions computed from APCS-MLR, PMF and Unmix models more sound and stringent water quality administration designs can be actualized to basic contamination zones causing groundwater quality pollution.

7.2 RECOMMENDATIONS

- The study concludes that there is insufficient drainage facilities in the study region. Substitution of impaired pipelines and coating of sewer channels is important to keep the spillage of sewage in pipes and drainage through unlined channels and keep the intermixing of sewage and groundwater. The consideration of concerned experts must be to make fitting strides to supply of safe drinking water to the general population.
- The wastewater produced from the industrial activities ought to be legitimately treated and arranged off. Also, strict enactment on enterprises setting up and working their effluent treatment plants ought to be authorized obligatorily and reasonable measures ought to be taken against the businesses abusing the ETP standards. Any laxity with respect to the experts may prompt further disintegration in the nature of groundwater.
- Solid waste disposal on a landfill is high aggravation and turns into a hotspot for ground water defilement and water-borne infections. A legitimate administration of these exercises by the concerned specialists must be made for safe transfer of solid wastes.
- Intense industrial activities coupled with poor disposal methods are the major threats for groundwater pollution of this region.

7.3 LIMITATIONS OF THE STUDY

- For good results the models require large data sets on physical and chemical analysis.
- Receptor models cannot identify any point source as well as cannot deal with source co-linearity very efficiently.
- The fate of the model results can be no better than the input data supplied to the model.
- Drop in the accuracy of the results generated by the model because of low signal to noise ratio can remove certain parameters which carries useful information and are prominent indicators of pollution.
- Numerous solutions can exist with the PMF, and results rely upon decision of number of sources
- Receptor models are not considered to be a cure-all for all circumstances. They are restricted spatially and temporally to the specific arrangement of tests being examined. Clearly they can't be utilized to foresee the effect of future sources.

7.4 SCOPE FOR FUTURE WORK

The results in this thesis are based receptor oriented source apportionment techniques which is top-down approach which provides data on the kinds of discharge sources and their respective contributions to measured contamination, which thus recognizes and quantifies the sources that would be best to control. Future research can be done using Bottom-up methods which identify pollution sources and estimate emission factors using dispersion models.

Other multivariate receptor models include specific rotation factor analysis, target transformation factor analysis, three-mode factor analysis, source profiles by unique ratios (SPUR), and receptor model applied to patterns in space (RMAPS) can be adopted for future work.

Leaching of pollutants into the subsoil and groundwater from the surface streams might be examined. Tracer procedure might be utilized to decide the movement of significant pollutants.

The ANN strategy is amazing in forecast of future information basing on the past information provided the quantity of data information is more. So appropriate techniques can be developed to anticipate the various parameters for any future years.

REFERENCES

Adam, M. J. (1998). “The principle of multivariate data analysis in P. R Ashurst & M. J Dennis (Eds)”. *Analytical methods of food authentication* (p. 350) London: Blackie Academic professional.

Alberto, W. D., del Pilar, D. M., Valeria, A. M., Fabiana, P. S., Cecilia, H. A., and de los Ángeles, B. M. (2001). “Pattern Recognition Techniques for the Evaluation of Spatial and Temporal Variations in Water Quality. A Case Study: Suquía River Basin (Córdoba–Argentina)”. *Water Research*, *35*(12), 2881-2894.

Ali, R., McFarlane, D., Varma, S., Dawes, W., Emelyanova, I., and Hodgson, G. (2012). “Potential climate change impacts on the water balance of regional unconfined aquifer systems in south-western Australia”. *Hydrology and Earth System Sciences*, *16*(12), 4581-4601.

Amadi, A.N., Ameh, M.I and Olasehinde P. I. (2010). “Effect of Urbanization on Groundwater Quality within Makurdi Metropolis, Benue State” .Proceedings on Annual Conference of the Nigerian Association of Hydrogeologists on Water Resources Development and Climate Change.p49.

Anderson, M. J., Daly, E. P., Miller, S. L., and Milford, J. B. (2002). “Source apportionment of exposures to volatile organic compounds: II. Application of receptor models to TEAM study data”. *Atmospheric Environment*, *36*(22), 3643-3658.

Andrade, E.M., Pala´ cio, H.A.Q., Criso´ stomo, L.A., Souza, I.H., Teixeira, A.S., 2005. “I´ndice de qualidade de a´ gua, uma proposta para o vale do rio Trussu. Ceara”. *Rev Cieˆnc Agron*, 36 (2), 135–142 (in Portuguese).

Anttila, P., Paatero, P., Tapper, U., and Järvinen, O. (1995). "Source identification of bulk wet deposition in Finland by positive matrix factorization". *Atmospheric Environment*, *29*(14), 1705-1718.

APHA (2005). Standard methods for the examination of water and wastewater, 20th edn. American Public Health Association, Washington, DC.

Basappa Reddy, M. (2003). "Status of groundwater quality in Bangalore and its Environs. Report, Department of Mines and Geology, Bangalore.

Bengraine, K., and Marhaba, T.F. (2003). "Using principal component analysis to monitor spatial and temporal changes in water quality". *J Hazard Mater*, 100(1-3):179-195.

Boamponsem, L. K., Adam, J. I., Dampare, S. B. E., Owusu-Ansah, E., and Addae, G. (2010). "Heavy metals level in streams of Tarkwa gold mining area of Ghana". *J Chem Pharm Res*, 2, 504-527.

Bhuiyan, M. A. H., Dampare, S. B., Islam, M. A., and Suzuki, S. (2015). "Source apportionment and pollution evaluation of heavy metals in water and sediments of Buriganga River, Bangladesh, using multivariate analysis and pollution evaluation indices". *Environmental monitoring and assessment*, *187*(1), 4075.

Bhutiani, R., Kulkarni, D. B., Khanna, D. R., and Gautam, A. (2016). "Water quality, pollution source apportionment and health risk assessment of heavy metals in groundwater of an industrial area in North India". *Exposure and Health*, 8(1), 3-18.

Bu, H., Tan, X., Li, S., and Zhang, Q. (2010). "Water quality assessment of the Jinshui River (China) using multivariate statistical techniques". *Environmental Earth Sciences*, *60*(8), 1631-1639.

Central Groundwater Board (1999). “Hydrogeolgical report on Bangalore Metropolitan area, Karnataka”, 12–14.

Central Ground Water Board (2008). “Ground Water Information Booklet of Bangalore urban district, Karnataka”.

Central Ground Water Board (2012). “Ground Water Information Booklet Bangalore urban district, Karnataka”.

Central pollution Control Board (2014). “Report on Assessment of Environmental Compliance in Electroplating Units at Bengaluru”.

Chabukdhara, M., Gupta, S. K., Kotecha, Y., and Nema, A. K. (2017). “Groundwater quality in Ghaziabad district, Uttar Pradesh, India: Multivariate and health risk assessment”. *Chemosphere*, 179, 167-178.

Chen, H. Y., Teng, Y. G., Wang, J. S., and Song, L. T. (2013). “Source apportionment of water pollution in the Jinjiang River (China) using factor analysis with nonnegative constraints and support vector machines”. *Environmental Forensics*, 13(2), 175–184.

Chene, J. M. (1996, March). “Planning and managing urban water needs: the new challenge in Managing water resources for large cities and towns”, *HABITAT II Conference* (pp. 18-21).

Chow, J. C., and Watson, J. G. (2002). “Review of PM2.5 and PM10 apportionment for fossil fuel combustion and other sources by the chemical mass balance receptor model”. *Energy & Fuels*, *16*(2), 222-260.

Cloutier, V., Lefebvre, R., Therrien, R., and Savard, M. M. (2008). “Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system”. *Journal of Hydrology*, *353*(3), 294-313.

Department of Mines and Geology. (2003). "Status of ground water quality in Bangalore and its environs".

Department of Mines and Geology. (2011). "Groundwater hydrology and groundwater quality in and around Bangalore city".

Drever, I.J., (1997). "The Geochemistry of Natural Waters". 3rd ed. Prentice Hall, Inc., Englewood Cliffs, NJ.

Federation, W. E., and American Public Health Association. (2005). "Standard methods for the examination of water and wastewater". *American Public Health Association (APHA):* Washington, DC, USA.

Foster, S. S. D.(1990). "Impacts of urbanization on groundwater". *IAHS Publication,* Vol 198, pp187-207.

García, J. H., Li, W. W., Cárdenas, N., Arimoto, R., Walton, J., and Trujillo, D. (2006). "Determination of PM 2.5 sources using time-resolved integrated source and receptor models". *Chemosphere*, *65*(11), 2018-2027.

Gholizadeh, M. H., Melesse, A. M., and Reddi, L. (2016). "Water quality assessment and apportionment of pollution sources using APCS-MLR and PMF receptor modeling techniques in three major rivers of South Florida". *Science of the Total Environment*, *566*, 1552-1567.

Gildemeister, A. E., Hopke, P. K., and Kim, E. (2007). "Sources of fine urban particulate matter in Detroit, MI". *Chemosphere*, *69*(7), 1064-1074.

Güler, C., Thyne, G. D., McCray, J. E., and Turner, K. A. (2002). "Evaluation of graphical and multivariate statistical methods for classification of water chemistry data". *Hydrogeology Journal*, *10*(4), 455-474.

Guo, H., and Wang, Y. (2004). "Hydrogeochemical processes in shallow quaternary aquifers from the northern part of the Datong Basin, China". *Applied Geochemistry*, *19*(1), 19-27.

Guo, X., Zuo, R., Shan, D., Cao, Y., Wang, J., Teng, Y., and Zheng, B. (2017). "Source Apportionment of Pollution in Groundwater Source Area using Factor Analysis and Positive Matrix Factorization Methods". *Human and Ecological Risk Assessment*: An International Journal, (just-accepted).

Hassen, I., Hamzaoui-Azaza, F., and Bouhlila, R. (2016). "Application of multivariate statistical analysis and hydrochemical and isotopic investigations for evaluation of groundwater quality and its suitability for drinking and agriculture purposes: case of Oum Ali-Thelepte aquifer, central Tunisia". *Environmental Monitoring and Assessment*, *188*(3), 135.

Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J. M., and Fernandez, L. (2000). "Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis". *Water Research*, 34(3), 807-816.

Henry, R. C., and Kim, B. M. (1990). "Extension of self-modeling curve resolution to mixtures of more than three components: Part 1. Finding the basic feasible region". *Chemometrics and Intelligent Laboratory Systems*, *8*(2), 205-216.

Henry, R. C. (2003). "Multivariate receptor modeling by N-dimensional edge detection". *Chemometrics and Intelligent Laboratory Systems*, *65*(2), 179-189.

Hassen, I., Hamzaoui-Azaza, F., and Bouhlila, R. (2016). "Application of multivariate statistical analysis and hydrochemical and isotopic investigations for evaluation of groundwater quality and its suitability for drinking and agriculture purposes: case of Oum

Ali-Thelepte aquifer, central Tunisia". *Environmental Monitoring and Assessment*, *188*(3), 135.

Hopke, P. K., Ito, K., Mar, T., Christensen, W. F., Eatough, D. J., Henry, R. C., and Liu, H. (2006). "PM source apportionment and health effects: 1. Intercomparison of source apportionment results". *Journal of Exposure Science and Environmental Epidemiology*, *16*(3), 275.

Hopke, P. K. (2016). "Review of receptor modeling methods for source apportionment". *Journal of the Air & Waste Management Association*, *66*(3), 237-259.

Huang, F., Wang, X., Lou, L., Zhou, Z., and Wu, J. (2010). "Spatial variation and source apportionment of water Pollution in Qiantang River (China) using statistical techniques". *Water Research*, 44(5), 1562–1572.

Jaiswal, R. K., Mukherjee, S., Krishnamurthy, J., and Saxena, R. (2003). "Role of remote sensing and GIS techniques for generation of groundwater prospect zones towards rural development--an approach". *International Journal of Remote Sensing*, *24*(5), 993-1008.

Juahir, H., Zain, S. M., Yusoff, M. K., Hanidza, T. T., Armi, A. M., Toriman, M. E., and Mokhtar, M. (2011). Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. *Environmental Monitoring and Assessment*, *173*(1-4), 625-641.

Kazi, T. G., Arain, M. B., Jamali, M. K., Jalbani, N., Afridi, H. I., Sarfraz, R. A., and Shah, A. Q. (2009). "Assessment of water quality of polluted lake using multivariate statistical techniques: A case study". *Ecotoxicology and Environmental Safety*, 72(2), 301-309.

Kazakis, N., Kantiranis, N., Kalaitzidou, K., Kaprara, M., Mitrakas, M., Frei, R.,and Filippidis, A. (2017). "Origin of hexavalent chromium in groundwater: the example of Sarigkiol Basin, Northern Greece". *Science of the Total Environment*, *593*, 552-566.

Khan, T. A. (2015). "Groundwater quality evaluation using multivariate methods, in parts of Ganga Sot Sub-Basin, Ganga Basin, India". *Journal of Water Resource and Protection*, 7(09), 769.

Khanam, Z., and Singh, V. (2014). "Groundwater quality assessment near polluted canal area in Kichha town, Uttarakhand, India". *International Journal of Recent Scientific Research*, *5*(2), 362-368.

Kim, B. M., and Henry, R. C. (2000). "Application of SAFER model to the Los Angeles PM 10 data". *Atmospheric Environment*, *34*(11), 1747-1759.

Kim, J. H., Kim, R. H., Lee, J., Cheong, T. J., Yum, B. W., and Chang, H. W. (2005). Multivariate statistical analysis to identify the major factors governing groundwater quality in the coastal area of Kimje, South Korea. *Hydrological Processes*, *19*(6), 1261-1276.

Kovács, J., Nagy, M., Czauner, B., Kovács, I. S., Borsodi, A. K., and Hatvani, I. G. (2012). "Delimiting sub-areas in water bodies using multivariate data analysis on the example of Lake Balaton (W Hungary)". *Journal of Environmental Management*, *110*, 151-158.

Kowalkowski, T., Zbytniewski, R., Szpejna, J., and Buszewski, B. (2006). "Application of chemometrics in river water classification". *Water Research*, *40*(4), 744-752.

Kumar, M., Ramanathan, A. L., Rao, M. S., & Kumar, B. (2006). "Identification and evaluation of hydrogeochemical processes in the groundwater environment of Delhi, India". *Environmental Geology*, 50(7), 1025-1039.

Kumar M, Ramanathan AL, Tripathi R, Farswan S, Kumar D, Bhattacharya P (2017). "A study of trace element contamination using multivariate statistical techniques and health risk assessment in groundwater of Chhaprola Industrial Area, Gautam Buddha Nagar, Uttar Pradesh, India". *Chemosphere,* 166, 135-145.

Lawrence, A. R., Morris, B. L., and Foster, S. S. D. (2005). "Hazards induced by groundwater recharge under rapid urbanization". *Geological Society, London, Engineering Geology Special Publications*, *15*(1), 319-328.

Lewis, C. W., Norris, G. A., Conner, T. L., and Henry, R. C. (2003). "Source apportionment of Phoenix PM2. 5 aerosol with the Unmix receptor model". *Journal of the Air & Waste Management Association*, 53(3), 325-338.

Ielpo, P., Cassano, D., Lopez, A., Pappagallo, G., Uricchio, V. F., and De Napoli, P. A. (2012). "Source apportionment of groundwater pollutants in Apulian agricultural sites using multivariate statistical analyses: case study of Foggia province". *Chemistry Central Journal*, *6*(S2), S5.

Li, S., Zhang, Q. (2011). "Response of dissolved trace metals to land use/land cover and their source apportionment using a receptor model in a subtropic river, China". *J. Hazard. Mater*. 190, 205-213.

Liu, C. W., Lin, K. H., & Kuo, Y. M. (2003). "Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan". *Science of the Total Environment*, 313(1), 77-89.

Llamas, R., and E. Custodio, eds. 2003. *Intensive Use of Groundwater—Challenges and Opportunities*. Leiden, Netherlands: Balkema.

Londhe, A; Talati, J; Singh, L. K.; Vilayasseril, M.; Dhaunta, S; Rawlley, B.; Ganapathy, K. K.; Mathew, R. P.(2004), "Urban hinterland water transactions: A scoping study of six class one.

Mahmood A and Kundu Amitabh (2005), "India's demography in 2050: size, structure and habitat, Discussionpaper, IWMI-TATA partners meet - 2005, Anand, Gujarat.

Machiwal, D., and Jha, M. K. (2015). "Identifying sources of groundwater contamination in a hard-rock aquifer system using multivariate statistical analyses and GIS-based geostatistical modeling techniques". *Journal of Hydrology: Regional Studies*, *4*, 80-110.

Mavukkandy, M. O., Karmakar, S., and Harikumar, P. S. (2014). "Assessment and rationalization of water quality monitoring network: a multivariate statistical approach to the Kabbini River (India)". *Environmental Science and Pollution Research*, *21*(17), 10045-10066.

Mustapha, A., Aris, A. Z., Juahir, H., Ramli, M. F., and Kura, N. U. (2013). "River water quality assessment using environmentric techniques: case study of Jakara River Basin". *Environmental Science and Pollution Research*, *20*(8), 5630-5644.

Monteiro, V. P., and Pinheiro, J. C. V. (2004). "Criterion for implantation of technologies destined to desalinated the drinking waters reservoirs that provide the cities of the state of Ceara affected by high levels of salt". *Revista de Economia e Sociologia Rural (Brazil)*.

MOWR (Ministry of water resources) 2000. Annual Report -1999-2000. New Delhi : Government of India

Morris, B. L., Lawrence, A. R., Chilton, P. J. C., Adams, B., Calow, R. C., and Klinck, B. A. (2003). “Groundwater and its susceptibility to degradation: a global assessment of the problem and options for management”. (Vol. 3). United Nations Environment Programme.

Nasir, M. F. M., Samsudin, M. S., Mohamad, I., Awaluddin, M. R. A., Mansor, M. A., Juahir, H., and Ramli, N. (2011). “River water quality modeling using combined principle component analysis (PCA) and multiple linear regressions (MLR): a case study at Klang River, Malaysia”. *World Applied Sciences Journal*, *14*, 73-82.

NIUA (2005), “Status of water supply, sanitation and solid waste management in urban areas”, New Delhi, URL:http://urbanindia.nic.in/moud/what'snew/main.htm.

Nosrati, K., and Van Den Eeckhaut, M. (2012). “Assessment of groundwater quality using multivariate statistical techniques in Hashtgerd Plain, Iran”. *Environmental Earth Sciences*, 65(1), 331-344.

Olson, D. A., and Norris, G. A. (2008). “Chemical characterization of ambient particulate matter near the World Trade Center: source apportionment using organic and inorganic source markers”. *Atmospheric Environment*, *42*(31), 7310-7315.

Omo-Irabor, O. O., Olobaniyi, S. B., Oduyemi, K., and Akunna, J. (2008). “Surface and groundwater water quality assessment using multivariate analytical methods: a case study of the Western Niger Delta, Nigeria”. *Physics and Chemistry of the Earth, Parts A/B/C*, *33*(8), 666-673.

Ouyang, Y., Nkedi-Kizza, P., Wu, Q. T., Shinde, D., and Huang, C. H. (2006). “Assessment of seasonal variations in surface water quality”. *Water Research*, *40*(20), 3800-3810.

Paatero, P., & Tapper, U. (1993). "Analysis of different modes of factor analysis as least squares fit problems". *Chemometrics and Intelligent Laboratory Systems*, 18(2), 183-194.

Paatero, P., and Tapper, U. (1994). "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values". *Environmetrics, 5*(2), 111-126.

Paatero, P. (1997). "Least squares formulation of robust non-negative factor analysis". *Chemometrics and Intelligent Laboratory Systems*, 37(1), 23-35.

Paatero, P. (1999). "The multilinear engine—a table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model". *Journal of Computational and Graphical Statistics*, *8*(4), 854-888.

Paatero, P.. 2007. "User's Guide for Positive Matrix Factorization Programs PMF2 and PMF3, Part 448 1-2: tutorial", University of Helsinki, Helsinki, Finland.http://www.lancs.ac.uk/people/cpaap/pfe/.

Panda, U. C., Sundaray, S. K., Rath, P., Nayak, B. B., and Bhatta, D. (2006). "Application of factor and cluster analysis for characterization of river and estuarine water systems–a case study: Mahanadi River (India)". *Journal of Hydrology*, *331*(3), 434-445.

Pekey, H., Karakaş, D., and Bakoglu, M. (2004). "Source apportionment of trace metals in surface waters of a polluted stream using multivariate statistical analyses". *Marine Pollution Bulletin*, *49*(9), 809-818.

Phansalkar S J, Kher V and Deshpande P (2005), "Expanding Rings of Dryness: Water Imports from Hinterlands to Cities and the Rising Demands of Mega-Cities", in IWMI-Tata Annual Partner's Meet, Anand.

Pius, A., Jerome, C., & Sharma, N. (2012). "Evaluation of groundwater quality in and around Peenya industrial area of Bangalore, South India using GIS techniques". *Environmental Monitoring and Assessment*, 184(7), 4067-4077.

Polissar, A. V., Hopke, P. K., Paatero, P., Malm, W. C., and Sisler, J. F. (1998). "Atmospheric aerosol over Alaska: 2. Elemental composition and sources". *Journal of Geophysical Research: Atmospheres*, *103*(D15), 19045-19057.

Puripus, Soonthornnonda., Erik, R. Christensen. (2008) "Source apportionment of pollutants and flows of combined sewer wastewater". *Water Research*, 42(8), 1989-1998.

Qian, J., Wang, L., Ma, L., Lu, Y., Zhao, W., and Zhang, Y. (2016). "Multivariate statistical analysis of water chemistry in evaluating groundwater geochemical evolution and aquifer connectivity near a large coal mine, Anhui, China". *Environmental Earth Sciences*, *75*(9), 747.

Ramesh, A. (2014). "Remediation of Heavy Metal Contaminated Soil and Ground Water" "Doctoral dissertation", Bangalore University., Bangalore.

Ramesh, A., Prakash, B. N., Sivapullaiaih, P. V., & Sadhashivaiah, A. S. (2012). "Assessment of Ground Water Quality in Designated Peenya Industrial Area and Estate, Bangalore, India–A Case Study". *International Journal of Environmental Protection*.

Rao, N. S., Devadas, D. J., & Rao, K. S. (2006). "Interpretation of groundwater quality using principal component analysis from Anantapur district, Andhra Pradesh, India". *Environmental Geosciences*, 13(4), 239-259.

Rao, Venkateshwara, B.(2016). "Physicochemical analysis of selected ground water samples of Vijayawada rural and urban in Krishna district, Andhra Pradesh, India". *International Journal of Environmental Sciences, 2(2), 722.*

Ravikumar, P., Somashekar, R. K., and Angami, M. (2011). "Hydrochemistry and evaluation of groundwater suitability for irrigation and drinking purposes in the Markandeya River basin, Belgaum District, Karnataka State, India". *Environmental Monitoring and Assessment*, *173*(1), 459-487.

Reisenhofer, E., Adami, G., and Barbieri, P. (1998). "Using chemical and physical parameters to define the quality of karstic freshwaters (Timavo River, North-eastern Italy): a chemometric approach". *Water Research*, *32*(4), 1193-1203.

Roser, D. J., Ashbolt, N., Ho, G., Mathew, K., Nair, J., Ryken-Rapp, D., & Toze, S. (2005). "Hydrogen sulphide production tests and the detection of groundwater faecal contamination by septic seepage". *Water Science and Technology*, 51(10), 291-300.

Sampat, P. (2000). Deep trouble: the hidden threat of groundwater pollution.

Selvaraju, N., Pushpavanam, S., and Anu, N. (2013). "A holistic approach combining factor analysis, positive matrix factorization, and chemical mass balance applied to receptor modeling". *Environmental Monitoring and Assessment*, 185(12), 10115-10129.

Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., and Kouimtzis, T. (2003). "Assessment of the surface water quality in Northern Greece". *Water Research*, *37*(17), 4119-4124.

Singh, K. P., Malik, A., Mohan, D., and Sinha, S. (2004). "Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India): A case study". *Water Research, 38*, 3980–3992.

Singh, K. P., Malik, A., and Sinha, S. (2005). "Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques: A case study". *Analytica Chimica Acta, 538*, 355–374.

Shah, T. (2005). "Groundwater and human development: challenges and opportunities in livelihoods and environment". *Water Science and Technology*, *51*(8), 27-37.

Sharma, M., Marsalek, J., and McBean, E. (1994). "Migration pathways and remediation of urban runoff for PAH control". *Journal of Environmental Management*, *41*(4), 325-336.

Shrestha, S., and Kazama, F. (2007). "Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan". *Environmental Modelling & Software*, 22(4), 464-475.

Shrestha, S., Kazama, F., and Nakamura, T. (2008). "Use of principal component analysis, factor analysis and discriminant analysis to evaluate spatial and temporal variations in water quality of the Mekong River". *Journal of Hydroinformatics*, 10(1), 43-56.

Shukla, S. P., and Sharma, M. (2008). "Source apportionment of atmospheric PM10 in Kanpur, India". *Environmental Engineering Science*, *25*(6), 849-862.

Singh, K. P., Malik, A., and Sinha, S. (2005). "Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques—a case study". *Analytica Chimica Acta*, *538*(1), 355-374.

Skidmore, A. K., Varekamp, C., Wilson, L., Knowles, E., and Delaney, J. (1997). "Remote sensing of soils in a eucalypt forest environment". *International Journal of Remote Sensing*, *18*(1), 39-56.

Song, Y., Xie, S., Zhang, Y., Zeng, L., Salmon, L. G., and Zheng, M. (2006). "Source apportionment of PM2. 5 in Beijing using principal component analysis/absolute principal component scores and UNMIX". *Science of the Total Environment*, 372(1), 278-286.

Su, S. L., Li, D., Zhang, Q., Xiao, R., Huang, F., and Wu, J. P. (2011). "Temporal trend and source apportionment of water pollution in different functional zones of Qiantang River, China". *Water Research*, 45, 1781–1795.

Sujatha, D., and Reddy, B. R. (2003). "Quality characterization of groundwater in the south-eastern part of the Ranga Reddy district, Andhra Pradesh, India". *Environmental Geology*, *44*(5), 579-586.

Suvedha, M., Gurugnanam, B., Suganya, M., and Vasudevan, S. (2009). "Multivariate statistical analysis of geochemical data of groundwater in Veeranam catchment area, Tamil Nadu". *Journal of the Geological Society of India*, *74*(5), 573-578.

Tiwari, T. N., and Mishra, M. A. (1985). "A preliminary assignment of water quality index of major Indian rivers". *Indian J Environ Prot*, *5*(4), 276-279.

Thurston, G. D., and Spengler, J. D. (1985). "A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston". *Atmospheric Environment (1967)*, *19*(1), 9-25.

Tziritis, E., Skordas, K., and Kelepertsis, A. (2016). "The use of hydrogeochemical analyses and multivariate statistics for the characterization of groundwater resources in a complex aquifer system. A case study in Amyros River basin, Thessaly, central Greece". *Environmental Earth Sciences*, *75*(4), 339.

Vega, M., Pardo, R., Barrado, E., and Debán, L. (1998). "Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis". *Water Research*, *32*(12), 3581-3592.

Venugopal, T., Giridharan, L., and Jayaprakash, M. (2008). “Groundwater quality assessment using chemometric analysis in the Adyar River, South India”. *Archives of Environmental Contamination and Toxicology*, *55*(2), 180-190.

Wang, Y., Wang, P., Bai, Y., Tian, Z., Li, J., Shao, X., and Li, B. L. (2013). Assessment of surface water quality via multivariate statistical techniques: A case study of the Songhua River Harbin region, China. *Journal of Hydro-Environment Research*, *7*(1), 30-40.

Xia, L. L., Liu, R. Z., and Zao, Y. W. (2012). “Correlation analysis of landscape pattern and water quality in Baiyangdian watershed”. *Procedia Environmental Sciences*, *13*, 2188-2196.

Yidana. S., Ophori. D., and Yakubo. B. (2008) “A multivariate statistical analysis of surface water chemistry data—The Ankobra Basin, Ghana”. *Journal of Environmental Management*, 86, 80–87.

Yidana, S. M. (2011). “Groundwater flow modeling and particle tracking for chemical transport in the southern Voltaian aquifers”. *Environmental Earth Sciences*, *63*(4), 709-721.

Yang, L., Mei, K., Liu, X., Wu, L., Zhang, M., Xu, J., and Wang, F. (2013). “Spatial distribution and source apportionment of water pollution in different administrative zones of Wen-Rui-Tang (WRT) river watershed, China”. *Environmental Science and Pollution Research*, *20*(8), 5341-5352.

Zaporozec, A., Miller, J. C. (2000). Ground-Water Pollution. UNESCO, Paris, France, 24 pp.

Zhang, Q., Li, Z., Zeng, G., Li, J., Fang, Y., Yuan, Q., and Ye, F. (2009). “Assessment of surface water quality using multivariate statistical techniques in red soil hilly region: a case study of Xiangjiang watershed, China”. *Environmental Monitoring and Assessment*, *152*(1), 123-131.

Zhang, X., Wu, J., and Song, B. (2011). “Application of principal component analysis in groundwater quality assessment. In Water Resource and Environmental Protection (ISWREP”), 2011 International Symposium (Vol. 3, pp. 2080-2083). IEEE.

Zhang, Q., Wang, H., Wang, Y., Yang, M., and Zhu, L. (2017). “Groundwater quality assessment and pollution source apportionment in an intensely exploited region of northern China”. *Environmental Science and Pollution Research*, 1-12.

Zhou, F., Huang, G., Guo, H., Zhang, W., and Hao, Z. (2007). “Spatio-temporal patterns and source apportionment of coastal water pollution in eastern Hong Kong”. *Water Research*, 41(15), 3429–3439.

Zhou, F., Liu, Y., and Guo, H. (2007). “Application of multivariate statistical methods to water quality assessment of the watercourses in Northwestern New Territories, Hong Kong”. *Environmental Monitoring and Assessment*, 132(1), 1-13.

LIST OF PUBLICATION

INTERNATIONAL JOURNAL PAPER

Gulgundi, M. S., & Shetty, A. (2016). "Identification and apportionment of pollution sources to groundwater quality". *Environmental Processes*, *2*(3), 451-461, DOI: 10.1007/s40710-016-0160-4.

(Publisher: Springer Link, Journal: Environmental Processes)

Gulgundi, M. S., & Shetty, A. (2017). "Groundwater Quality Assessment of Urban Bengaluru Using Multivariate Statistical Techniques". *Applied water science*. 8-43, DOI : 10.1007/s13201-018-0684-z.

(Publisher: Springer Link, Journal: Applied Water Science)

Gulgundi, M. S., & Shetty, A. (2017). "Source Apportionment of Groundwater Quality using Unmix and Positive Matrix Factorization". Under review.

(Publisher: Springer Link, Journal: Water, Air, & Soil Pollution)

BIODATA

Name :	Mr. Mohammad Shahid Gulgundi
D.O.B :	14-12-1984
Adress :	R. S. Number 154/1, Plot No 102
	Imam Hussain Colony
	Near Imam Hussain Mosque
	Gadag – 582101
	Telephone : +91 9986220504
Email :	shahid.env@gmail.com
Qualification :	Diploma, B.E, M.Tech (Environmental Engineering)
Area of Specilization :	Multivariate Statistical Techniques, Source Apportionment, water and Wastewater Pollution
Journal Publications :	03